Redes sociales: una introducción

Juan Julián Merelo Guervós, jj@merelo.net,

Depto. Arquitectura y Tecnología de Computadores
Universidad de Granada (Spain)

1. Introducción

Las redes no son sólo eso que se echa al mar para coger acedías, ni siquiera eso que se usa para chatear con los coleguis. Una red es una forma abstracta de visualizar una serie de sistemas, y, en general, casi todos los sistemas complejos.

Las redes, al fin y al cabo, están compuestas de *nudos*, que se llaman habitualmente *nodos*, y de enlaces entre ellos; que se llaman *aristas*, si es que son *flechas* que van de un nodo al otro, con un sentido definido, o bien *arcos*, si es que la relación es recíproca, o por decirlo de otro modo, las flechas tiene puntas en los dos extremos.

Efectivamente, Internet es una red. Simplificando un poco, los *arcos* son los diferentes medios que sirven para enlazar dos ordenadores conectados a la red (inalámbricos o alámbricos), mientras que los nodos son, efectivamente, los diferentes chismes computacionales conectados a la red. Pero también es una red un grupo de páginas web, que usen hiperenlaces para referirse unas a otras. En general, en este caso se tratará de *aristas*, porque los hiperenlaces tienen una dirección definida (de la página que enlaza a la enlazada).

Si esas páginas web están escritas por una sola persona, o son directamente páginas web personales, los enlaces pueden reflejar una relación *social* entre los creadores de la web, que se podría expresar vagamente como *conoce-a*, el escritor de una web que ha incluido un enlace a la segunda web *conoce-al* autor de esa segunda web. Las redes sociales son también redes complejas, aunque usan una terminología ligeramente diferente: los nodos son *agentes*, porque hacen algo, mientras que las aristas o arcos expresan, habitualmente, una relación social tal como *conoce-a*, *es-amigo-de*, o *han-comido-spaghetis-juntos*.

2. Calculando relaciones sociales

El reducir las relaciones sociales a un grafo (es decir, un conjunto de nodos con unas relaciones explícitas entre ellos) permite hacer una serie de estudios sobre esa maraña, de la cual se pueden extraer conclusiones desde simples (cuántas personas intermedias harían falta para conseguir el número de móvil de Beckham) hasta complejas (quién es el *agente* con más influencia dentro de una red social). Veamos una red relativamente simple: la de algunos jugadores de fútbol, con la relación *han jugado en el mismo equipo* en la siguiente figura 1.

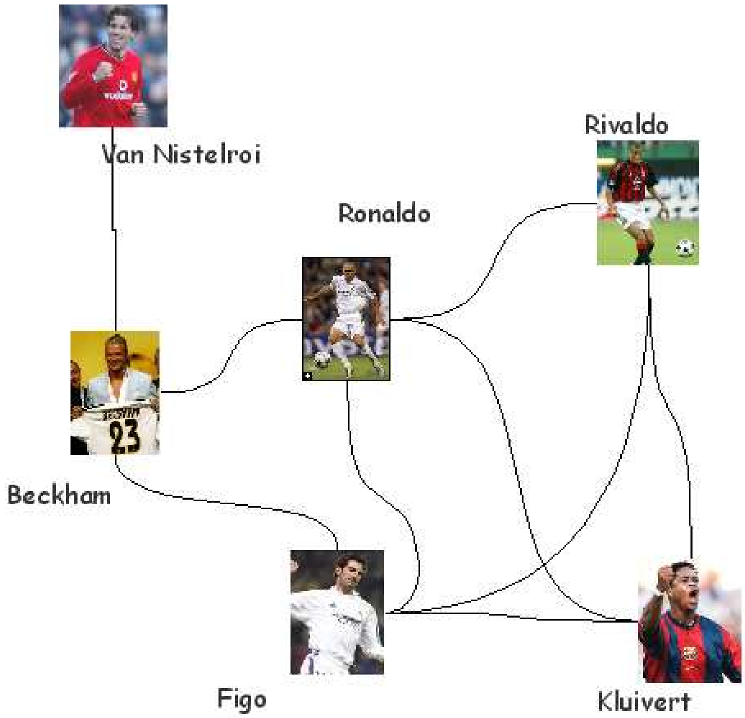

Figura1. Red de relaciones entre 6 jugadores de fútbol: Figo, Van Nistelrooy, Beckham, Ronaldo y Rivaldo. Algunos han jugado juntos en el Real Madrid, otros en el Manchester United, y otros en el Barcelona. Supongo que, esperando suficiente tiempo, todos habrán jugado en el mismo equipo que todos, porque Rivaldo anda ahora en no sé qué equipo griego, Kluiwert en el Valencia, y Figo ya no está en el Madrid. Se excluye también la *diagonal*, porque el hecho de que se relacionen consigo mismo no es interesante.

Lo primero que hay que hacer para analizar esta red es expresarla como una *matriz de contacto*, que tenga como filas y columnas los *actores* o *agentes* de esta red social; esta matriz se muestra en la tabla 1.

Esa tabla se puede introducir en una hoja de cálculo, o en un fichero CSV (*comma separated value*, valores separados por comas (u otro separador, tal como el punto y coma), es decir, cada fila de la matriz en una línea, sus valores separados por comas), de esta forma:

```
Van Nistelrooy ; 0 ; 1 ; 0 ; 0 ; 0 ; 0
```

y usarla en alguno de los programas de análisis de redes sociales que permitirán, para empezar, visualizarla, tal como se ve en la figura 2.

Entre otras cosas que permite apreciar la figura 2 es el hecho de que la red (o *grafo*; en realidad se trata de eso) está conectada; es decir, siguiendo vínculos puedes ir de cualquier jugador a cualquier otro jugador. Si quisieras el móvil de

	Van Nistelrooy	Beckham	Figo	Ronaldo	Rivaldo	Kluiwert
Van Nistelrooy	0	1	0	0	0	0
Beckham	0	0	1	1	0	0
Figo	0	0	0	1	1	1
Ronaldo	0	0	0	0	1	1
Rivaldo	0	0	0	0	0	1
Kluiwert	0	0	0	0	0	0

Cuadro1. Matriz de relaciones entre los diferentes jugadores de la red *social* futbolística. Si existe un enlace entre un jugador y otro, el elemento correspondiente de la matriz tendrá el valor 1. Se muestra (convencionalmente) sólo el triángulo superior de la matriz, aunque se entiende que, tratándose de una matriz simétrica, los elementos por debajo de la diagonal tienen el mismo valor que los que hay por encima de la diagonal.

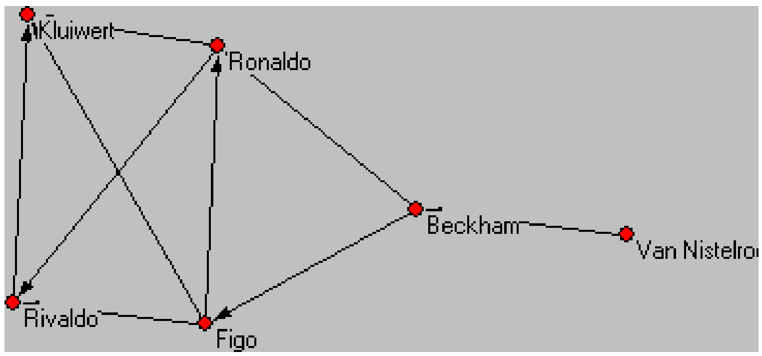

Figura2. Red futbolística representada por el programa Pajek a partir de un fichero tal como el anterior. La simple representación de la red de forma óptima (es decir, de forma que el círculo que representa a cada agente está más cercano a aquellos con los que está unido, y más lejos de aquellos con los que no tiene ninguna relación) permite apreciar estructuras tales como el cuadrado que forman Kluiwert, Ronaldo, Rivaldo y Figo.

