

# Minería de Medios Sociales

Máster en Ciencia de Datos e Ingeniería de Computadores

Bloque I: Redes Sociales y Ciencia de Datos en Redes

Sesión I.1: Fundamentos de Redes y Sistemas Complejos. Redes Sociales

Oscar Cordón García

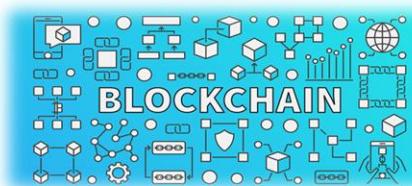
Dpto. Ciencias de la Computación e Inteligencia Artificial. Universidad de Granada  
[ocordon@decsai.ugr.es](mailto:ocordon@decsai.ugr.es)

# **COMPLEJIDAD, SISTEMAS COMPLEJOS Y REDES COMPLEJAS**

# Hay Sistemas Complejos en Todo Nuestro Entorno...

que son muy difíciles de comprender y analizar como:

- Las **sociedades biológicas** (p.e., colonias de hormigas, abejas, manadas, etc.) y la propia **sociedad humana**, cuyo funcionamiento requiere cooperación entre billones de individuos
- El **cerebro humano**, que requiere una actividad coherente de billones de neuronas
- Los **genomas** de los individuos, basados en la interacción constante entre miles de genes
- Los **sistemas financieros**, como por ejemplo el mercado de valores (la bolsa), en los que interaccionan millones de componentes a nivel mundial
- Los **sistemas de transmisiones**, como las redes de telecomunicaciones (ej: Internet) o las **redes de distribución eléctrica**
- etc.



# Complex

[adj., v. kuh m-pleks, kom-pleks; n. kom-pleks]  
—adjective

1.  
composed of many interconnected parts;  
compound; composite: a complex highway  
system.

2.  
characterized by a very complicated or  
involved arrangement of parts, units, etc.:  
complex machinery.

3.  
so complicated or intricate as to be hard to  
understand or deal with: a complex problem.

*Source: Dictionary.com*

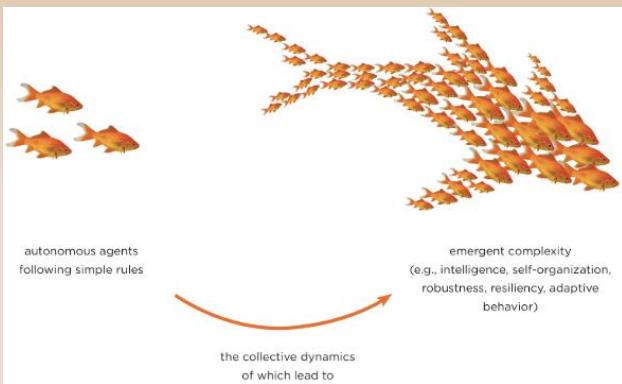
Complexity, a **scientific theory** which asserts that some systems display behavioral phenomena that are completely inexplicable by any conventional analysis of the systems' constituent parts. These phenomena, commonly referred to as emergent behaviour, seem to occur in many complex systems involving living organisms, such as a stock market or the human brain.

*Source: John L. Casti, Encyclopædia Britannica*

# Complexity

# SISTEMAS COMPLEJOS (2)

- Compuesto de agentes simples que interactúan entre sí y con el entorno (**auto-organizado**)
- **Descentralizado:** No existe un supervisor, no hay un control central
- **Comportamiento emergente:** No hay un plan global, se producen acciones complejas e inesperadas a partir de reglas simples
- **Robusto:** Las actuaciones se completan, aunque un individuo falle

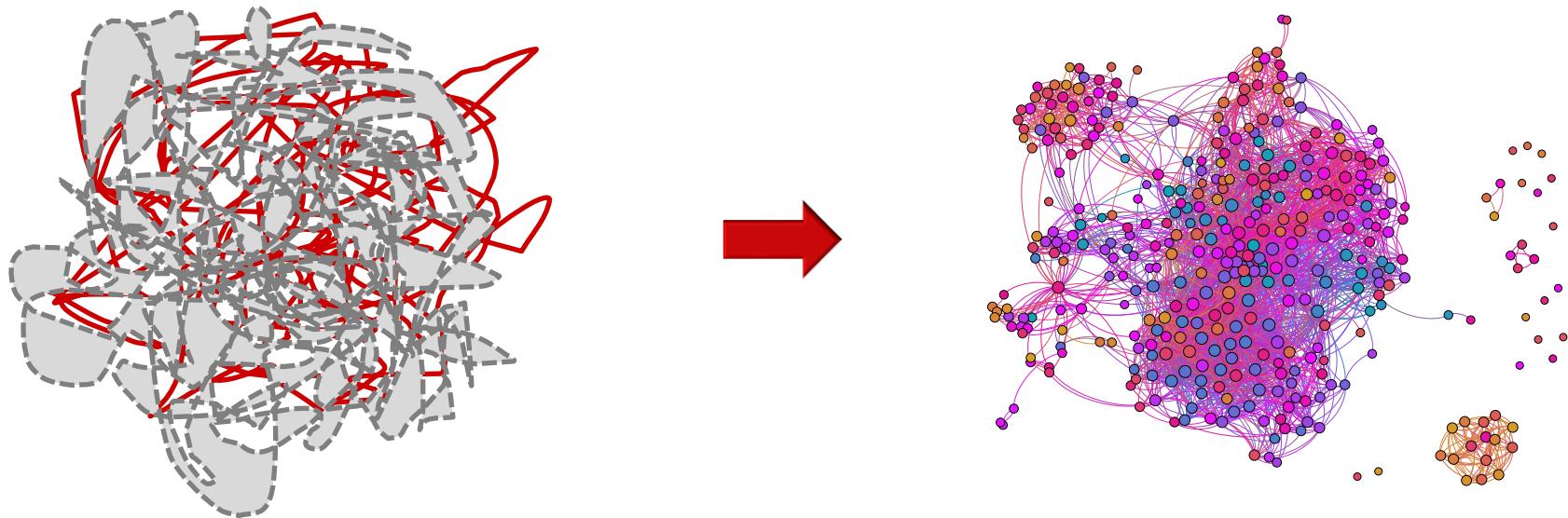




*“I think the next century  
will be the century  
of complexity.”*

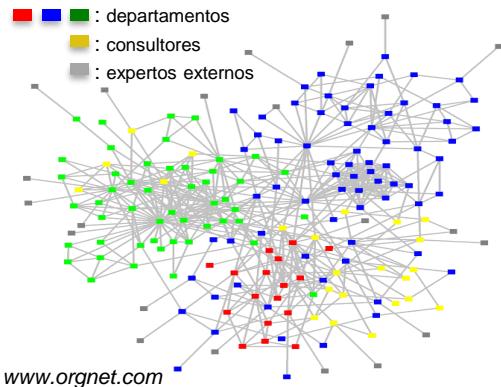
Stephen Hawking  
January 23, 2000

Detrás de cada sistema complejo siempre hay una **red**, que define las interacciones entre sus componentes

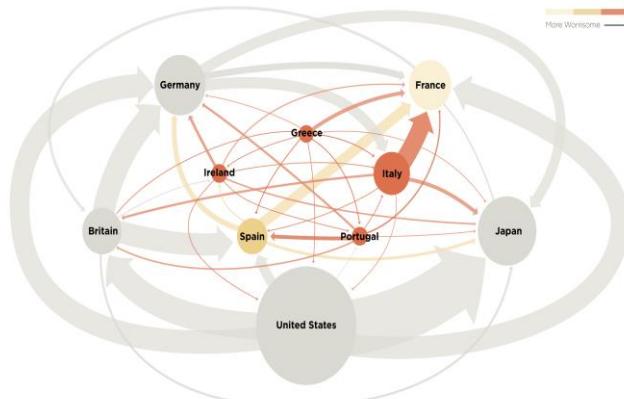


# EJEMPLOS DE REDES COMPLEJAS REALES

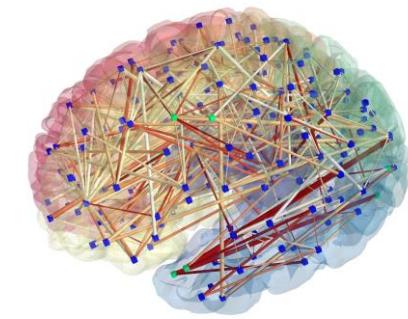
## Organizaciones



## Redes financieras

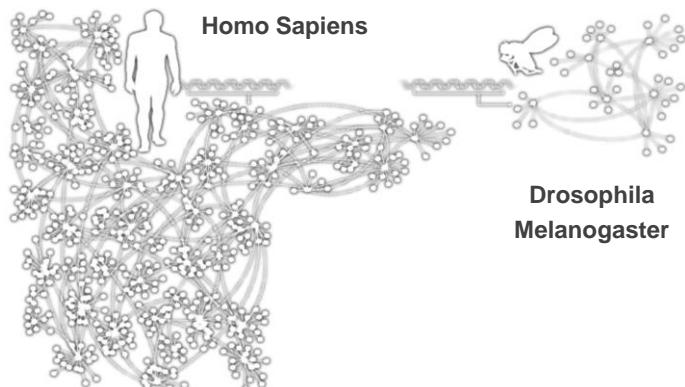


## Redes cerebrales

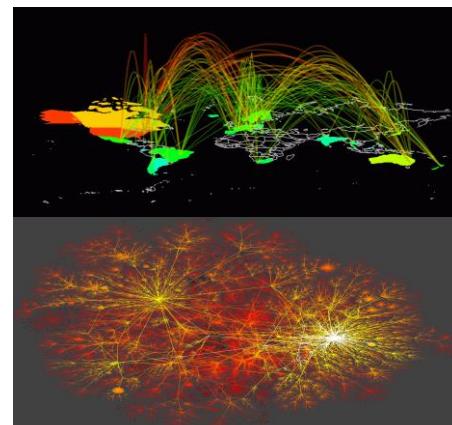


<http://www.humanconnectomeproject.org/>

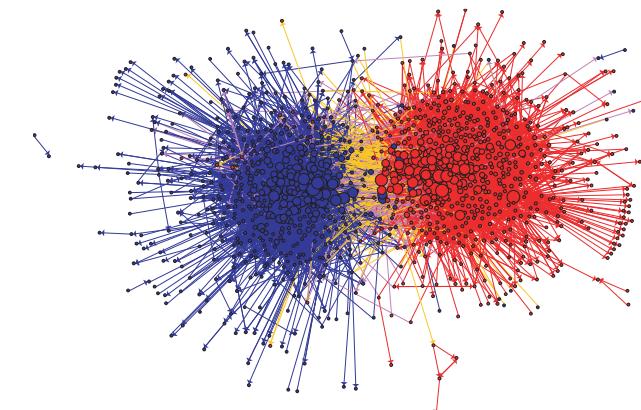
<http://www-personal.umich.edu/~mejn/netdata/>



Redes genéticas



Red física de Internet



Redes políticas

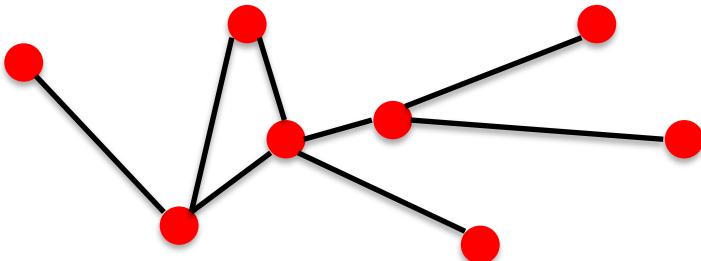
# EL ROL DE LAS REDES

Detrás de cada sistema complejo estudiado siempre hay un diagrama de conexiones, una **red**, que define las interacciones entre sus componentes

**No seremos capaces de entender los sistemas complejos a menos que podamos mapear y comprender las redes que los soportan**

A pesar de las diferencias aparentes en componentes e interacciones, **las redes que regulan los distintos sistemas complejos existentes en nuestro mundo son similares, siguen unas leyes comunes y presentan mecanismos reproducibles**

# COMPONENTES Y NOMENCLATURA DE UN SISTEMA COMPLEJO



- **componentes:** nodos, vértices      N
- **interacciones:** enlaces, arcos      L
- **sistema:**      red, grafo      (N,L)

**red** se suele referir a sistemas reales

- www: red de páginas conectadas por URLs,
- red social: grupo de individuos conectados por una relación familiar, de amistad o profesional,
- red metabólica: conjunto de las reacciones químicas que tienen lugar en una célula

