

# Minería de Medios Sociales

Máster en Ciencia de Datos e Ingeniería de Computadores

Bloque I: Redes Sociales y Minería de Datos en Redes



# **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 (**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

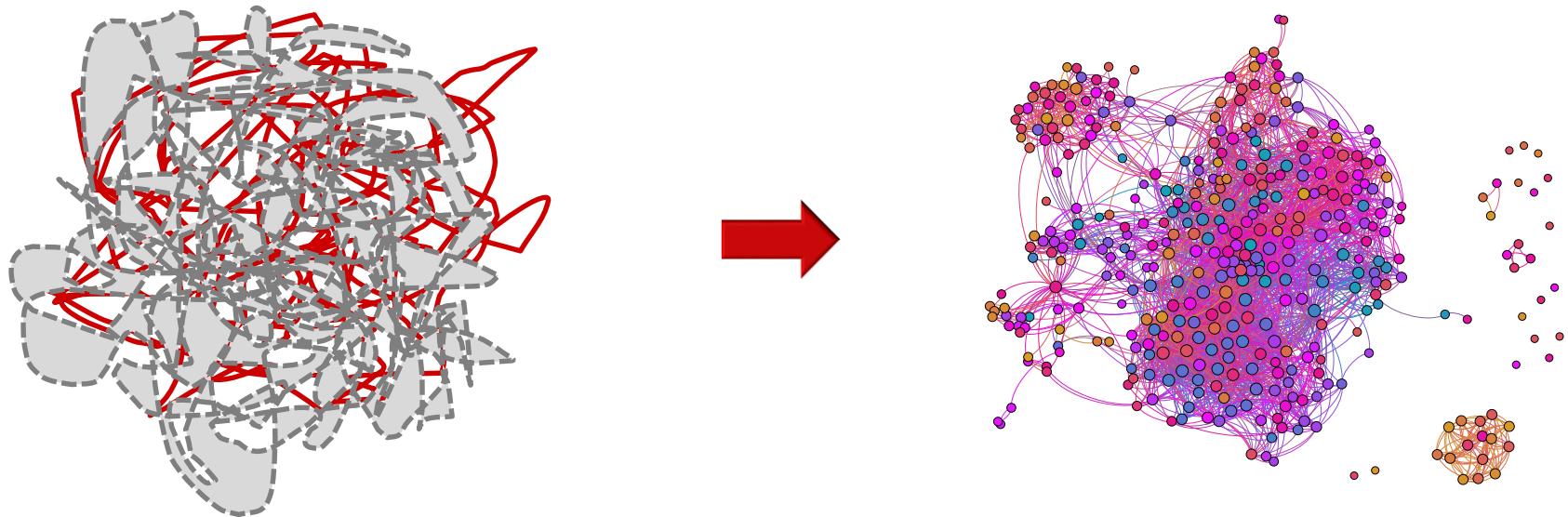




*“I think the next century  
will be the century  
of complexity.”*

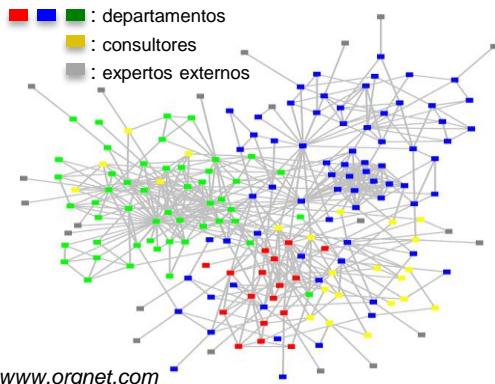
Stephen Hawking  
January 23, 2000