Van Nistelrooy, no tendrías más que preguntar a cualquiera de ellos, y acabarían dándotelo, antes o después. Pero evidentemente, el número de pasos intermedios variará dependiendo de a quién le preguntes. ¿Cuánto variará? En media, un programa de análisis tal como el Pajek [Batagelj and Mrvar, 2003] o UCINET [Borgatti et al., 2002] dará un resultado tal como el siguiente:

```
Average distance (among reachable pairs) = 1.533
Distance-based cohesion = 0.389
```

Es decir, que la distancia media entre dos agentes cualesquiera es de 1.533; habrá que usar entre una y dos conexiones para alcanzar a un futbolista cualesquiera (siempre que seas uno de los otros, claro).

Se supone, claro está, que se sigue el camino más corto. A este camino más corto se le denomina *geodésica*, por similitud a las *geodésicas* que son las curvas que siguen el camino más corto entre dos puntos del mapa. Esa distancia media es la media de las distancias más cortas entre todos los elementos de la red, tomados uno por uno. Para calcularlo habría que hacer una matriz similar a la anterior, lo que queda como ejercicio para el lector.

Otra cantidad interesante es el diámetro de la red. ¿Qué tamaño tiene? Así, a ojo de buen cubero, sería 4: la distancia mayor entre dos elementos de la red (siempre que se siga una de esas geodésicas, claro). Esta cantidad es interesante, porque refleja lo *grande* que es la red. Más que tomada de forma aislada, lo que interesa es ver cómo evoluciona el diámetro de la red cuando se añaden nuevos nodos. Podría pensarse que siempre que se añada un nodo nuevo aumentará el diámetro; pero, mirando esta red, sólo aumentará de diámetro si se añade un nodo que esté unido solo a Van Nistelrooy (pongamos, por ejemplo, Michael Owen).

También es interesante ver cómo se agrupan los enlaces de la red. En el caso de Figo y Ronaldo se puede decir aquello de los amigos de mis amigos son tus amigos: todos los que reciben un enlace de uno, reciben un enlace de otro; es decir, el enlazado es *transitivo*. Si Ronaldo enlaza a Kluiwert, y Figo enlaza a Ronaldo, Figo también enlaza a Kluiwert. Sin embargo, no siempre es así: Beckham ha estado en un equipo con Figo, y Van Nistelrooy con Beckham, pero evidentemente, y por la presente, Van Nistelrooy no ha coincidido con Figo (aunque acabarán coincidiendo en algún equipo de Qatar antes de retirsrse, seguro). Esta tendencia a agruparse se denomina *coeficiente de clustering*, y representa la tendencia natural de la gente a *transmitir* relaciones. Un grafo con un coeficiente de clustering alto tendrá relaciones sociales muy tupidas, con todos más o menos relacionados con todos; por el contrario, un coeficiente de clustering bajo representaría relaciones de un tipo particular, generalmente no *transmisibles*.

3. Grafos bipartitos

En realidad, el gráfico con el que hemos estado tratando anteriormente está compuesto de dos tipos de *actores*: los futbolistas y los equipos de fútbol. Cada futbolista estaría relacionado con el equipo de fútbol en el que ha estado, y sólo habría enlaces entre futbolistas y equipos, pero no de los futbolistas o de los equipos entre si. A Este tipo de grafos se les llama *grafos bipartitos*. Los grafos bipartitos tienen dos tipos de nodos, tal como aparece en la figura 3

Los grafos bipartitos, sin embargo, son más fáciles de estudiar convirtiéndolos en unipartitos o grafos *modo-uno*: para ello simplemente se *proyectan*, eliminando los nodos de uno de los dos tipos, y sustituyéndolos por la relación *estar conectados al mismo nodo*: dos nodos *rojos* estarán enlazados si, y solo si, están enlazados al mismo nodo en el grafo bipartito. El gráfico 3 se convierte en el grafo que aparece en la figura 4, de modo 1:

La distinción entre los gráficos de uno u otro tipo es importante, sobre todo, a la hora de calcular un gráfico aleatorio que tuviera las mismas propiedades que el grafo estudiado. Por ejemplo, el coeficiente de clustering, mencionado

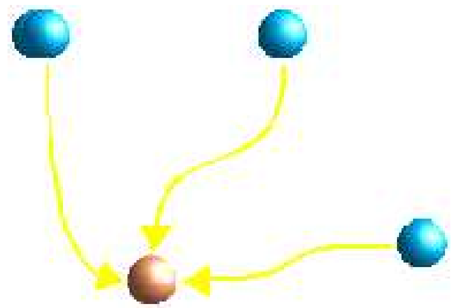

Figura3. Grafo bipartito con nodos de tipo *azul* y nodos de tipo *rojo*. Los enlaces van solamente de los nodos azules a los rojos.

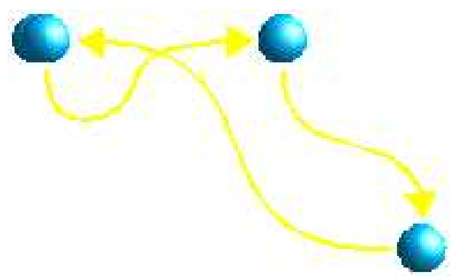

Figura4. Grafo unipartito, o de modo 1, generado a partir del gráfico bipartito representado en la figura 3.

anteriormente, de un grafo generado aleatoriamente será diferente si procede de un grafo bipartito que si procede de uno con nodos de un solo tipo; en el caso del grafo bipartito, evidentemente, será mayor.

4. Mundo pequeño

La casualidad de que dos futbolistas que han jugado en los juveniles del Racing se encuentren de repente jugando en un equipo de la segunda división china hace que se hable de que el mundo es un pañuelo, lo que corresponde a la expresión inglesa *It's a small world*. Quién no se ha montado en un avión, o un autobús, y descubierto que el que está sentado al lado de uno es primo del cuñado del barbero del pueblo donde hizo la mili, o algo así, lo que conduce a la creencia popular de que todo el mundo está a 6 grados de separación, 6 *apretones de manos* de cualquier otra persona, sea un masai de Kenya o un Hmong de las montañas del Yemen.

Al parecer, fue Guglielmo Marconi, el inventor del telégrafo, el que dijo (y esto es posiblemente apócrifo) que dos personas cualesquiera en el mundo estaban separadas por una media de 583 estaciones de telégrafo. Y alguien, en los años sesenta, se puso a probar si tal cosa era cierta. Fue Stanley Milgram, en el primero del que acabó siendo una serie de experimentos [Milgram, 1967], el que intentó medir cuántas personas, efectivamente separaban a todo el mundo de un agente de Bolsa que residía en una ciudad de Masachussets, por el procedimiento de dar a un grupo de personas una carta, y pedir que se la hicieran llegar a ese agente de bolsa de la forma más corta posible, con la condición de que tenían que entregarla en mano. Y vio que, efectivamente, la *moda estadística*, es decir, el número de personas por las que había pasado la carta en la mayor parte de los casos, era (más o menos) cinco (véase figura 5, lo que correspondería a seis grados de separación. De ahí la expresión de los 6 grados de separación, que se convertiría posteriormente en una obra de teatro y película (con Will Smith, Donald Sutherland y Stockard Channing), y en una página web. *6 grados de separación hasta Kevin Bacon*.