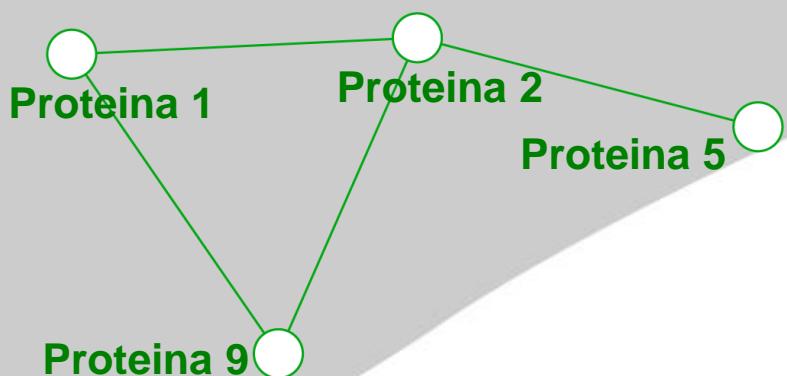
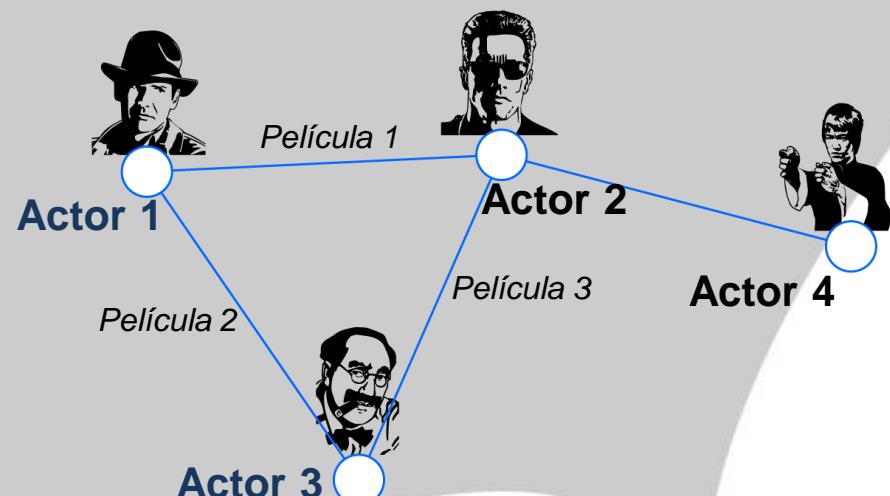
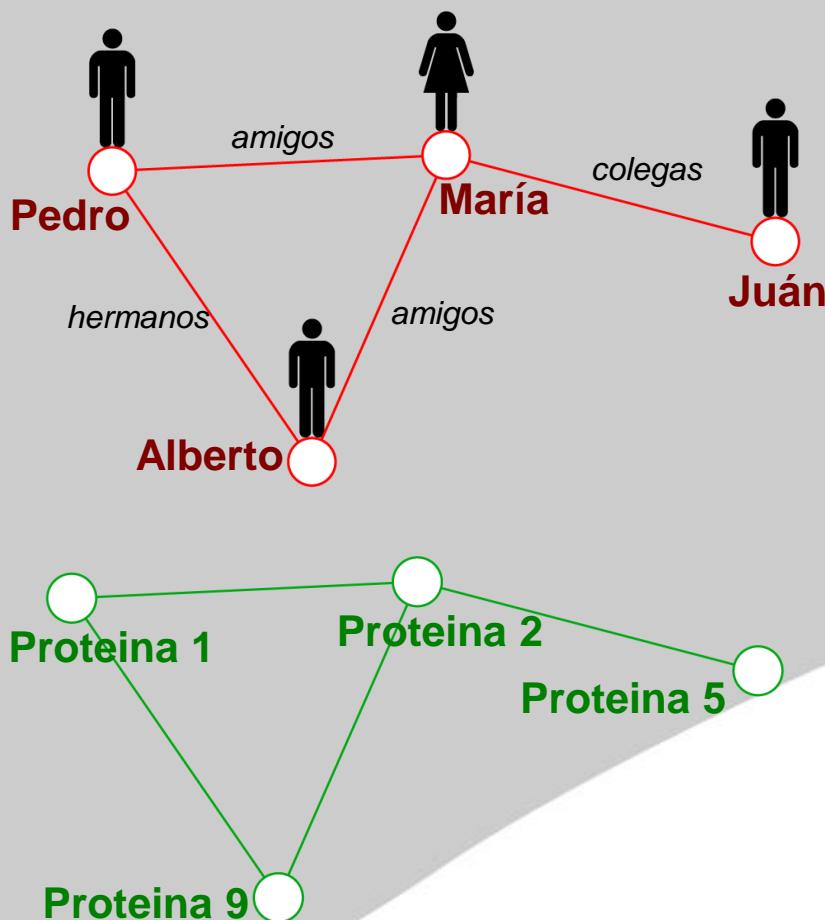
Nomenclatura: (Red, nodo, enlace)

**grafo:** representación matemática de una red

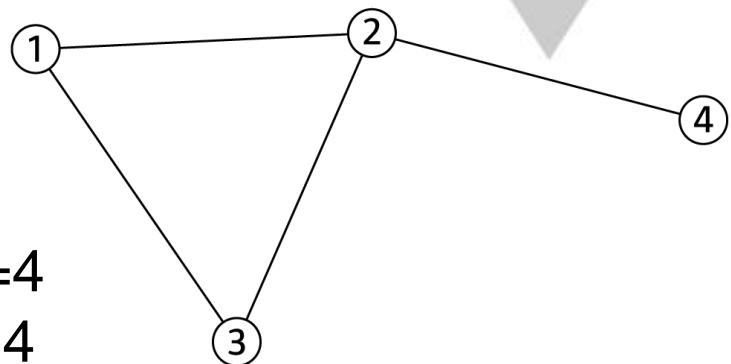
- grafo de la web,
- grafo social (término de Facebook)

Nomenclatura: (Grafo, vértice, arco)

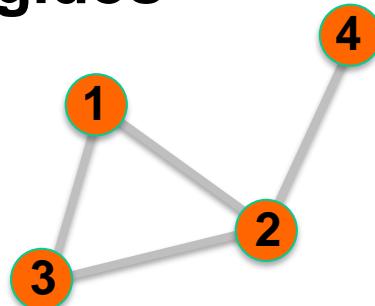
# LAS REDES OFRECEN UN LENGUAJE COMÚN PARA ESTUDIAR SISTEMAS



$$\begin{aligned} N &= 4 \\ L &= 4 \end{aligned}$$



## No dirigidos



$$A_{ij} = \begin{pmatrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}$$

$$A_{ii} = 0$$

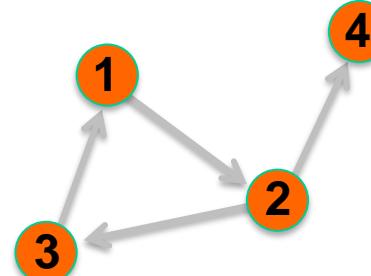
$$A_{ij} = A_{ji}$$

$$L = \frac{1}{2} \sum_{i,j=1}^N A_{ij} \quad \langle k \rangle = \frac{2L}{N}$$

Red de actores de Hollywood

Redes de interacciones entre proteínas

## Dirigidos



$$A_{ij} = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

$$A_{ii} = 0$$

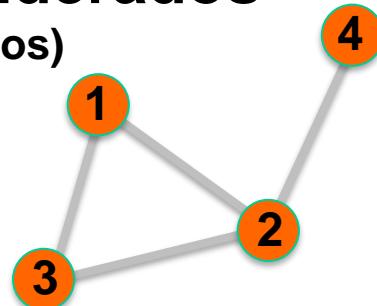
$$A_{ij} \neq A_{ji}$$

$$L = \sum_{i,j=1}^N A_{ij} \quad \langle k \rangle = \frac{L}{N}$$

WWW

Redes de citas en artículos científicos

## No ponderados (no dirigidos)



$$A_{ij} = \begin{pmatrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}$$

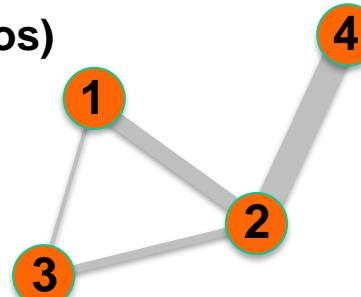
$$A_{ii} = 0$$

$$A_{ij} = A_{ji}$$

$$L = \frac{1}{2} \sum_{i,j=1}^N A_{ij} \quad \langle k \rangle = \frac{2L}{N}$$

Redes de interacciones entre proteínas  
WWW

## Ponderados (no dirigidos)



$$A_{ij} = \begin{pmatrix} 0 & 2 & 0.5 & 0 \\ 2 & 0 & 1 & 4 \\ 0.5 & 1 & 0 & 0 \\ 0 & 4 & 0 & 0 \end{pmatrix}$$

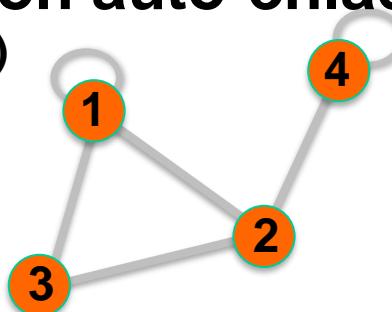
$$A_{ii} = 0$$

$$A_{ij} = A_{ji}$$

$$L = \frac{1}{2} \sum_{i,j=1}^N \text{nonzero}(A_{ij}) \quad \langle k \rangle = \frac{2L}{N}$$

Redes de llamadas  
Redes metabólicas

## Grafos con auto-enlaces (no dirigidos)



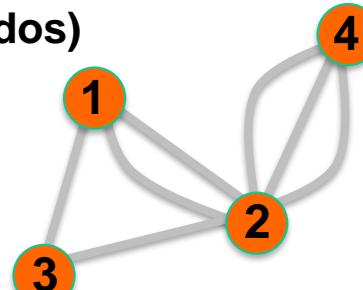
$$A_{ij} = \begin{pmatrix} 1 & 1 & 1 & 0 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 \end{pmatrix}$$

$$A_{ii} \neq 0$$

$$L = \frac{1}{2} \sum_{i,j=1}^N A_{ij} + \sum_{i=1}^N A_{ii} \quad ?$$

Redes de interacciones entre proteínas  
WWW

## Multigrafos (no dirigidos)



$$A_{ij} = \begin{pmatrix} 0 & 2 & 1 & 0 \\ 2 & 0 & 1 & 3 \\ 1 & 1 & 0 & 0 \\ 0 & 3 & 0 & 0 \end{pmatrix}$$

$$A_{ii} = 0$$

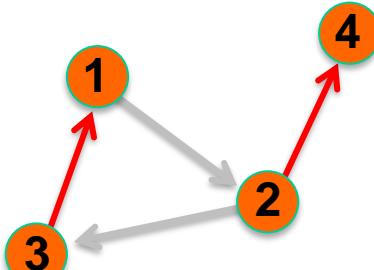
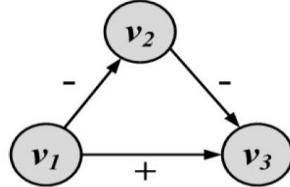
$$L = \frac{1}{2} \sum_{i,j=1}^N \text{nonzero}(A_{ij}) \quad \langle k \rangle = \frac{2L}{N}$$

Redes sociales  
Redes de colaboración

- Representa las conexiones de los sitios web
- Los nodos representan los sitios y los enlaces, hiperenlaces entre sitios web
- Es un **multigrafo dirigido** (pueden existir varios hiperenlaces entre dos sitios) **con auto-enlaces** (puede haber hiperenlaces al mismo sitio web)



## Ponderados con signo (dirigidos)

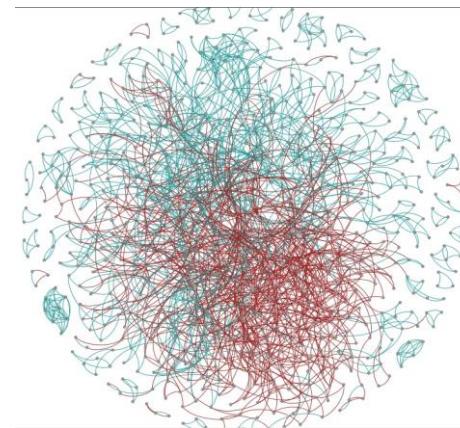


$$A_{ij} = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & -1 \\ -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

$$A_{ii} = 0 \quad A_{ij} \neq A_{ji}$$

$$L = \sum_{i,j=1}^N |A_{ij}| \quad \langle k \rangle = \frac{L}{N}$$

Redes de confianza (+1=amigos, -1=enemigos)  
Redes de status social (+1=mayor, -1=menor)