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

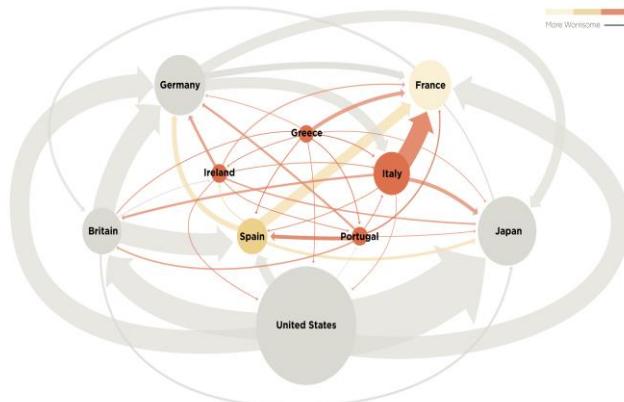


# EJEMPLOS DE REDES COMPLEJAS REALES

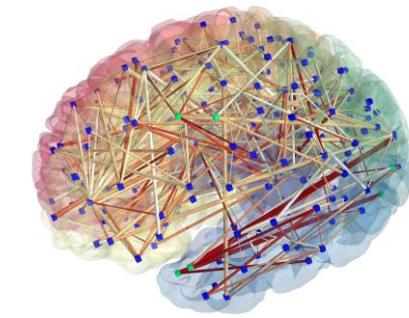
## Organizaciones



## Redes financieras

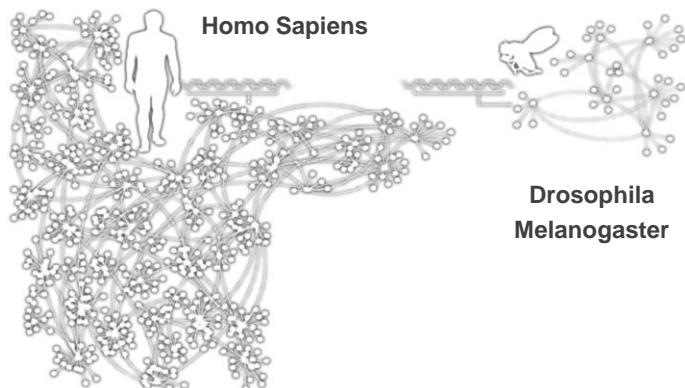


## Redes cerebrales

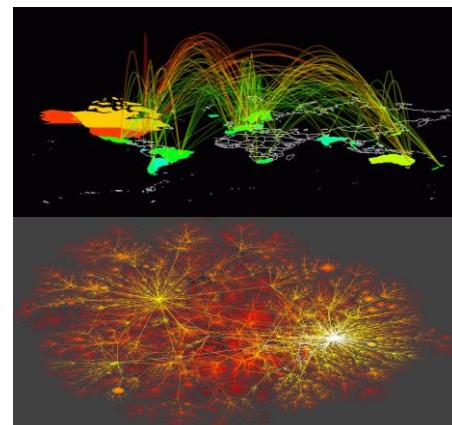


<http://www.humanconnectomeproject.org/>

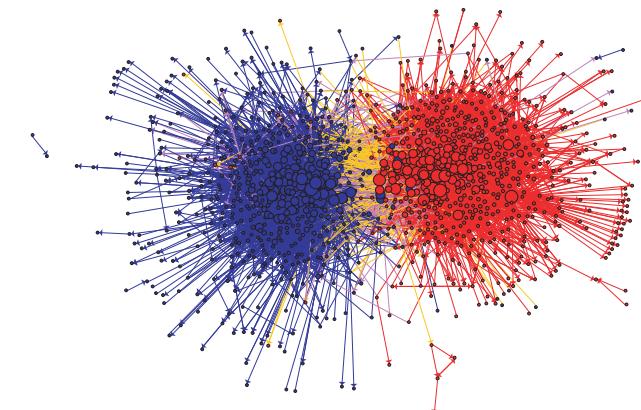
<http://www-personal.umich.edu/~mejn/netdata/>