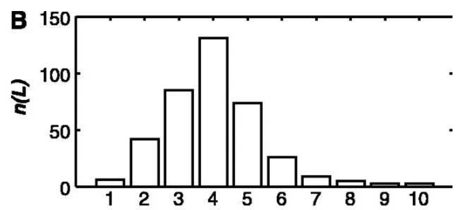

Figura5. Histograma del número de personas por las que pasó la carta en uno de los experimentos de Milgram. De los que efectivamente llegaron a su destino, la moda fue 4 personas, correspondiente a 5 grados de separación.

En este último *juego* se usaba el grafo bipartito de películas y actores, proyectado al grafo homogéneo de actores y la relación *ha participado en la misma película*. El diámetro de este grafo es bastante mayor que los célebres 6 grados, pero Kevin Bacon tiene una posición relativamente buena dentro del mismo: su grado medio de separación a todo el resto del grafo está entre 3 y 4; por lo que, en media, sólo harán falta 3 o 4 películas para encontrar la relación entre Kevin Bacon y cualquier otro actor, sea John Travolta o Tita Merelo[1].

Las redes pequeñas son un pañuelo. Pero redes muy grandes (como la de actores, por ejemplo, o la Internet) pueden serlo también. Partiendo de una red regular, Watts y Strogatz [Watts and Strogatz, 1998] probaron que, añadiendo unos pocos enlaces, el diámetro de la red disminuía drásticamente, mientras que el coeficiente de clustering no variaba sustancialmente; sin embargo, el coeficiente de clustering de una red mundo pequeño es mayor que en una red aleatoria; de forma que, si no se pueden hacer mediciones a gran escala de la red, partiendo del coeficiente (conocido) de clustering de una red aleatoria con el número de nodos y enlaces de la red que se está midiendo, y comparándolo con el coeficiente

[1] La página está en la dirección http://oracleofbacon.org, y Tita Merello (o Merelo) está separada de Kevin Bacon por tres grados de separación solamente

de clustering de esta, se puede tener una primera aproximación a si la red es de tipo *mundo pequeño* o no.

Este tipo de medidas se pueden aplicar también a mundos paralelos. ¿Es un mundo pequeño la red de *conocimientos* de los superhéroes (entendiéndose como tal haber aparecido en el mismo número de un cómic)? Miró et al [Alberich et al., 2002] hicieron diferentes medidas sobre el universo Marvel (que aparece en los cómics de este sello), y encontró que, a pesar de tratarse de un mundo pequeño, su coeficiente de clustering era relativamente pequeño, aunque había ciertas super-estrellas sociométricas: el Capitán América y Spiderman, que habían aparecido en cientos de *team-up* con otros superhéroes. Especialmente el Capi, que ha sido miembro de los Vengadores desde que era cabo.

Las redes mundo pequeño tienen una característica importante: cuando aumentan de tamaño, su diámetro aumenta *lentamente*. ¿Cómo de lentamente? Si pensamos que aumenta linealmente con el número de nuevos nodos, ya nos estamos pasando de rápidos. Incluso si pensamos en la mitad, o en la raíz cuadrada. Aumento *lento*, en este caso, significa que aumenta logarítmicamente. Bastante lento: si se añaden 1000 nuevos nodos, el diámetro podrá aumentar en 3.

A las redes sociales les suele suceder esto; casi todas son de tipo *mundo pequeño*. Otras redes tienden a serlo: por ejemplo, la red de conexiones aéreas de un país como los Estados Unidos, o la red de carreteras de España. Y como en estas, claro está, hay atascos.

En los nodos de las redes mundo pequeño se suele producir un efecto denominado *canalización*. Todos los nodos están conectados a todos los demás a partir de unas pocas conexiones, pero su *alcance* al resto del mundo no está repartido de forma equitativa entre todas ellas. Hay una, o unas pocas, a través de las cuales pasan la mayoría de las geodésicas. De hecho, en el experimento de Milgram se vio algo similar: la mayoría de las cartas le llegaron al agente de Bolsa a través de un colega suyo.

Otra red que se suele someter a estudios habitualmente es la denominada *red de coautorías* en trabajos científicos, en las que los nodos son personas, y los enlaces representan coautoría en un trabajo científico publicado; todos los científicos autores de un trabajo tendrán enlaces entre ellos [Kretschmer, 1997]. Pues bien, en esta red de coautorías de trabajos científicos sucede algo similar: Newman [Newman, 2001] halló que la mayor parte de los *contactos* le llegaban a través de un coautor suyo. Y seguramente tú mismo, lector, si piensas como conociste a tu círculo de amistades te darás cuenta que la mayoría te los presentaron sólo una o dos personas. Este fenómeno lleva a personas como Malcolm Gladwell de hablar del fenómeno de los *conectores* en su libro *The Tipping Point* [Gladwell, 2003]: estos conectores efectivamente unen a gran parte de su red social con el resto del mundo (por el simple hecho de que tienen más conexiones que nadie).

5. Evolución de las redes

Sin embargo las redes se hacen, no nacen, y dependiendo de cómo vayan creciendo, el tipo de red y sus propiedades serán diferentes. Por eso, se han propuesto diferentes modelos de crecimiento de redes. El más antiguo es el de Erdös-Renyi, que es un modelo de crecimiento de redes aleatorias, en el que cada vez que se añade un nodo nuevo, se enlaza a uno aleatorio. Como modelo, no está mal, y tiene propiedades interesantes, pero hay pocas redes en el mundo real que se comporten así, siempre hay nodos más chulos que otros. Por eso, Barabási [Barabási and Albert, 1999,Barabási, 2002] y otros propusieron un modelo denominado de *enlazado preferencial*: los nodos *mejores* se enlazan con más probabilidad que los peores; aunque en realidad, sólo se sabe si son los mejores por el número de enlaces que ya tienen; por lo tanto, es un modelo poco equitativo: da a los que tiene más, aunque por lo menos no quita a los que tienen menos.

Las redes que resultan de estos dos modelos se diferencian, al menos, en lo que se denomina el *componente gigante*, es decir, un grupo de nodos enlazados entre si, y que agrupan a la mayoría de los nodos de la red. Se produce también un efecto *percolación*: llega un punto en el que los diferentes componentes aislados se unen. Sin embargo, en las redes con enlace preferencial el componente gigante aparece con muchos menos nodos, ya que los nodos *entrantes* se conectan con más probabilidad a los nodos que ya tienen muchas conexiones, que, evidentemente, estarán conectados al resto del componente.

En grafos dirigidos se produce una situación similar, la que se muestra en la figura 6

El componente gigante aparece también en casi todas las redes sociales. Por ejemplo, en el estudio de Newman de las redes de coautorías [Newman, 2001] todos los campos tienen un componente gigante que agrupa más del 50 % de los autores; en algunos campos, agrupa casi al 50 % de los mismos. En cuanto a la web, los estudios de Huberman han descubierto que el componente principal gran parte de los sitios web.