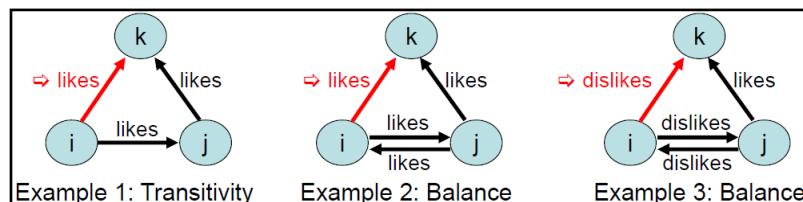
Muestreo de las opiniones positivas y negativas de la red Epinions

<http://blogs.cornell.edu/info2040/2011/09/28/epinions-com-and-the-web-of-trust/>

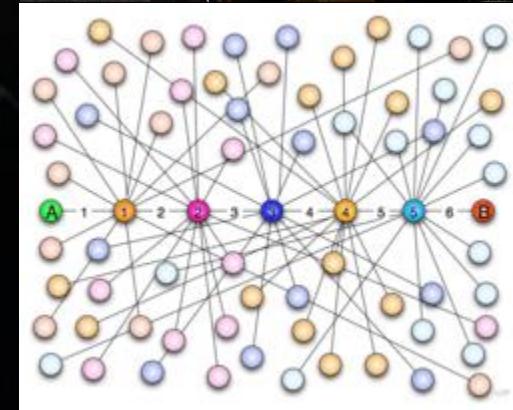
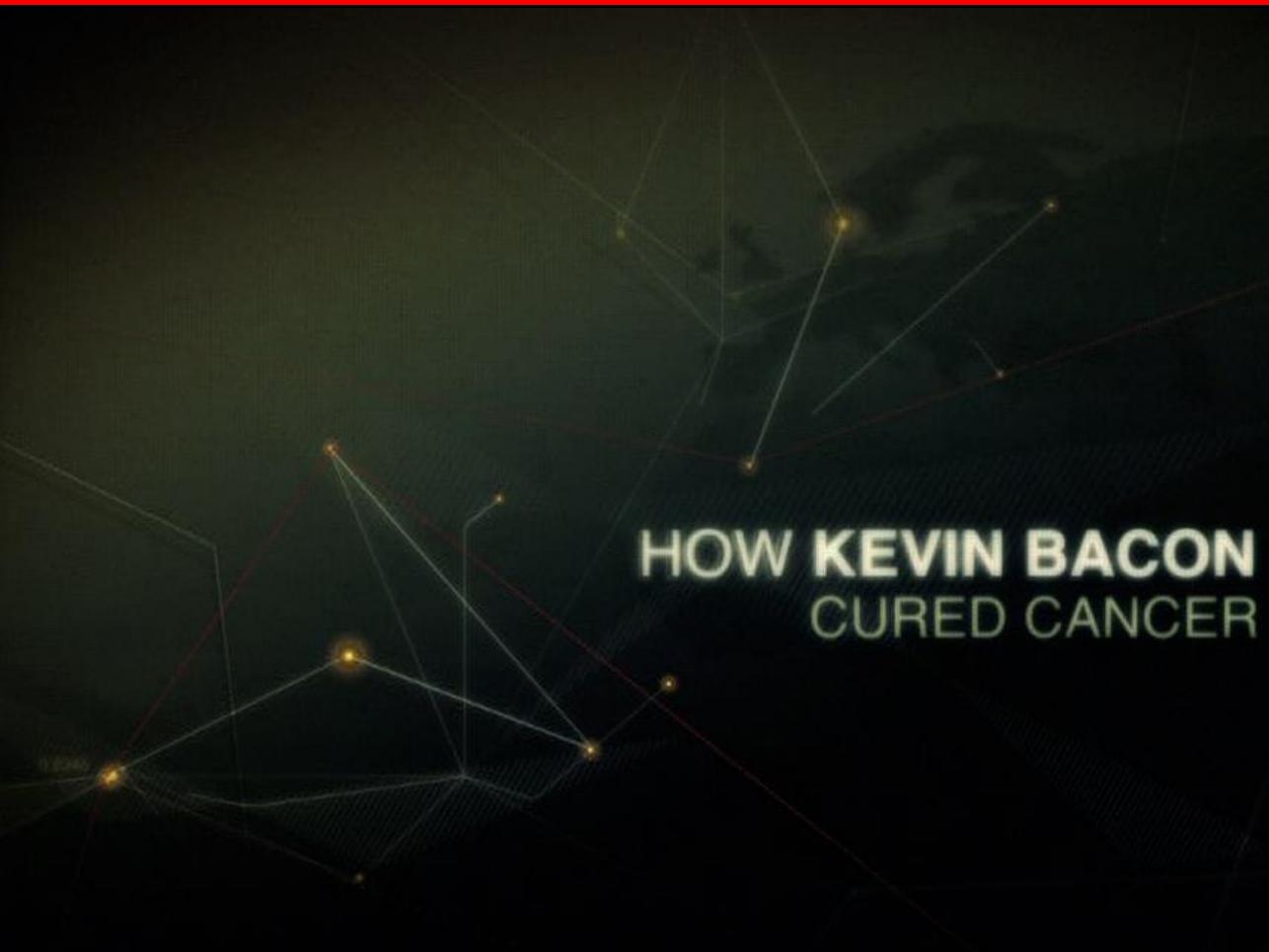
<http://snap.stanford.edu/data/soc-sign-epinions.html>

Opinión de la confianza de una persona sobre otra en el antiguo sitio web de revisión de productos *Epinions.com*

**RETO:** Determinar cómo se propagan los sentimientos negativos en la red. ¿Es mi amigo el enemigo de mi enemigo?



# DIFUSIÓN GENERAL: DOCUMENTAL



Kevin Bacon y la  
Teoría de los seis  
grados de separación

<https://youtu.be/zK1Cb9qj3qQ>

# LAS REDES REALES SON DISPERSAS

La mayoría de las redes observadas en sistemas reales (y en particular las de medios sociales) son dispersas:  $L \ll L_{\max}$  o  $\langle k \rangle \ll N-1$

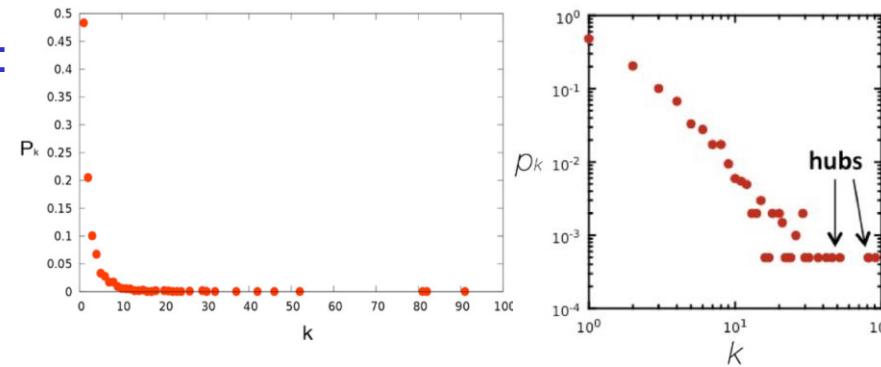
NETWORK NAME	NODES	LINKS	DIRECTED/ UNDIRECTED	N	L	$\langle k \rangle$
Internet	routers	Internet Connections	Undirected	192,244	609,066	2.67
WWW	webpages	links	Directed	325,729	1,497,134	4.60
Power Grid	power plants, transformers	cables	Undirected	4,941	6,594	2.67
Mobile-Phone Calls	subscribers	calls	Directed	36,595	91,826	2.51
Email	email addresses	emails	Directed	57,194	103,731	1.81
Science Collaboration	scientists	co-authorships	Undirected	23,133	186,936	16.16
Actor Network	actors	co-acting	Undirected	212,250	3,054,278	28.78
Citation Network	papers	citations	Directed	449,673	4,707,958	10.47
E. coli Metabolism	metabolites	chemical reactions	Directed	1,039	5,802	5.84
Yeast Protein Interactions	proteins	binding interactions	Undirected	2,018	2,930	2.90

# ANÁLISIS Y MODELADO DE REDES COMPLEJAS

La **Ciencia de Redes** se dedica al análisis de redes reales. Se usan medidas como:

Distribución (de probabilidad) de grados  $p_k$ :

Probabilidad de encontrar nodos con  $k$  enlaces  
(grado medio, **presencia de hubs**, etc.)



Distancia media (de los caminos entre todos los nodos de la red)  $\langle d \rangle$

Coeficiente de Clustering  $C_i$ : Mide la densidad local de la red, ¿qué proporción de los vecinos de cada nodo están conectados?

$$C_i = \frac{2L_i}{k_i(k_i - 1)}$$

$$\langle C \rangle = \frac{1}{N} \sum_{i=1}^N C_i$$

**Los valores de estas medidas en una red concreta determinan el comportamiento dinámico del sistema complejo**

Existen distintos tipos de redes en función de estas características:  
regulares, aleatorias, mundos pequeños, **libres de escala**, ...

$$\langle d \rangle = \frac{\log N}{\log \langle k \rangle}$$

En sociología, una persona cualquiera conoce directamente a otras mil,  $k \sim 1000$

La población mundial está actualmente en torno a los 8200 millones ( $8.2 \cdot 10^9$ )

$$\langle d \rangle = \frac{\ln(8.2 \cdot 10^9)}{\ln(1000)} = 3$$

Experimento basado en e-mails:

**Dodds, Muhamad, Watts (2003).**  
**Science 301**

- 18 objetivos
- 13 países distintos
- Más de 60,000 participantes
- 24,163 cadenas de mensajes
- 384 alcanzaron el objetivo

**Distancia media = 4.0**



Fuente de la imagen: NASA  
[http://visibleearth.nasa.gov/view\\_rec.php?id=2429](http://visibleearth.nasa.gov/view_rec.php?id=2429)

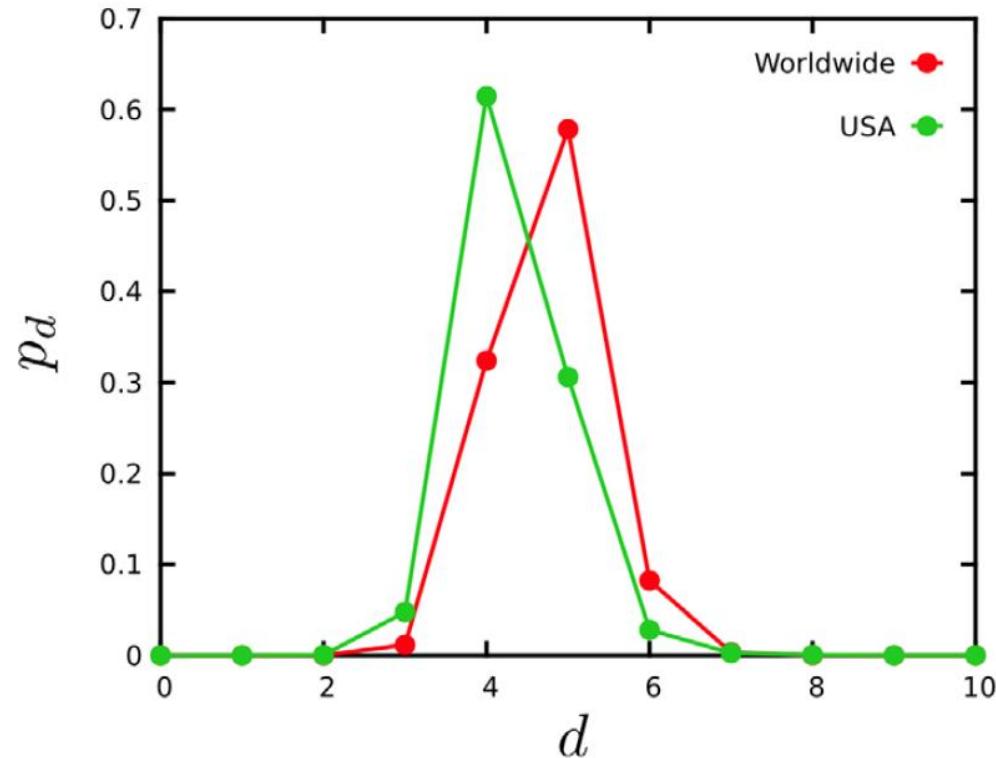
Experimento basado en Facebook:

Backstrom, L., Boldi, P., Rosa, M., Ugander, J. & Vigna, S. (2011). Four degrees of separation. CoRR, abs/1111.4570