Redes genéticas



Red física de Internet



Redes políticas

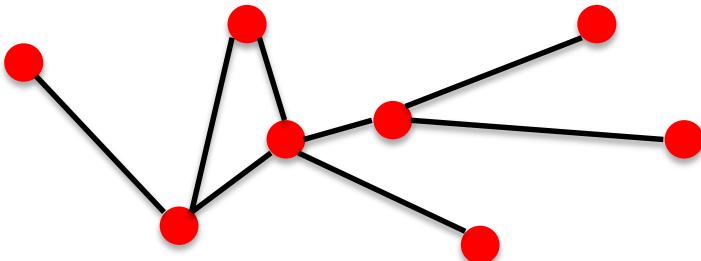
## EL ROL DE LAS REDES

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

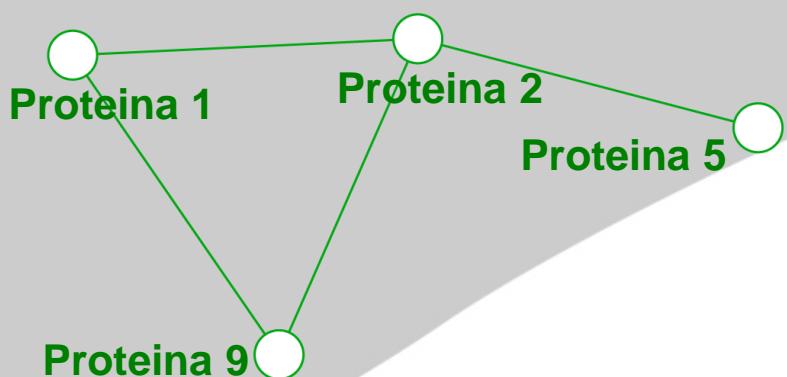
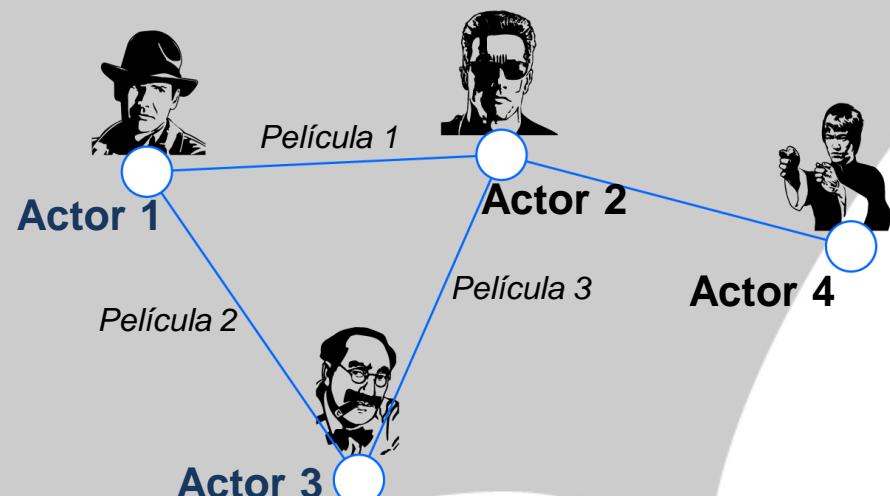
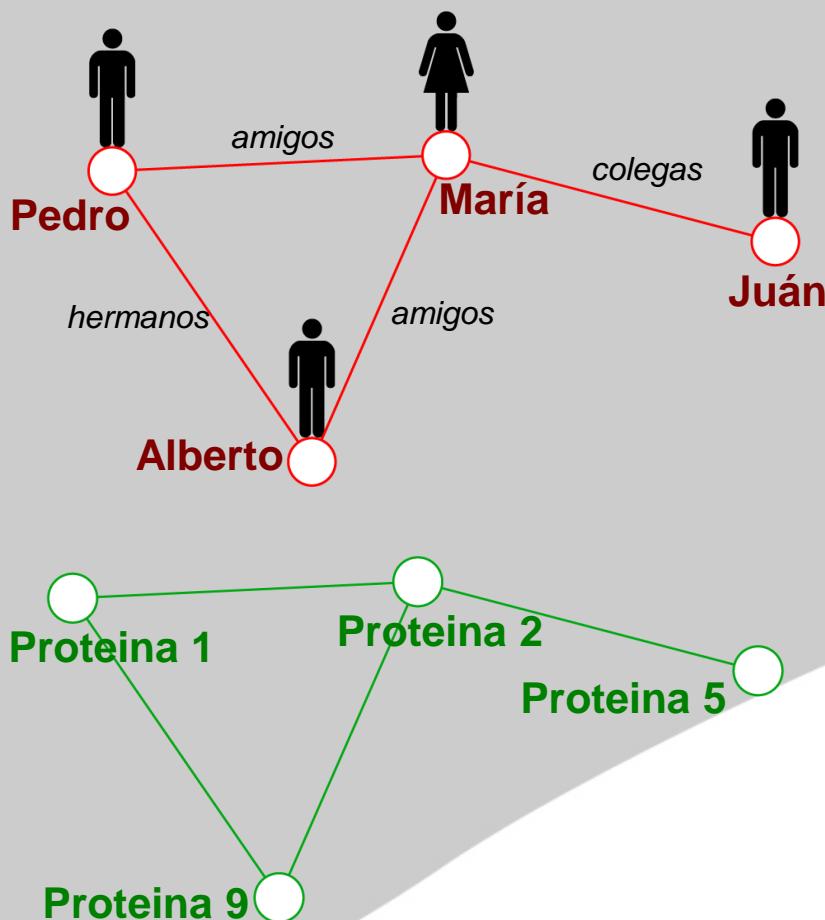
Nomenclatura: (Red, nodo, enlace)