6. Leyes de potencias

El que los ricos se hagan más ricos, una situación habitual en muchas redes sociales, lleva a la denominada Ley de Pareto, que se suele enunciar como "El 20 % de la población tiene el 80 % de las riquezas". En realidad, la situación se puede generalizar a lo que se denomina una *ley de potencias*: si la variable independiente y la dependiente se representan en escala logarítmica, la línea que mejor se adapta a los puntos es una línea recta. En el caso de que la variable dependiente sea el *orden* en el que aparecen las cantidades que se representan, se le suele denominar ley de Zipf.

Por ejemplo, muchas redes siguen una ley de Zipf en el número de enlaces. Los enlaces del nodo más enlazado son un múltiplo fijo del segundo más enlazado

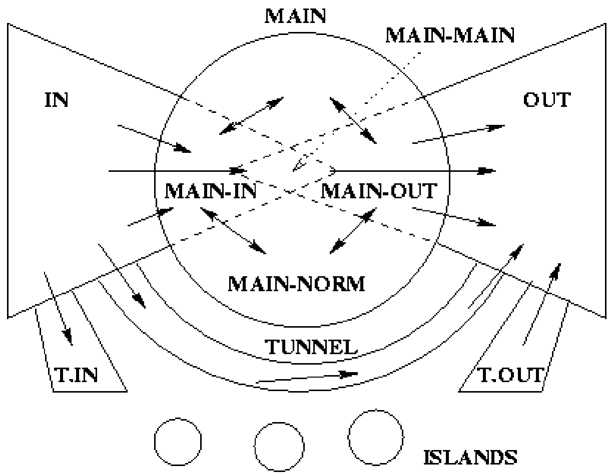

Figura6. Esquema de un grafo dirigido *maduro*. Aparte del componente gigante o principal (en el centro, etiquetado con *main*), hay una parte que sólo enlaza o es enlazado desde él (las *asas*, pequeños componentes que enlazan o son enlazados, pero que están fuera del componente principal, islas y *tentáculos* o *túneles* que enlazan componentes *lejanos*.

(digamos, 2 veces más), éste a su vez del tercero más enlazado, y así sucesivamente. El primero más enlazado tendrá, por tanto, 8 enlaces más que el cuarto más enlazado.

Esto provoca a su vez una serie de fenómenos: nodos más enlazados serán más conocidos, y en páginas webs, o específicamente en bitácoras[2], el número de visitas (de las cuales muchas proceden de las otras bitácoras que le enlazan, o proceden de un buscador y son causadas por el *pagerank* de Google elevado debido a esos mismos enlaces) también seguirá una ley de Zipf, tal como se muestra en la figura 7, extraida de la primera encuesta sobre lectores/autores de bitácoras en la blogosfera (http://tintachina.com/archivo/los_weblogs_mas_leidos.php).

Pero también aparecen leyes de potencias en otros sitios: en la distribución de nodos con un número de enlaces determinado. Representando en abcisas el número de enlaces, y en ordenadas el número de nodos con ese número de enlaces, en las redes con enlazado preferencial, y en muchas redes sociales (por ejemplo, en la blogosfera) se sigue una ley de potencias: habrá muchos nodos con ningún enlace, menos (en una proporción fija) con un enlace, y así sucesivamente. Pero una característica importante de las leyes de potencias es que su valor disminuye más lentamente que una exponencial, lo que da lugar a una *cola larga*, o *fat tail*. Esta cola larga, al contrario de lo que se suele pensar habitualmente, lo que dice es que "sucesos" extremos (en este caso, nodos con un número extremo de enlaces) suceden con una probabilidad no nula. La forma habitual de verlo es darle la

[2] Para una introducción más completa a las bitácoras, ver en este mismo volumen los capítulos escritos por José Luis Orihuela y Fernando Tricas

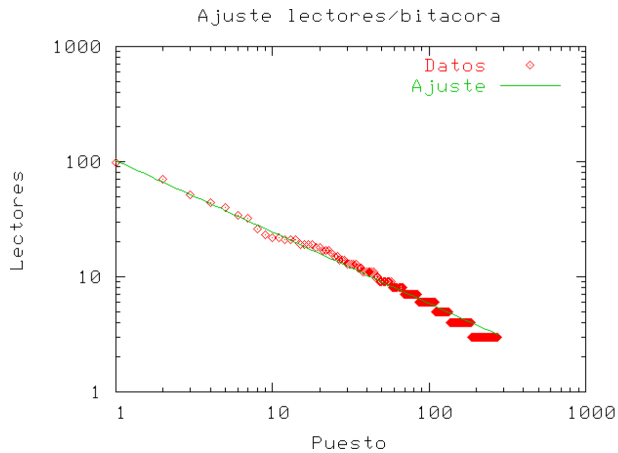

Figura7. Gráfico que representa el número de lectores de un grupo de blogs frente al lugar que ocupan, puestos por orden. Los datos experimentales se representan con un rombo (de color rojo, si ve en color), y el ajuste a una ley de potencias se representa mediante una línea verde. Los datos están extraidos de la *Primera encuesta de lectores/escritores de blogs*, preparada por Tintachina (http://tintachina.com) y Blogpocket (http://blogpocket.com).

vuelta a esta gráfica, y mostrar en abscisas el número de nodos y en ordenadas el número de enlaces: la cola larga nos diría que hay muchos nodos con ningún o pocos enlaces, pero, lo más importante, que si tomamos en consideración el número total de enlaces (o de compras, o de visitas) en los nodos más enlazados, éste será mayor que el número de enlaces total de los nodos más enlazados. Sin embargo, eso depende del exponente de la ley de potencias.

Como se ve en la figura 8, las leyes de potencias se comportan de forma muy diferente dependiendo del exponente. En esta figura se presenta la suma acumulativa de los valores de una función, para x menor que el valor de la abscisa representado. Por ejemplo, se ve que los valores de $x < 100$ acumulan el 5 % del valor (aproximadamente) para la gráfica inferior (en la que el exponente es 1/2), aproximadamente el 20 % para la siguiente, pero casi el 100 % para las dos siguientes (que son leyes de potencias con exponentes mayores que uno). Lo que implica que el fenómeno de la *cola larga funciona* sólo para leyes de potencias con exponente menor que uno: los primeros "nodos" acumulan sólo un porcentaje pequeño del valor. Por ejemplo, en el caso de un exponente igual a 3/4, los valores mayores que 100 acumulan aproximadamente el 80 % del valor total. Una medida no demasiado reciente [Merelo, 2003] da el resultado que se muestra en la figura 9.