El experimento de Milgram no disponía de un mapa adecuado de la red social mundial. Hoy en día, FB es una buena aproximación

Mapa FB Mayo 2011: 721 millones de usuarios activos, 68 billones de relaciones: **distancia media = 4.74**

Este valor más cercano a la distancia media teórica: 3.90



# REDES LIBRES DE ESCALA:

# La Red Compleja de la WWW

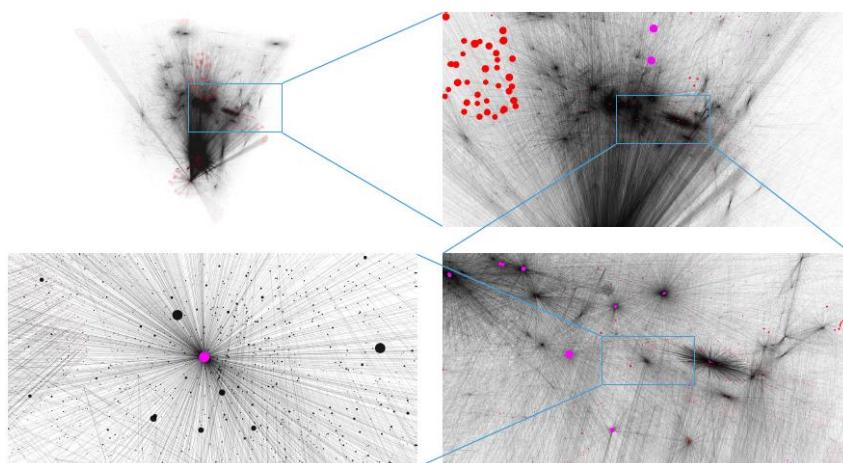
Nodos: **documentos WWW**

Enlaces: **hiperenlaces URL**

En torno a 1 billón ( $N \approx 10^{12}$ ) de documentos

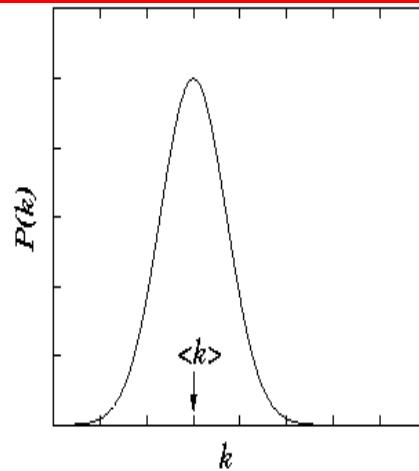
ROBOT: recopila todas las URLs encontradas en una página web y las sigue recursivamente

En 1998 se pensaba que era una red aleatoria:

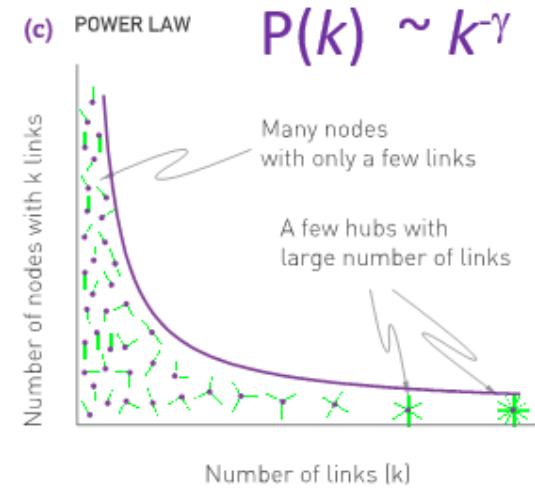


Mapa de la web (dominio nd.edu) con 325,725 páginas (1998)

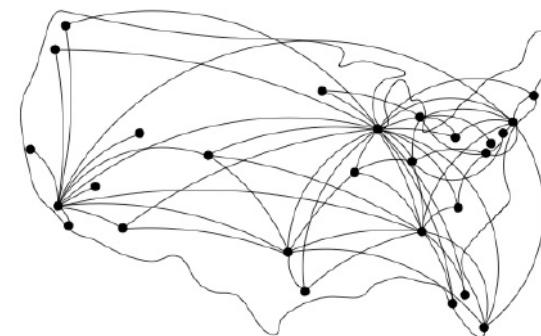
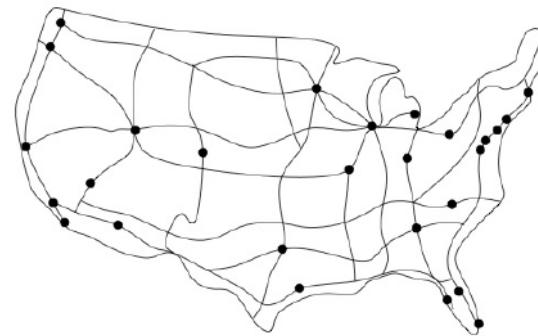
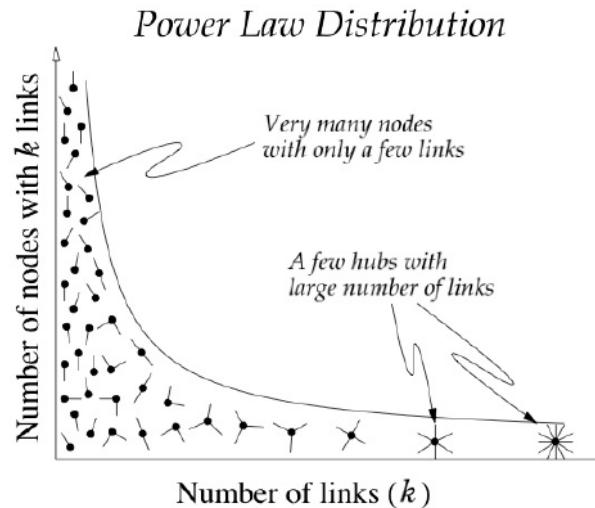
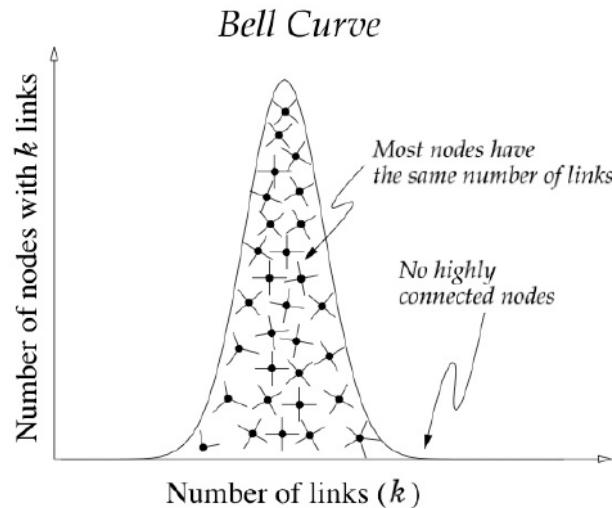
Nodos rojos → grado  $\geq 50$ ; nodos morados → grado  $\geq 500$



Esperada

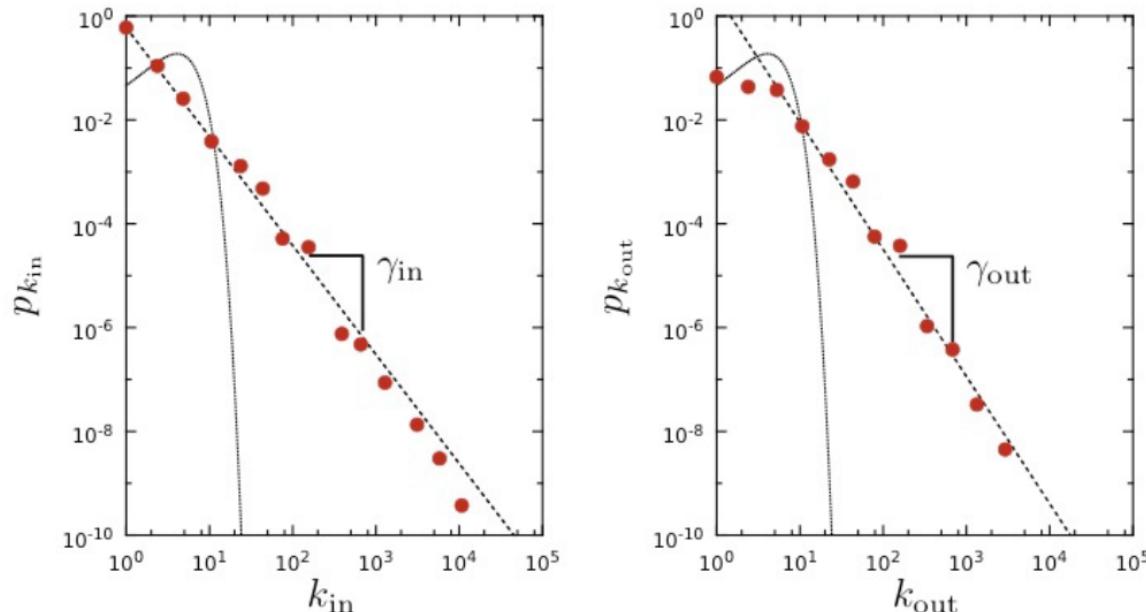


Observada



Once hubs are present, they change the way we navigate the network. For example, if we travel from Boston to Los Angeles by car, we must drive through many cities (nodes). On the airplane network, however, we can reach most destinations via a single hub, like Chicago. After [4].

[4] A.-L. Barabási. *Linked: The New Science of Networks*. (Plume, New York). ISBN 0-452-28439-2, 2002.



**Figure 4.2: The degree distribution of the WWW.**

The incoming (left panel) and outgoing (right panel) degree distribution of the WWW sample mapped in the 1999 study of Albert *et al.* [1]. The degree distribution is shown on double logarithmic axis (log-log plot), in which a power law is expected to follow a straight line. The symbols correspond to the empirical data and the dotted line corresponds to the power law fit, with degree exponents  $\gamma_{\text{in}} = 2.1$  and  $\gamma_{\text{out}} = 2.45$ . The degree distribution predicted by a Poisson function with average degree  $\langle k_{\text{in}} \rangle = \langle k_{\text{out}} \rangle = 4.60$ , representing the observed values for the WWW sample, is shown as a dotted line.

# REDES SOCIALES

# DEFINICION Y PARTICULARIDADES DE LAS REDES SOCIALES

Una **red social** es un **tipo concreto de red compleja** que modela las relaciones/interacciones existentes entre un conjunto de **entidades sociales**, tales como personas, grupos u organizaciones

Mathematics	Computer Science	Sociology
Vertex/vertices	Node	Actor/agent
Edge	Link/connection	Relational tie

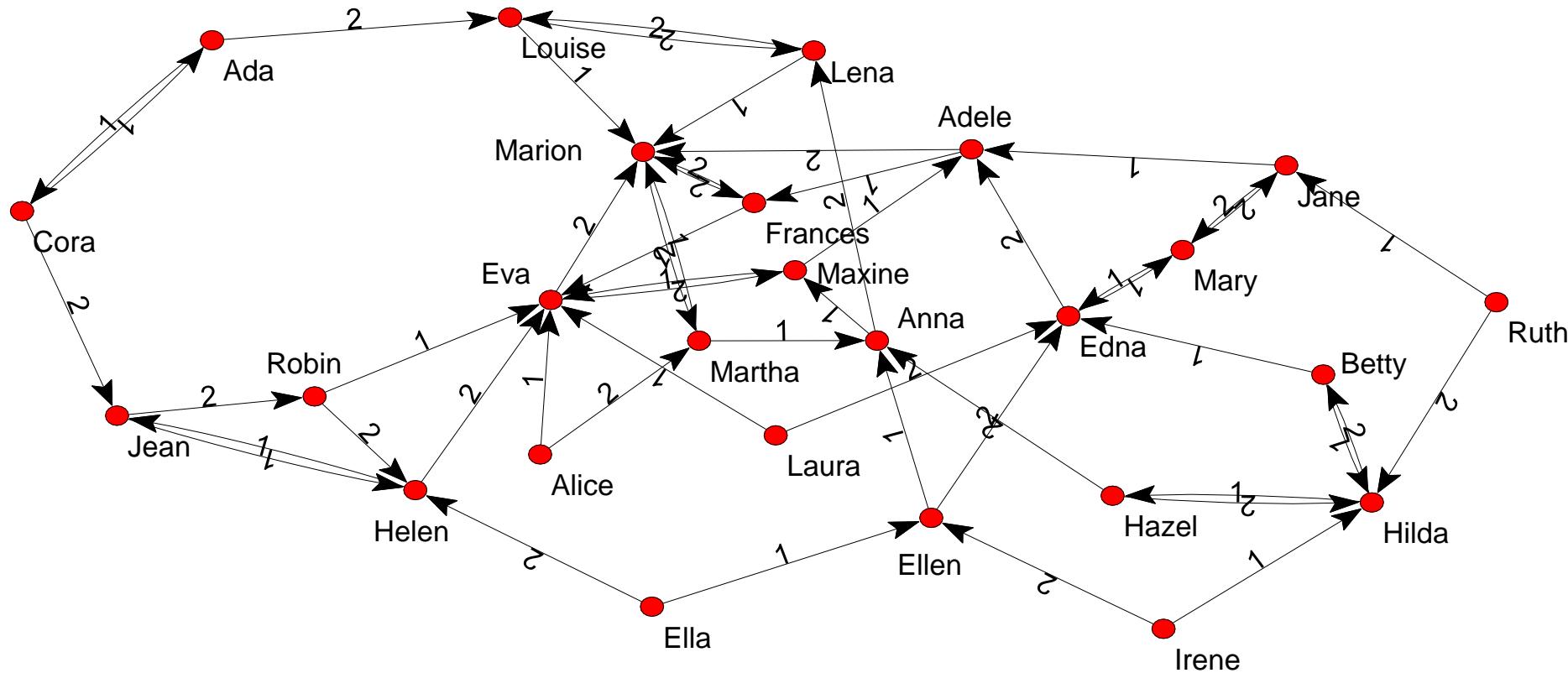
Según Newman, las redes reales se pueden categorizar en cuatro tipos principales:

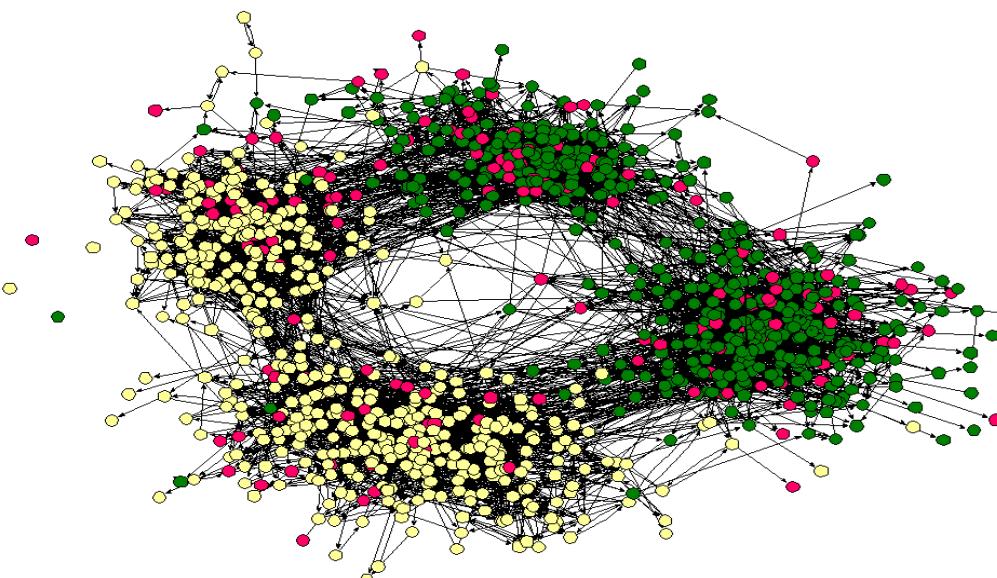
- redes sociales,
- redes de información o conocimiento,
- redes tecnológicas y
- redes biológicas

M.E.J Newman. *The structure and function of complex networks*. SIAM Reviews 2003, 45:167–228

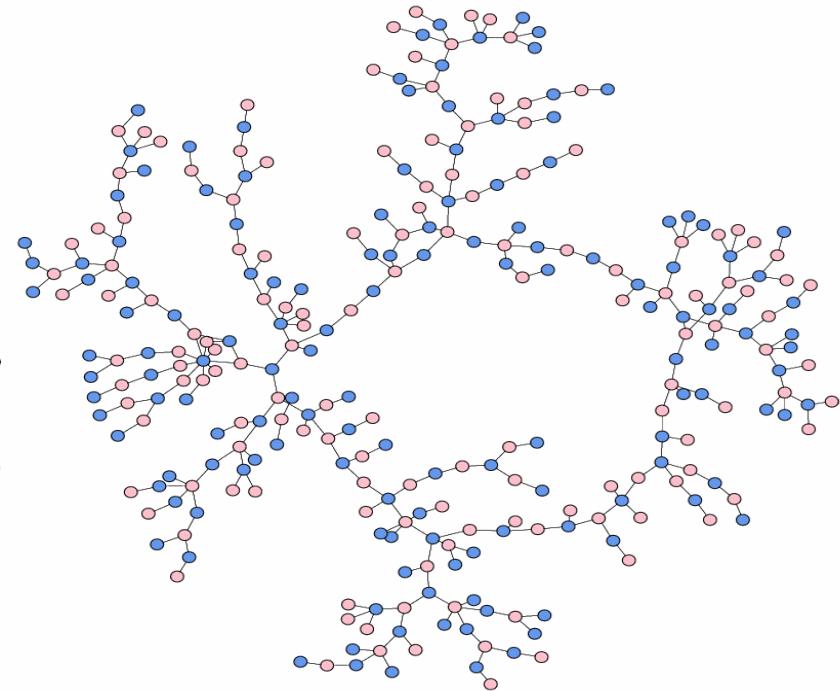
Las redes sociales son las que resultan de interacciones humanas y sociales, como las redes de amistad, las redes informales de comunicación entre empresas y las redes de colaboración (p.ej., la red de actores de Hollywood y las redes de coautoría científica), entre otras

1<sup>a</sup> y 2<sup>a</sup> elección de compañeras de mesa en una residencia femenina  
(Moreno, *The sociometry reader*, 1960)





**Relaciones de amistad en  
un instituto Americano**



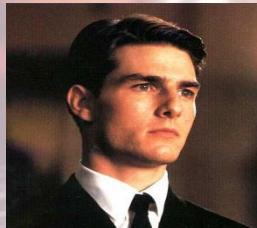
**Relaciones sexuales en un  
instituto Americano**

## EJEMPLOS (3):

## Red de Actores de Hollywood

Nodos: actores

Enlaces: actuaciones conjuntas



*Days of Thunder* (1990)  
*Far and Away* (1992)  
*Eyes Wide Shut* (1999)

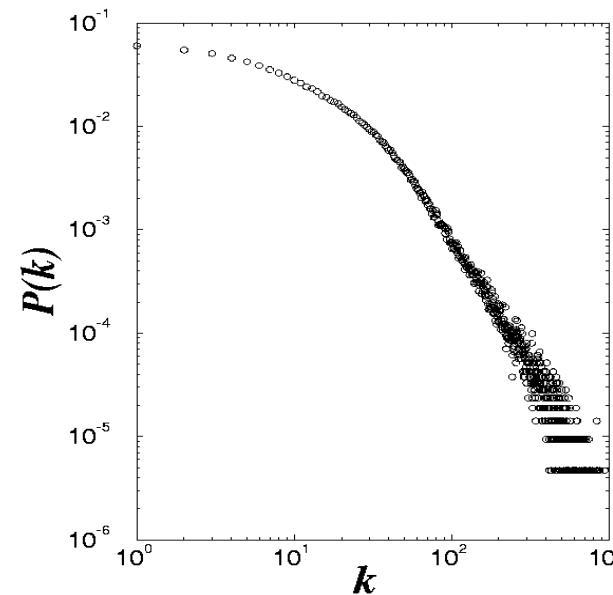


$N = 212,250$  actores

$\langle k \rangle = 28.78$

$P(k) \sim k^\gamma$

$\gamma = 2.3$



# EJEMPLOS DE REDES SOCIALES ON-LINE (1):

## Facebook

Busca personas, lugares y cosas

Oscar | Inicio | Chat (desactivado)

Oscar Cordón García  
Editar perfil

Últimas noticias

Mensajes

Eventos

Guardado

PÁGINAS

European Society f...

Descubrir páginas 20+

Noticias de páginas 20+

Crear un anuncio

Crear página

GRUPOS

PROMOCIÓN LICE... 1

Ex-Alumnos Institu...

evostar 7

Granada

ANTIGUOS ALUM...

ECSC

Reunión dia 25 de ...

Crear grupo

Ver grupos nuevos

Actualizar estado | Añadir fotos/video | Crear álbum de fotos

¿Qué estás pensando?

Juan Pedro Cordón Gutiérrez 14 de febrero a la(s) 20:58 ·

No soy azul pero no tendrás que besar a una rana . Pon un Juampe en tu vida !!! 😊

Ya no me gusta · Comentar · Compartir

Atí, Carmen Zarco Fernández, Caty Rguez Cordon, Eva McCormack y 3 personas más os gusta esto.

David Lanzarote Andrades jaja 14 de febrero a la(s) 21:23 · Me gusta · 1

Juana Gutiérrez García yo tengo uno\*\*\* 14 de febrero a la(s) 21:34 · Me gusta

Caty Rguez Cordon Y yo 14 de febrero a la(s) 22:00 · Me gusta

Escribe un comentario...

PUBLICACIONES

Pedro López Gallardo ha añadido 11 fotos nuevas. 2 h.

Large Bow Lamp - Shiny Chrome - £99.00

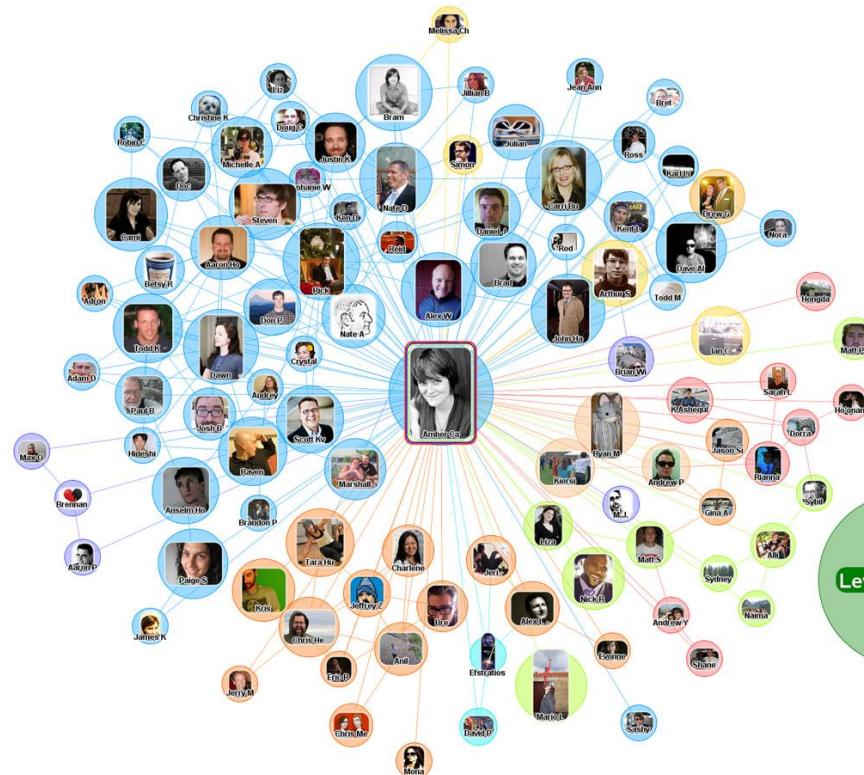
ANUNCIOS

Crea un anuncio

Large Bow Lamp - Shiny Chrome - £99.00

Chat (desactivado)

La **red social de amistad de Facebook** es un **grafo no dirigido**. Los nodos son usuarios y los enlaces indican que los dos usuarios son amigos. Es necesario tener el permiso de la otra persona para añadirla a tu red social.





El “Grafo Social” que hay detrás de Facebook

# EJEMPLOS DE REDES SOCIALES ON-LINE (4):

# Twitter (X)

Inicio

Explorar

Notificaciones

Mensajes

Grok

Listas

Guardados

Trabajos

Comunidades

Premium

Organizaciones v...

Perfil

Más opciones

Postear

Anuncios ni interrupciones en X.

Para ti

Siguiendo

Subirse

Ana @vida9ana · 24 feb.  
Si mi Comunidad tiene deuda 0, porque tengo que pagar como española la deuda de Cataluña?