Esos puntos experimentales se pueden ajustar según una ley de potencias con un exponente menor que uno: $338x^{-0,58}$, por lo que cabría suponer que en la blogosfera española sí se produce ese fenómeno de *colas largas* mencionado anteriormente. En la blogosfera americana, el exponente negativo es 0.8309, al-

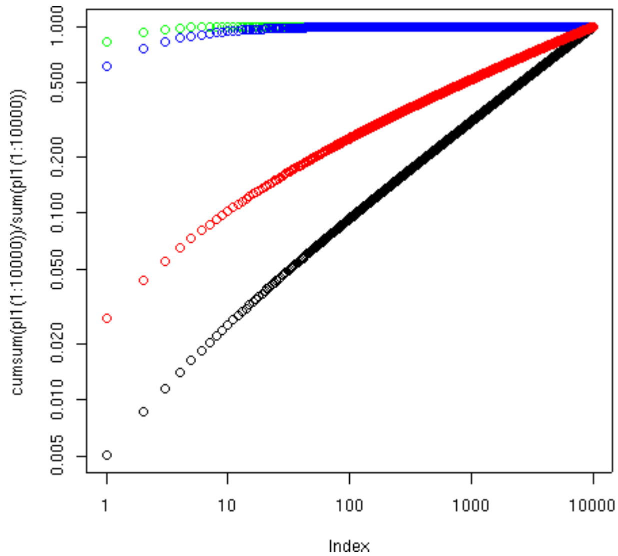

Figura8. Representación de leyes de potencias con diferente exponente, $f(x) = x^p$. En ordenadas se representa la suma acumulativa de los valores $f(x)$, para $x < x_0$. Los dos gráficos que aparecen en la parte baja de la gráfica corresponden a valores de p menores que 1, y los dos en la parte alta, a valores mayores que 1. Una explicación más completa en el sitio donde apareció publicado originalmente, http://atalaya.blogalia.com/historias/22196.

go superior. Es un mundo más cruel, sin lugar a dudas. Con plutócratas más plutócratas.

Por eso es interesante ver según qué ley de potencias se comporta redes tales como la blogosfera. Si el exponente es mayor que uno, aparecerán esas *colas largas*, la blogosfera será igualitaria, y valdrá más ser cola de león que cabeza de ratón. Sin embargo, si el exponente es mayor que uno, los plutócratas blogosféricos acumularán la mayor parte de prácticamente todo: enlaces, visitas y Nokias enviados para que se hable de ellos[3]. Esos plutócratas son los incluidos en la denominada *lista A*, o grupo de bitácoras acaparadoras que están siempre en los *top* de todo. Lo que está bien, pero no hay que olvidar que cualquier medida de la red da una idea estática de la red en cada momento, y que quien tiene más enlaces, o visitas, o comentarios hoy, puede dejar de tenerlos mañana. Así que ánimo, que entrar en la blogosfera es fácil, y cada uno tiene derecho a sus 5 minutos/5000 visitas/500 comentarios de fama (la cola larga proveerá). En todo caso, no está tan claro que la blogosfera, al menos la española, siga una ley de potencias. Más o menos la sigue, y se acerca cada vez más [Tricas and Merelo, 2004].

[3] En marzo de 2004, Nokia envió terminales Nokia 7710 a varios autores de bitácoras en España, Francia y Finlandia; entre los españoles, están varios de los autores de este volumen: Pedro Jorge Romero, José Luis Orihuela, y al autor de este texto.

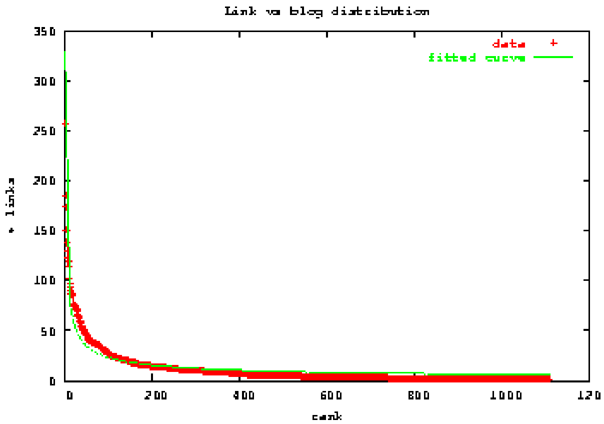

Figura9. Gráfica de puesto vs. número de enlaces entrantes, en la blogosfera española hace un año. Los datos fueron tomados del Blogómetro (http://blogometro.blogalia.com) [Tricas et al., 2003], y por lo tanto, en aquella época, eran los más exhaustivos disponibles. El ajuste según una ley de potencias no es demasiado bueno.

Lo que tampoco está nada claro es que las redes que aparentemente siguen una ley de potencias la sigan de verdad. Y lo es porque en el Mundo Real^{TM} es muy difícil tener todos los enlaces posibles; es difícil ser exhaustivo a la hora de hacer un mapa incluso de los hiperenlaces en un grupo de páginas web. Para empezar, porque un enlace es algo efímero: puede desaparecer en un momento determinado, o puede desaparecer el destino del enlace, o haber cambiado de dirección, o, como mucho, podrás abarcar un grupo finito de páginas web. Todo ello introduce sesgos en el muestreo, y hay quien afirma que muchas de las leyes de potencias medidas son en realidad exponenciales submuestreadas; es decir, que si se usan suficientes nodos y se toman en consideración todos los enlaces, las leyes de potencias desaparecerán. En realidad, el enlazado preferencias no se da en el Mundo Real^{TM}. Para empezar, un pobre nodo que entre en una red no suele tener una visión completa de la misma, por lo tanto no puede enlazar preferencialmente al que más enlaces tenga; o puede que, simplemente, no pueda ver el número de enlaces (lo que es totalmente cierto con el número de enlaces entrantes: es prácticamente imposible calcularlo de forma precisa). Y para seguir, los enlaces desaparecen y se transforman, por lo que, lo que tenemos en un momento determinado, es algo casi totalmente diferente de una red con ley de potencias. Por ejemplo, aparecen leyes log-normales (es decir, distribuciones en las que el logaritmo de la variable sigue una distribución normal, la clásica campana de Gauss: visitar la definición en MathWorld para una explicación más detallada y ver las gráficas) o simplemente exponenciales.

Las redes que siguen una ley de potencias se suelen denominar *libres de escala*, por alguna razón recóndita que indica que no hay una *escala* preferida, o número de enlaces preferido, con respecto a los demás. Las redes aleatorias siguen una distribución de *Poisson* en cantidades tales como el número de enlaces; sin embargo, las redes libres de escala siguen una ley de potencias en la que la *moda* estadística seguida es, en realidad, el número mínimo de enlaces medido (0 o 1), y la media depende del exponente de la ley de potencias, aunque no es un "lugar" ni una "escala" destacada dentro de la red.

En resumen, que las redes complejas, y entre ellas las redes sociales, están cubiertas casi por doquier por leyes de potencias, especialmente en el número de enlaces y cantidades relacionadas con los mismos. Pero esas leyes de potencias nunca están claras, y en algunos casos, pueden aparecer ciertas desviaciones con respecto a la ley de potencias *perfecta*.

7. ¿Cuál es tu red preferida?

Que es como preguntar a quién se quiere más, si al padre o a la madre. ¿Qué redes son más interesantes: las *mundo pequeño* o las *libres de escala*?

En principio, las libres de escala. Muchas redes son mundo pequeño, y eso no las hace especialmente interesantes. Es cuestión sólo de poner unos cuantos enlaces bien dirigidos, unas cuantas "circunvalaciones", y casi cualquier red se puede convertir en una red mundo pequeño.