802 703 5 mil 525 mil

DaSCI-Instituto Andaluz de Inteligencia Artificial @DaSCIes · 22h  
El próximo miércoles 5 de marzo tenemos nuevo #DaSCI Breakfast con el Dr. @geo\_jrc Jesús Rodrigo-Comino. ¿Sabía que la IA también se puede aplicar a la agricultura? Trataremos de buscar sinergias entre los proyectos de #EGEMAP y @DaSCIes  
Os esperamos

«DaSCI Breakfast»  
Buscando sinergias entre los proyectos desarrollados en EGEMAP y DaSCI  
Anteriormente, desde el laboratorio de investigación del IGR-DaSCI del grupo IRMAS-IRYI (www.egemap.es) estamos desarrollando tres proyectos (uno europeo - H2020) destinados a la optimización de la agricultura mediante la integración de datos satelitales, vuelos UAS, experimentación en campo, sensorización, nuevas técnicas de análisis y aprendizaje automático para la obtención de una mayor eficiencia en la agricultura. Los resultados obtenidos hasta la fecha han sido presentados en numerosas conferencias y publicados en revistas científicas internacionales. Actualmente, se está trabajando en la obtención de un gran volumen de datos. Esta presentación tendrá como objetivo presentar las principales conclusiones obtenidas y las perspectivas futuras de trabajo en el campo de la IA y las plataformas Geobebés. EGEMAP y DaSCI pueden colaborar estrechamente desde plena libertad.

Próximo Miércoles 5 de marzo  
Café a las 11h.  
Charla a las 11:30h.

3 74

Fijados por las personas a las que sigues

Real Madrid ... @realmad... · 3 feb.  
Comunicado Oficial.

EL PAÍS @el\_pais · 24 ene.  
TV en DIRECTO #ELPAÍS400V somos 400.000. Sigue en directo 'Historias de una redacción' el evento EL PAÍS para celebrar sus 400.000 suscriptores tinyurl.com/3pfss55

Patrycia Centeno @PoliticyModa · 25 feb.  
La relación entre Trump y Macron siempre ha sido curiosa gestualmente. El galo entendió ya en el primer mandato de Trump que un alfa tóxico sólo entiende el lenguaje corporal de otro alfa tóxico. Y en la visita de ayer lo tuvo en cuenta, pero perfeccionó (sófisticó) la técnica...  
Mostrar más



Oscar Cerdón García @oscarcordonogr ...

Buscar

Suscríbete a Premium

Suscríbete para desbloquear nuevas funciones y, si eres elegible, recibir un pago de cuota de ingresos.

Suscríbete

En directo en X

EL PAÍS es anfitrión Sesión de control al Gobierno en el Congreso d... +1,3 mil

A quién seguir

Javier Melero @fj.... Seguir tú también

Natalia Díaz Rodríguez @NataliaDiazRodr Seguir

ETSIT -UGR- @ETSITUGR Seguir

Mostrar más

Qué está pasando

Tendencia en España Hamás Madrid 2.696 publicaciones

Tendencia en España Alarma 5.090 publicaciones

Tendencia en España Mercedes Milá 1.136 publicaciones

Tendencia en España #CallejaEnElEspacio 8.763 publicaciones

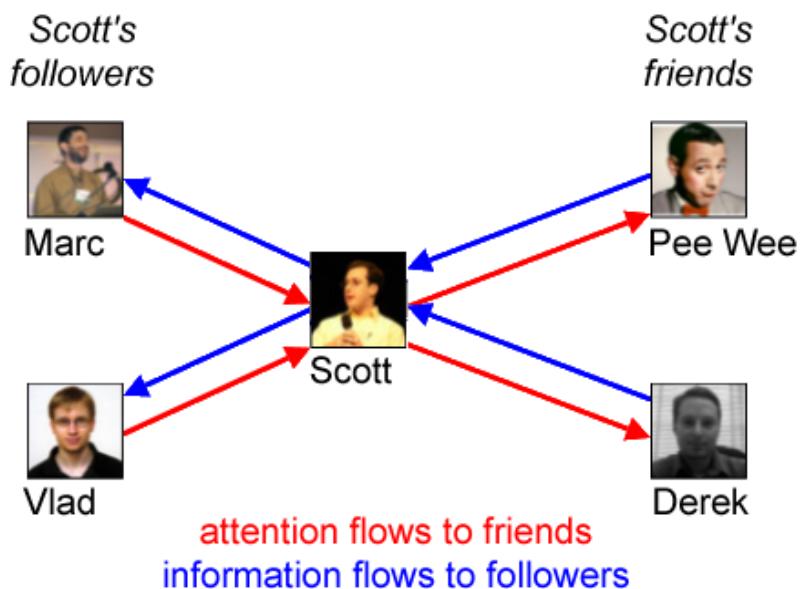
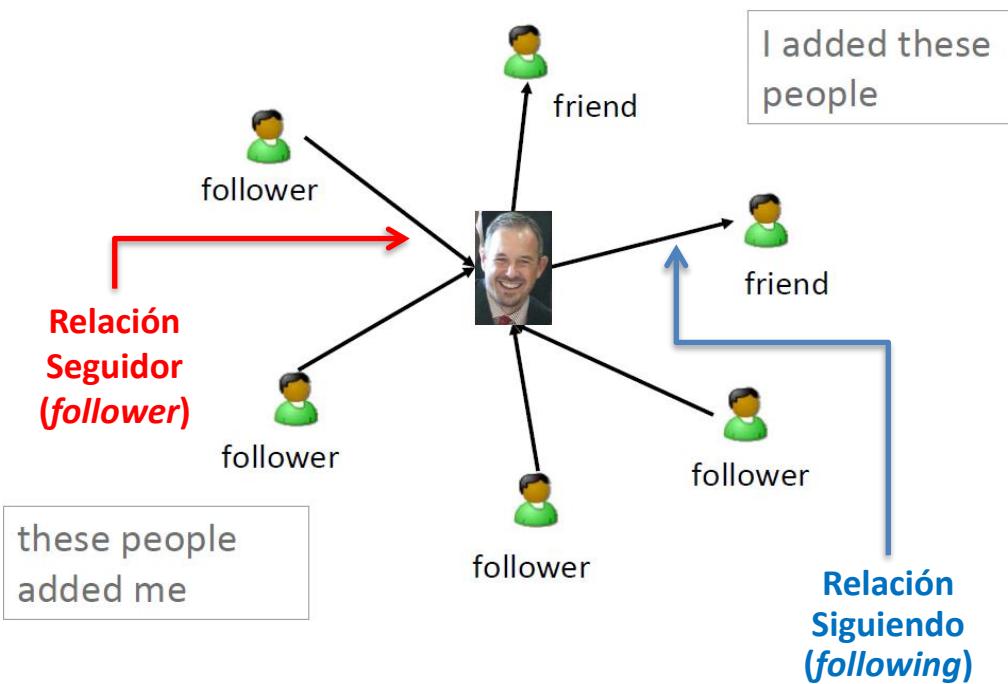
Mostrar más

Condiciones de Servicio Política de Privacidad  
Política de cookies Accesibilidad  
Información de anuncios Más opciones ...  
© 2025 X Corp.

## EJEMPLOS DE REDES SOCIALES ON-LINE (5):

## Twitter (X)

La **red social de amistad (friend-follower) de Twitter** es un **grafo dirigido**. Los nodos son usuarios y los enlaces (dirigidos) indican “quién sigue a quién”. Tu ves lo que publican tus amigos, tus seguidores ven lo que tu publicas

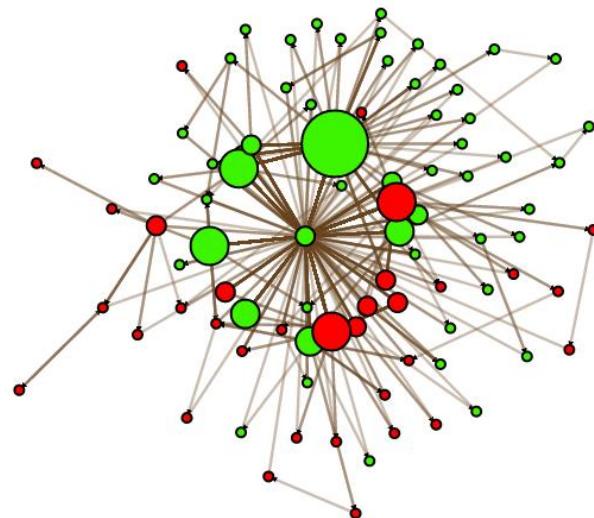


Ejemplo de red de amistad de Twitter

Existen otros tipos de redes generables a partir de datos de Twitter:

- **Redes de flujo de información** (*information flow*): Los nodos son usuarios y los enlaces (dirigidos) indican el camino seguido por el tweet (“quién ha retuiteado a quién”). La visualización de los *retweets* refleja el flujo de información en Twitter
- **Redes de co-ocurrencia de hashtags**: Los nodos son *hashtags* y los enlaces indican que dichos *hashtags* aparecen conjuntamente en el mismo tweet. Los enlaces están ponderados, indicando el número de co-ocurrencias

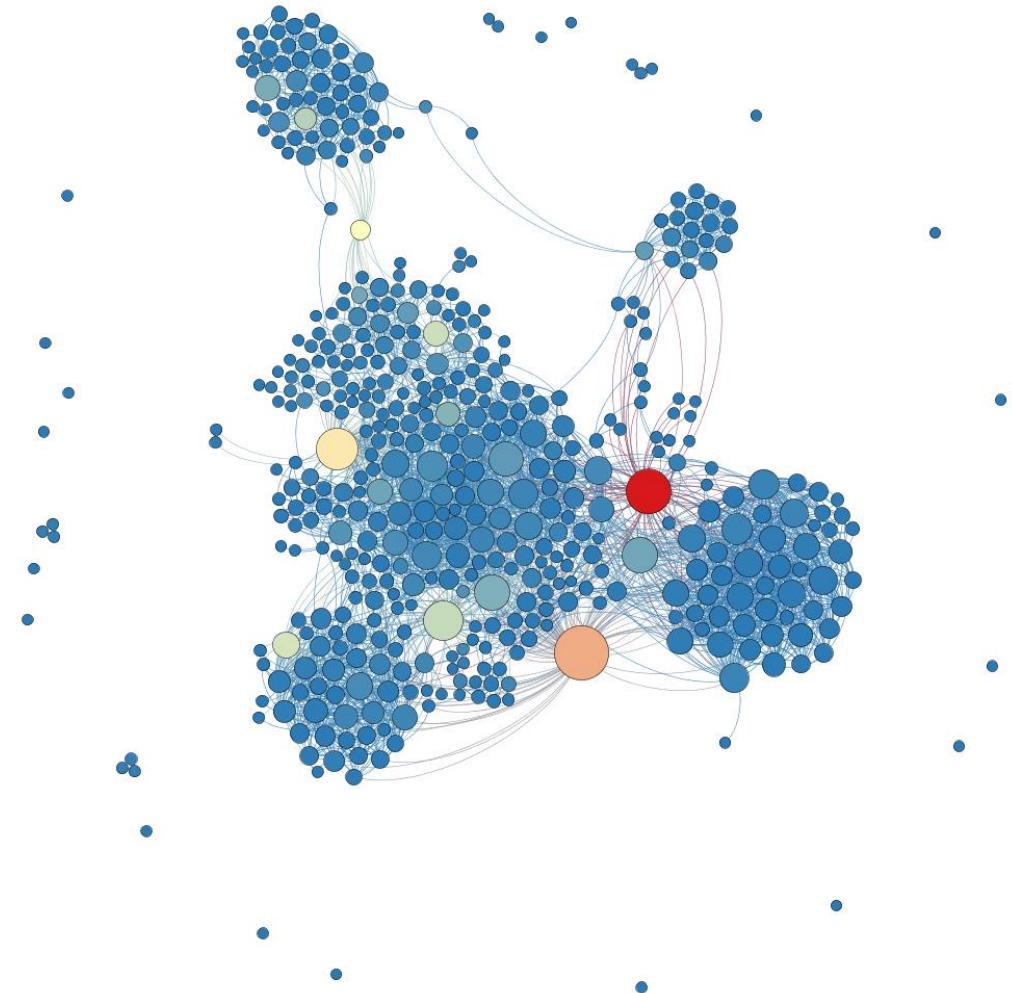
Red de *retweets* de los tópicos “#zuccotti” y “#nypd”



# ANÁLISIS BÁSICO DE LA RED DE FACEBOOK DE OSCAR CORDÓN (1)

**Red social de amistad en Facebook de Oscar Cordón 2014**

**¡Incluso siendo una red pequeña, es demasiado compleja para analizarla visualmente o con la matriz de adyacencia!**

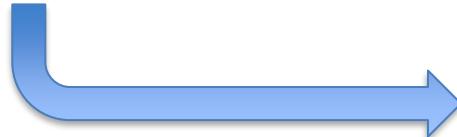


# ANÁLISIS BÁSICO DE LA RED DE FACEBOOK DE OSCAR CORDÓN (2)

**Valores de las medidas:**  $N=483$  nodos,  $L=3323$  enlaces,  $L_{max} = \frac{483 \cdot 482}{2} = 116403$  enlaces,  $D = \frac{3323}{116403} = 0.029$  (**2.9% de los enlaces posibles, valor muy bajo, red muy poco densa, muy dispersa**)

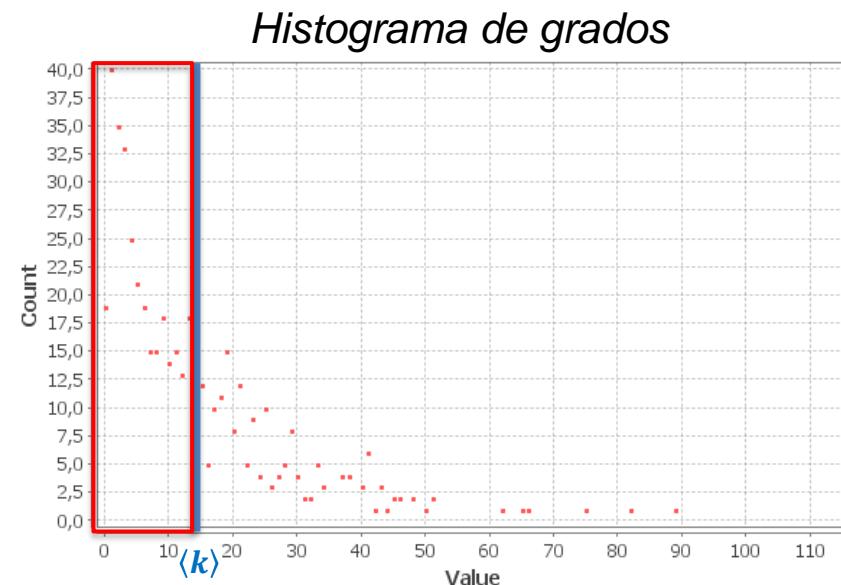
**Grado medio:**  $\langle k \rangle = \frac{2L}{N} = 13.76 \rightarrow$  cada nodo de la red tiene casi 14 amigos en media (**valor muy bajo, cada nodo podría tener hasta 482 amigos**)

La distribución indica que un 62.11% de los nodos tienen menos de 14 amigos, grado medio ( $k_i < \langle k \rangle$ )



Así, **tenemos muchos más nodos de grado bajo que de grado alto**. Hay 19 nodos de grado 0, 40 de grado 1, 35 de grado 2, 33 de grado 3, etc. En cambio, solo hay 1 nodo de grado 115, 1 nodo de grado 89, etc.

Se da la **propiedad libre de escala (scale-free)**, muy común en redes reales



# ANÁLISIS BÁSICO DE LA RED DE FACEBOOK DE OSCAR CORDÓN (3)

Existen unos pocos nodos fuertemente conectados (**hubs**), el mayor con grado 115

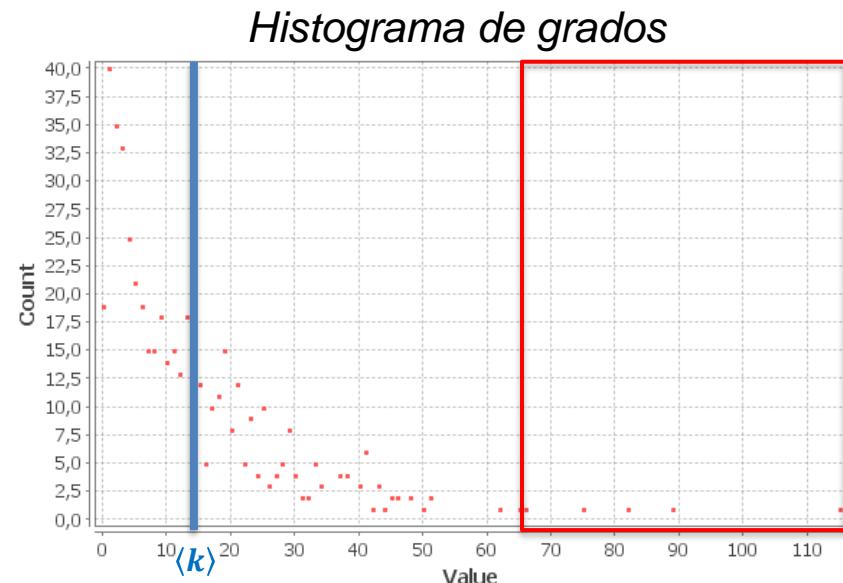
Este grado es más de 8 veces mayor que la media (13.76), mucho para una red tan pequeña

Sólo el 1% de los nodos (5 nodos) tienen 66 amigos o más:

- grado 66 (casi 5 veces mayor que la media),
- grado 75 (5 veces mayor que la media),
- grado 82 (casi 6 veces mayor que la media),
- grado 89 (6 veces mayor que la media), y
- grado 115 (más de 8 veces mayor que la media)

El sistema social no está equilibrado y hay unos pocos actores con muchas más conexiones que otros

**Red libre de escala:** Muchos más nodos por debajo de la media que por encima. Pocos nodos con un grado significativamente mayor que la media (hubs). **Sistema desequilibrado hacia la izquierda**



# ANÁLISIS BÁSICO DE LA RED DE FACEBOOK DE OSCAR CORDÓN (4)

Al haber  $N=483$  nodos, la distancia entre dos nodos podría ser de hasta 482 ( $N-1$  enlaces)

Viendo la red, pensaríamos que hay variaciones grandes en las distancias entre nodos, pero la distribución tiene una **distancia media** ( $\langle d \rangle = 3.35$ ) y una desviación típica **muy bajas** (muchos nodos con distancia cercana a la media)

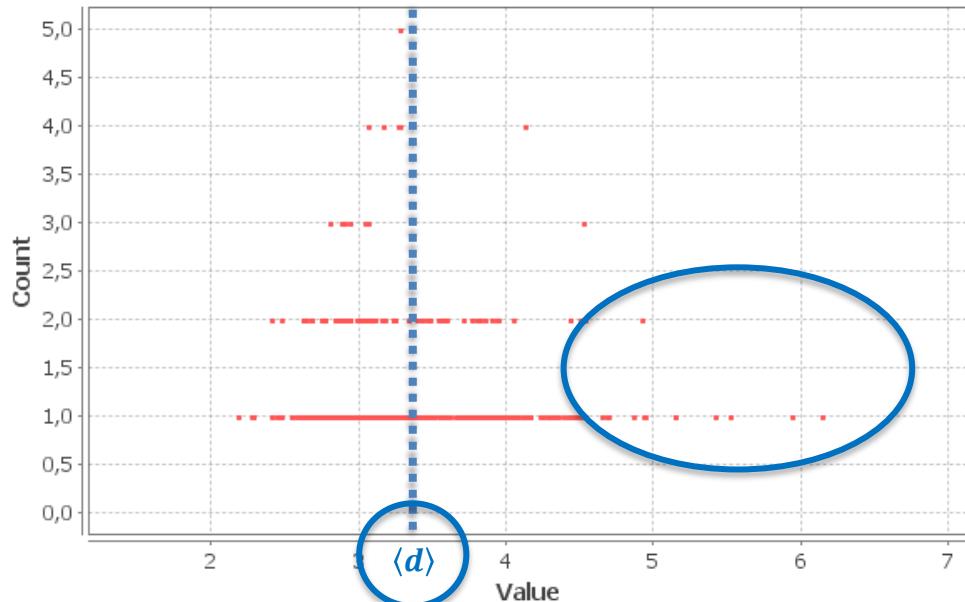
El **diámetro** es  $d_{max}=9$ , es decir, con **sólo 9 saltos** en el peor de los casos llegamos a **todos** los nodos de la red. También es **muy bajo**

La frecuencia cae fuertemente para las distancias grandes, indicando que no existen distancias de este tipo en la red



Consecuencia de la **propiedad de mundos pequeños (small-world)**, también común en redes reales

Histograma de distancias



# ANÁLISIS BÁSICO DE LA RED DE FACEBOOK DE OSCAR CORDÓN (5)

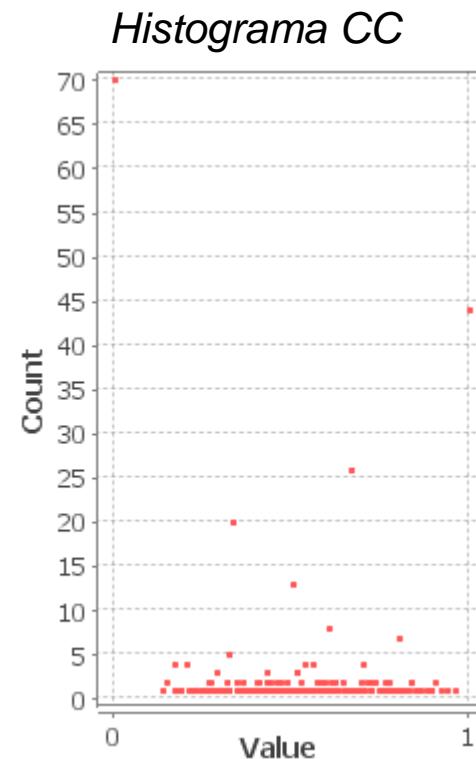
**Conectividad:** la red presenta 24 componentes conexas, muchas de ellas (19) formadas por una sola persona aislada que no tiene relación con el resto

Vemos una componente gigante que agrupa el 93.79% de los nodos: 453 de los 483. Hay también algunas componentes aisladas muy pequeñas (1 de 2 nodos y 3 de 3 nodos). Esta **fragmentación** también es común en redes reales

**Coeficiente de clustering medio:**  $\langle C \rangle = 0.558$ . **Es un valor medio-alto** (habitual en redes sociales), indicando un grado significativo de clustering local (**el 56% de los amigos de un nodo son amigos entre sí**)



La distribución del coeficiente de clustering está **muy desequilibrada**. Hay muchos nodos con valor 0 (70) y con valor 1 (44). También hay bastantes con valores en torno a 0.4 (20) y en torno a 0.6 (26). Para el resto de valores hay muy pocos nodos



# ANÁLISIS BÁSICO DE LA RED DE FACEBOOK DE OSCAR CORDÓN (6)

Comparación de la densidad y el coeficiente de clustering medio:  $D=0.029$  y  $\langle C \rangle=0.558$

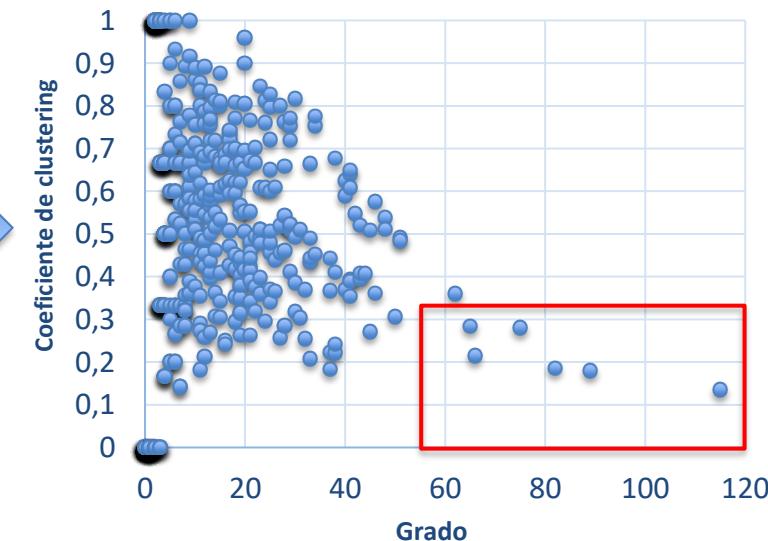
La densidad es mucho más baja, indicando que hay pocos enlaces en la red

En cambio, el coeficiente de clustering es muchísimo más alto, casi 20 veces más (2.9% contra 55%)

Esto indica que, aunque haya pocos enlaces, están muy agrupados localmente, en zonas concretas.