Sin embargo, las redes que siguen una ley de potencias son bastante más interesantes, porque hablan de unos procesos de evolución algunas veces evidentes, pero en otros caso difíciles de encontrar. Si además tienen la propiedad de ser mundos pequeños, se convierten en doblemente interesantes. Las redes sociales que aparecen en la blogosfera son habitualmente de este tipo, y eso las hace más o menos interesantes.

Estas redes son también interesantes desde otro punto de vista: la vulnerabilidad a *ataques*. Una red ley de potencias es bastante más vulnerable: eliminando sistemáticamente los nodos más conectados, se acaba rápidamente con la conectividad global de la vez, y con menos rapidez, va aumentando el diámetro de la red. Sin embargo, en una red mundo pequeño ese aumento del diámetro es mucho más lento, y hay que eliminar muchos nodos para notar una disminución notable en el diámetro. Por ejemplo, se ha sugerido que para librar a una red de virus [Dezsö and Barabási, 2002] basta con desinfectar y mantener desinfectados los nodos más conectados, es decir, aquellos que tengan más gente en la libreta de direcciones. Sin embargo, la efectividad de esa medida es relativamente pequeña si, además de tratarse de una red libre de escala (como es el caso), es una ley de potencias. Lo mismo ocurre con la red de carreteras: su vulnerabilidad es relativamente grande, dado que se trata generalmente de una red libre de escala. Si se produce una congestión en alguno de los nodos más conectados, la red se viene abajo rápidamente.

8. Buscando el ombligo

Se cree como se cree una red, hay nodos de una red que son más iguales que otros. Los que tienen más enlaces, entrantes, salientes o indiferentes, están destacados, pero no siempre tienen porqué ser los más importantes. Imaginemos una red como una red viaria, por donde tienen que pasar cosas (lo que sucede bastante a menudo). Los nodos con más importancia serán los más inevitables, es decir, aquellos por los que hay que pasar más inevitablemente cuando se vaya de un punto a otro de la red. Lo mismo ocurrirá con los enlaces: los más inevitables tendrán, forzosamente, más importancia, porque, en caso de ser eliminados, dividirán o incomunicarán una parte mayor de la red.

A esta *inevitabilidad* se le suele denominar *centralidad* o *betweenness* (lo que cabría traducir más literalmente como *enmediedad* o *juevidad*, si es que tal palabro existiera en castellano) [Freeman, 1977]. La definición de *enmediedad*, o *betweenness centrality*, es la proporción de geodésicas (recordarlas de 2) que pasan por el nodo o arista.

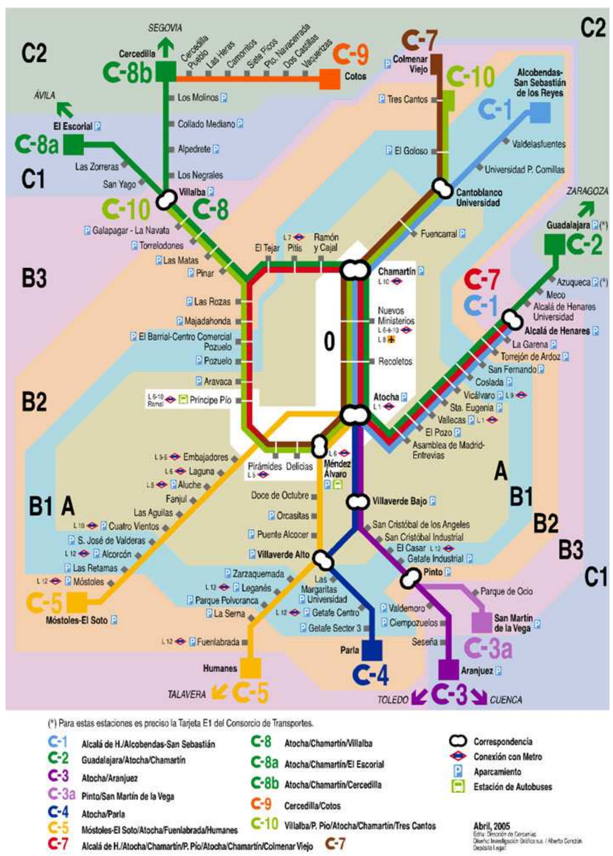

Figura10. Mapa zonal de cercanías en Madrid. Las estaciones con mayor tráfico son también las que tienen más enlaces: las de Atocha y la de Chamartín. Y de hecho, esas son las estaciones con mayor centralidad.

La centralidad es una cantidad bastante intuitiva; sobre todo si hay efectivamente alguna forma de percibir esos "flujos" que pasan por la red. En centro de la ciudad, habitualmente, es la zona con una mayor inmediatez. Y también sucede en las redes de ferrocarriles, tal como se ve en la figura 10. Calculando la centralidad de las estaciones de cercanías, también denominada centralidad de Freeman, se obtiene la figura 11.

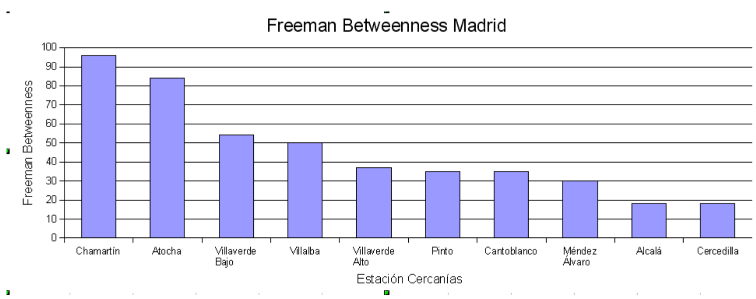

Figura11. Gráfico de betweenness de Freman no normalizada para las estaciones con más valor. Sobresale Chamartín, pero también Atocha y, curiosamente, las estaciones más cercanas a Atocha y Chamartín, que a su vez actúan como conexión entre varias líneas, tales como Villaverde Alto y Bajo y Villalba.

En la figura anterior se muestran los valores de betweenness para las diferentes estaciones. Atocha es la que mayor valor tiene, seguido por Chamartín. Posiblemente por eso los ataques terroristas del 11-M fueron dirigidos a la estación de Atocha, porque intuitivamente es la que mayor tráfico tiene, y además, al cerrarse, paraliza una parte mayor de la red (la línea amarilla o C5, azul o C4, moradao C3... pero también la roja o C7, la verde o C2: prácticamente todas las líneas se verían afectadas por su paralización). Teniendo en cuenta que Atocha es, además, una estación de metro y de ferrocarril, su centralidad con respecto a otras redes es también bastante alta.

En cuanto a personas, la centralidad a veces se suele igualar con *popularidad*, pero también con esos *conectores* de los que hemos hablado anteriormente (en la sección 4). Haciendo una pequeña encuesta en el curso de Nuevas Tecnologías en Internet, en el que se impartió por primera vez este tutorial, se encontró que la red de conexiones es la que aparece en la figura 12.

Y esa figura da lugar al siguiente gráfico, que confirma la impresión de que hay una persona que está *enmedio*: es la *estrella* que aparece en la gráfica 12 y, además, une partes separadas de la red. Suprimiendo ese nodo, cinco nodos (que son un porcentaje considerable de la red) se quedarían aislados, y otra zona (por ejemplo, Eduardo Sánchez) perdería un *atajo* a gran parte de la red. Y eso se ve reflejado en el cálculo de la centralidad, que es el que aparece en la figura 13

Lo de ser inevitable puede ser interesante en caso de que a alguien le interese montar una estación de servicio o un bar de carretera, pero en muchas redes

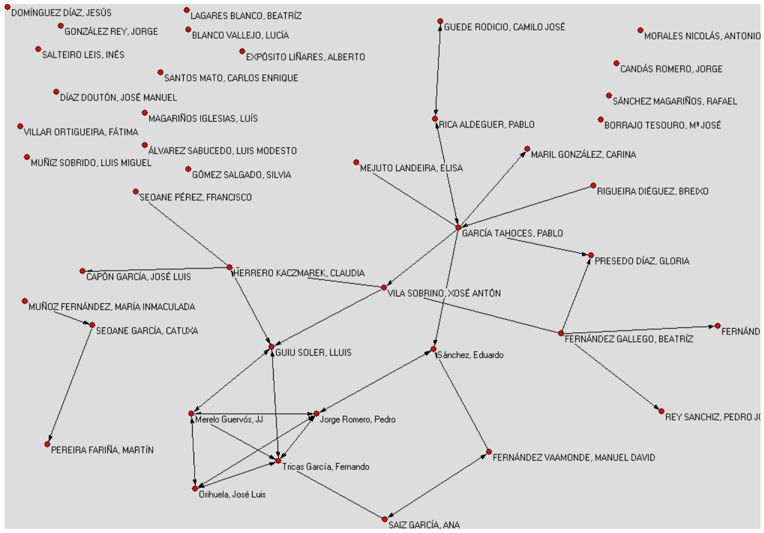

Figura12. Red social del curso de Nuevas Tecnologías en Internet, donde los nodos son el alumnado y profesorado del mismo, y los enlaces representan la relación *conoce a* (previo al curso). Por pura casualidad, los profesores son los que tienen el nombre en minúsculas y los alumnos están en mayúsculas. En redes tan pequeñas es difícil apreciar de qué tipo es, pero sí se aprecian ciertos grupos (como el que forman los profesores) y una *estrella* centrada en Pablo García Tahoces.

lo que interesa es colocarse *cerca* de la mayor parte de la red posible. En este caso el símil de las carreteras nos falla (porque en las carreteras sí interesa la distancia y en las redes, habitualmente, no), y tendremos que recurrir a redes de contactos personales o a redes de comunicaciones. Por ejemplo, a una empresa le puede interesar colocar un centro de datos *data center* lo más cerca posible de sus potenciales clientes, que pueden ser un país o una región completa. En la internet no interesa tanto lo largo que sea el cable sino los "saltos" que tenga que darse para llegar de un punto a otro: habrá, por tanto, que minimizar el número de saltos medio del centro de datos a todos los clientes. A ese número de saltos medio se le denomina *cercanía* en términos de redes, o centralidad de cercanía (*closeness centrality*), y es simplemente la distancia media de un nodo al resto de los nodos de la red. El nodo con una máxima cercanía estará, por decirlo así, en el *centro* de la red, y aunque esta cantidad es difícil de percibir si uno está *dentro* de la red, es más fácil de apreciar en un gráfico que la represente. Además, en redes dirigidas (como en el caso de la red del curso) los resultados de cercanía van en contra de la intuición, porque hay que tener en cuenta la dirección del enlace (una persona puede conocer a otra, pero no al contrario).

Sin embargo, en la red de trenes cercanías de Madrid más o menos la intuición funciona: Atocha y Chamartín son las más *cercanas*, y, curiosamente, están a la misma distancia media del resto de la red. Y aparte de las tres que hemos mencionado anteriormente, se cuela en el ranking Méndez Alvaro, que se sitúa

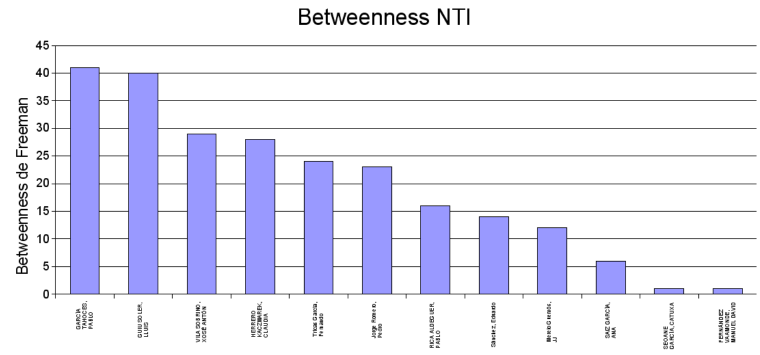

Figura13. Betweenness de Freeman de la red social del curso de Nuevas Tecnologías de Internet.

en un nudo de comunicaciones para conectar sobre todo con el sur, y está suficientemente cerca de Atocha como para estar sólo a un enlace más de distancia que ésta del resto de la red.

En grafos bipartitos se dan otro tipo de relaciones, y no sólo importa la posición dentro de la red, sino hacia dónde apuntan las flechas. Se habla habitualmente de *hubs*, que son personas que *lanzan* muchas flechas (el hub es el *cubo* de las ruedas, de donde salen los radios), o autoridades, que son a las que llegan las flechas. La terminología viene de la web o de la red de trabajos científicos, donde las páginas más apuntadas o los trabajos más citados son los que son considerados *autoridades* en un tema determinado; por otra parte, los *hubs* citan todo lo que debe ser citado. Puede darse en cierto tipo de redes que la *hubidad* y la autoridad se den en un mismo nodo (sucede muy a menudo en páginas web, por ejemplo). Este tipo de personas son a las que se denominan *conectores* en *The Tipping Point*, aunque un valor de autoridad alto correspondería también a quien Gladwell denomina en ese libro *mavens*. Aunque Gladwell hace una clara diferenciación, en realidad no tienen que ser conceptos separados, aunque puede suceder: en el gráfico 14 aparece la parte *contigua* del sociograma del curso, con las personas que actúan como hubs y autoridades. El grupo de autoridades se corresponde prácticamente con el profesorado, que se conocían entre sí, y conocían a pocos alumnos del curso; aunque en este grupo entra también Lluis Guiu, un alumno; y se sale Eduardo Sánchez, el organizador del curso, que aparece como un *hub* o conector, con un papel esencial en la conexión entre alumnos y profesores, y de los profesores entre si.