**Este sistema social muestra una tendencia a formar grupos cohesivos (comunidades)**

El coeficiente de clustering es mucho mayor en los nodos poco conectados que en los *hubs* → los nodos de grado bajo se sitúan en vecindarios localmente densos y viceversa. Consecuencia de la **jerarquía de redes**



# ANÁLISIS DE REDES SOCIALES ON-LINE (1)

**Nodos:** usuarios

**Enlaces:** relaciones

Estudio de varias redes sociales online, con un muestro de un tamaño significativo, que concluye que todas ellas tienen propiedades comunes

**Mislove et al., Measurement and Analysis of Online Social Networks, Proc. IMC'07, pp. 29-42 (2007)**

	Flickr	LiveJournal	Orkut	YouTube
Number of users	1,846,198	5,284,457	3,072,441	1,157,827
Estimated fraction of user population crawled	26.9%	95.4%	11.3%	unknown
Dates of crawl	Jan 9, 2007	Dec 9 - 11, 2006	Oct 3 - Nov 11, 2006	Jan 15, 2007
Number of friend links	22,613,981	77,402,652	223,534,301	4,945,382
Average number of friends per user	12.24	16.97	106.1	4.29
Fraction of links symmetric	62.0%	73.5%	100.0%	79.1%
Number of user groups	103,648	7,489,073	8,730,859	30,087
Average number of groups memberships per user	4.62	21.25	106.44	0.25

Table 1: High-level statistics of our social networking site crawls.

# ANÁLISIS DE REDES SOCIALES ON-LINE (2)

**Nodos:** usuarios

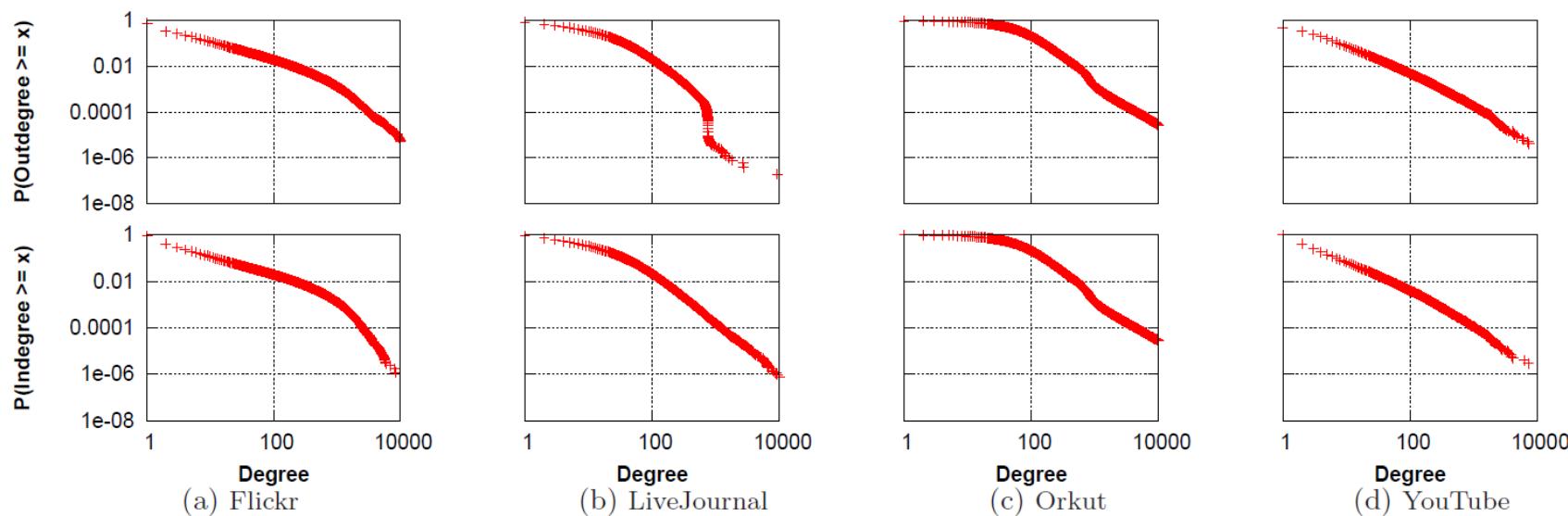
**Enlaces:** relaciones

Todas las distribuciones muestran un comportamiento de **larga cola (redes libres de escala)**: hay desviaciones de varios órdenes de magnitud en los grados

Mislove et al., Measurement and Analysis of Online Social Networks, Proc. IMC'07, pp. 29-42 (2007)

Network	Outdegree		Indegree	
	$\gamma$	D	$\gamma$	D
Web [12]	2.67	-	2.09	-
Flickr	1.74	0.0575	1.78	0.0278
LiveJournal	1.59	0.0783	1.65	0.1037
Orkut	1.50	0.6319	1.50	0.6203
YouTube	1.63	0.1314	1.99	0.0094

Table 2: Power-law coefficient estimates ( $\gamma$ ) and corresponding Kolmogorov-Smirnov goodness-of-fit metrics (D). The Flickr, LiveJournal, and YouTube networks are well approximated by a power-law.



# ANÁLISIS DE REDES SOCIALES ON-LINE (3)

Network	Avg. Path Len.	Radius	Diameter
Web [12]	16.12	475	905
Flickr	5.67	13	27
LiveJournal	5.88	12	20
Orkut	4.25	6	9
YouTube	5.10	13	21

Table 3: Average path length, radius, and diameter of the studied networks. The path length between random nodes is very short in social networks.

the path lengths and diameters for all four social networks are remarkably short. Interestingly, despite being comparable in size to the Web graph we considered, the social networks have significantly shorter average path lengths and diameters. This property may again result from the high degree of reciprocity within the social networks. Incidentally, Broder et al. [12] noted that if the Web were treated as an undirected graph, the average path length would drop from 16.12 to 7.

Network	$C$	Ratio to Random Graphs	
		Erdös-Rényi	Power-Law
Web [2]	0.081	7.71	-
Flickr	0.313	47,200	25.2
LiveJournal	0.330	119,000	17.8
Orkut	0.171	7,240	5.27
YouTube	0.136	36,900	69.4

Table 4: The observed clustering coefficient, and ratio to random Erdös-Rényi graphs as well as random power-law graphs.

The clustering coefficients of social networks are between three and five orders of magnitude larger than their corresponding random graphs, and about one order of magnitude larger than random power-law graphs. This unusually high clustering coefficient suggests the presence of strong local clustering, and has a natural explanation in social networks: people tend to be introduced to other people via mutual friends, increasing the probability that two friends of a single user are also friends.

# ANÁLISIS DE REDES SOCIALES ON-LINE (4)

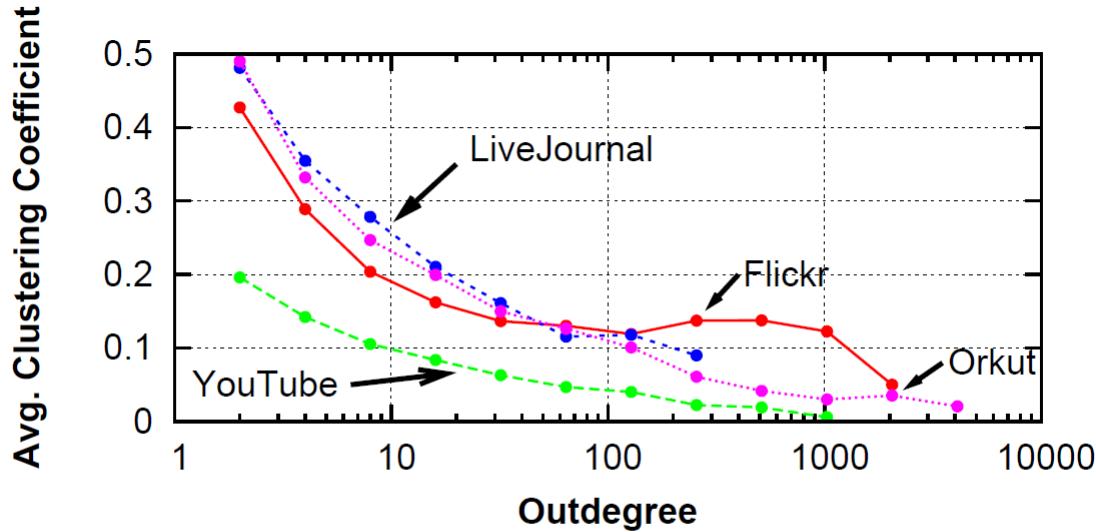


Figure 9: Clustering coefficient of users with different outdegrees. The users with few “friends” are tightly clustered.

Figure 9 shows how the clustering coefficients of nodes vary with node outdegree. The clustering coefficient is higher for nodes of low degree, suggesting that there is significant clustering among low-degree nodes. This clustering and the small diameter of these networks qualifies these graphs as small-world networks [52], and further indicates that the graph has scale-free properties.

# ANÁLISIS DE REDES SOCIALES ON-LINE (5)

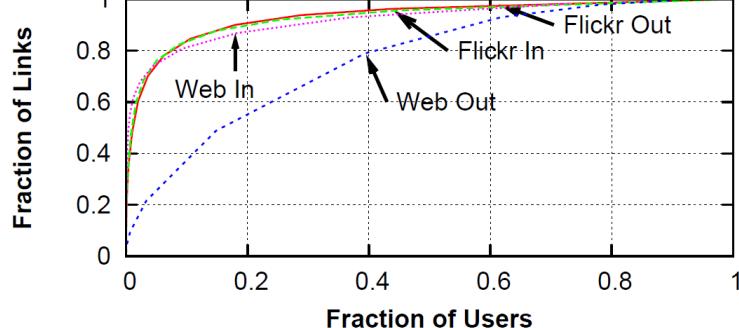


Figure 3: Plot of the distribution of links across nodes. Social networks show similar distributions for outgoing and incoming links, whereas the Web links shows different distributions.

In social networks, the nodes with very high outdegree also tend to have very high indegree. In our study, for each network, the top 1% of nodes ranked by indegree has a more than 65% overlap with the top 1% of nodes ranked by outdegree. The corresponding overlap in the Web is less than 20%. Hence, active users (i.e., those who create many links) in social networks also tend to be popular (i.e., they are the target of many links). Figure 4 shows the extent of the overlap between the top  $x\%$  of nodes ranked by indegree and outdegree.

The high correlation between indegree and outdegree in social networks can be explained by the high number of symmetric links.

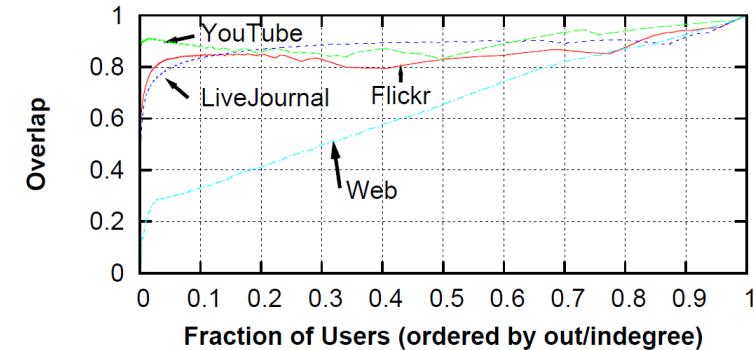


Figure 4: Plot of the overlap between top  $x\%$  of nodes ranked by outdegree and indegree. The high-indegree and high-outdegree nodes are often the same in social networks, but not in the Web.

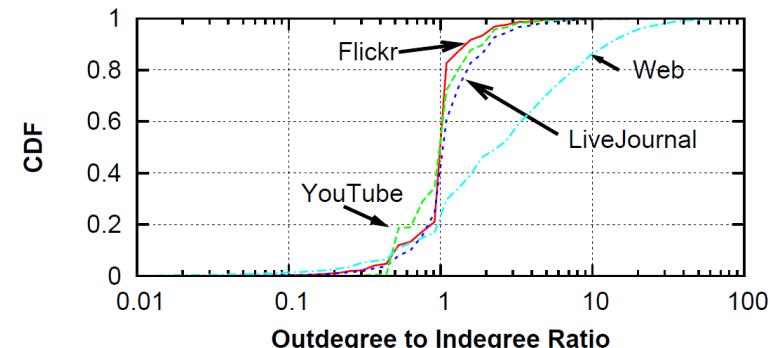


Figure 5: CDF of outdegree to indegree ratio. Social networks show much stronger correlation between indegree and outdegree than the Web.

# Referencias y Agradecimientos

Para elaborar las transparencias de este curso, he hecho uso de algunos materiales desarrollados por expertos en el área disponible en Internet:

- “Network Science Interactive Book Project” del Laszlo Barabasi Lab.  
Northeastern University: <http://barabasilab.com/networksciencebook>
- K. Lerman. “Social Media. A Responsible User’s Guide”. University of South California: <http://www.isi.edu/integration/people/lerman/talks.html>
- Curso on-line “Social Network Analysis” de Lada Adamic, Coursera, Universidad de Michigan: <https://www.coursera.org/course/sna>
- L. Cherny. “Simplifying Social Network Visualizations” . Ghostweather Research & Design, LLC: <http://es.slideshare.net/arnicas/simplifying-social-network-diagrams?related=1>