Dependiendo de la red, la autoridad y la hubidad pueden significar cosas diferentes. Por ejemplo, en la red de los equipos de fútbol durante la Eurocopa 2004 [Lee et al.,], los hubs deberían ser los medios-punta o los defensas centrales, y las autoridades los delanteros. Sin embargo, como se ve en la figura 15, si el hub es un defensa central y la autoridad un extremo como Xabi Alonso (lo que indica que los pases le llegan, pero no salen de él), en vez de un delantero como

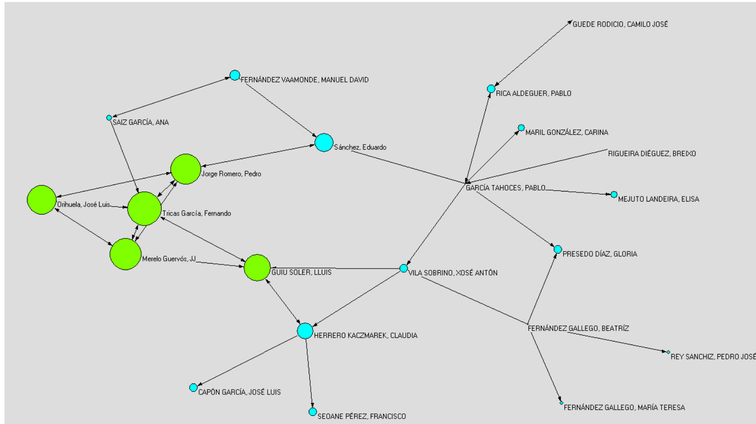

Figura14. Gráfico hecho usando Pajek que representa la red social *contigua* del curso NTI. Diferentes colores (o diferentes tonalidades) representan la cualidad de *hub* (azul) y autoridad (verde); el tamaño indica el valor de esa cantidad. El profesorado son los más conocidos (junto con Lluis Guiu), y por eso aparecen abajo a la izquierda y con círculos de color verde; entre el resto, Eduardo Sánchez que es parte del profesorado, actúa también como *hub*, labor en la que es ayudado por Claudia Herrero, una alumna.

Fernando Torres o Morientes, el equipo tiene un serio problema que explica su eliminación fulminante en la primera fase de la Eurocopa.

9. Resumen

En resumen, el estudio de las redes nos permite comprender cosas totalmente diferentes, desde el hecho de que surjan figuras en la blogosfera hasta la eliminación de la Eurocopa (a posteriori, claro). Estudiar de qué tipo es una red es interesante, porque puede explicar su origen y comportamiento, pero incluso en redes pequeñas estudiar la posición de cada nodo dentro de la misma nos puede ayudar a entender mucho mejor la dinámica de un grupo. Y en el fútbol, ya se sabe que son 11 contra 11, pero si lo miras a través del prisma de una red, puedes leer mejor que Benítez un partido de fútbol.

Agradecimientos

Agradezco a Eduardo Sánchez su invitación a impartir clase en este curso y su ánimo a la hora de escribir este tutorial; a Pedro Jorge Romero su presencia continua en Internet y la parte que le corresponda en la invitación, y a los lectores de mis blogs (Atalaya y BloJJ) sus comentarios en las historias que han sido origen de parte de los datos y los gráficos de este texto. También a José Luis Molina la idea de estudiar los equipos de la Eurocopa usando redes

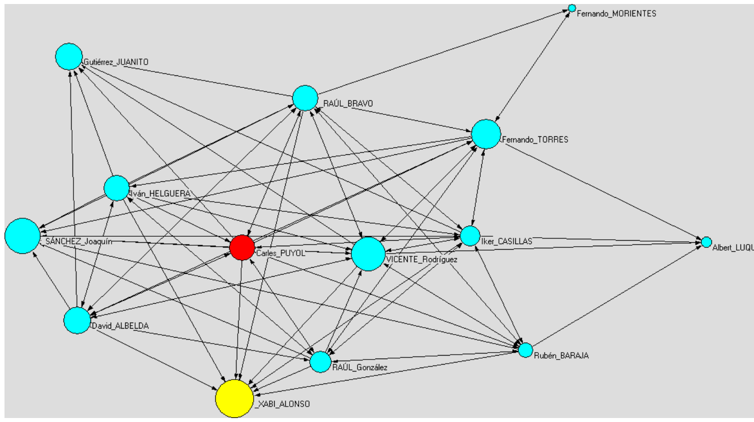

Figura15. Gráfico hecho usando Pajek que representa la red de pases del equipo español en el encuentro España-Portugal en la eurocopa 2004. En color diferenciado se representan los jugadores con el mayor valor de hub (rojo, corresponde a Puyol) y de autoridad (amarillo, corresponde a Xabi Alonso)

sociales. También agradezco a Fernando Tricas la revisión de un borrador, y sus sugerencias.

Referencias

[Alberich et al., 2002] Alberich, R., Miro-Julia, J., and Rossello, F. (2002). Marvel universe looks almost like a real social network. Available from URL: Arxiv.

[Barabási, 2002] Barabási, A.-L. (2002). *Linked-The new science of networks*. Perseus Publishing, Cambridge, MA.

[Barabási and Albert, 1999] Barabási, A.-L. and Albert, R. (1999). Emergence of scaling in random networks. *Science*, 286:509–512.

[Batagelj and Mrvar, 2003] Batagelj, V. and Mrvar, A. (2003). *Pajek. Program for Large Network Analysis*. University of Ljubljana, Slovenia, Ljubljana. To download from http://vlado.fmf.uni-lj.si/pub/networks/pajek/doc/pajekman.htm.

[Borgatti et al., 2002] Borgatti, S., Everett, M., and Freeman, L. (2002). *Ucinet for Windows: Software for Social Network Analysis*. Analytic Technologies, Harvard, MA.

[Dezsö and Barabási, 2002] Dezsö, Z. and Barabási, A.-L. (2002). Halting viruses in scale-free networks. *Physical Review E*, 65. 055103.

[Freeman, 1977] Freeman, L. (1977). A set of measures of centrality based upon betweenness. *Sociometry*, 40:35–41.

[Gladwell, 2003] Gladwell, M. (2003). *The Tipping Point*. Espasa. Cómo pequeñas cosas pueden provocar una gran diferencia.

[Kretschmer, 1997] Kretschmer, H. (1997). Patterns of behaviour in coauthorship networks of invisible colleges. *Scientometrics*, 40(3):579–591.

[Lee et al., 2005] Lee, J., Borgatti, S. P., Molina, J. L., and Guervos, J. J. M. (2005). Who passes to whom: Analysis of optimal network structure in soccer matches. Poster at the Sunbelt XXV conference,.

[Merelo, 2003] Merelo, J. J. (2003). Mapeando la blogosfera hispana II: Ley de potencias. URL: http://atalaya.blogalia.com/historias/7861.

[Milgram, 1967] Milgram, S. (1967). The small world problem. *Psychology Today*, 2:60–67.

[Newman, 2001] Newman, M. (2001). Who is the best connected scientist? A study of scientific coauthorship networks. *Physics Review*, 64(4). Available from http://arxiv.org/abs/cond-mat/0011144/.

[Tricas and Merelo, 2004] Tricas, F. and Merelo, J. J. (2004). The Spanish-speaking blogosphere: Towards the power law? In Kommers, P., Isaías, P., and Nunes, M. B., editors, *Web Based Communities 2004, Proceedings of the IADIS International Conference, Lisbon, Portugal, 24-26 March 2004*, pages 430–433. IADIS. Available from http://webdiis.unizar.es/f̃tricas/Articulos/tricasMereloWBC2004.pdf.

[Tricas et al., 2003] Tricas, F., Merelo, J. J., and Ruíz, V. R. (2003). Do we live in a small world? Measuring the Spanish-speaking blogosphere. In Burg, T.Ñ., editor, *Blogtalks, Proceedings of BlogTalk A European Conference on Weblogs, Viena, Austria, May 23-24, 2003*, pages 158–173. Available from http://www.blogalia.com/pdf/20030506blogtalk.pdf.

[Watts and Strogatz, 1998] Watts, D. J. and Strogatz, S. H. (1998). Collective dynamics of 'small-world' networks. *Nature*, 393:440–442.