**grafo:** representación matemática de una red

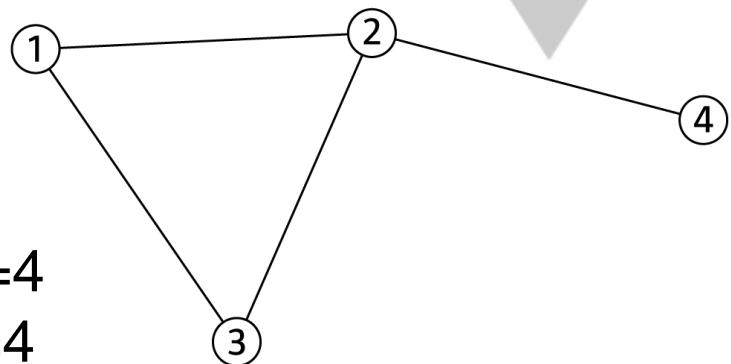
- grafo de la web,
- grafo social (término de Facebook)

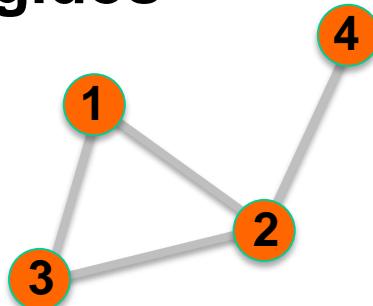
Nomenclatura: (Grafo, vértice, arco)

# LAS REDES OFRECEN UN LENGUAJE COMÚN PARA ESTUDIAR SISTEMAS



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



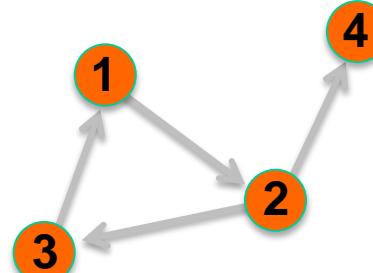
**No dirigidos**

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

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

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

*Red de actores de Hollywood*  
*Redes de interacciones entre proteínas*

**Dirigidos**

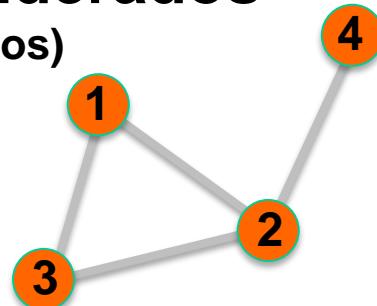
$$A_{ij} = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

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

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

*WWW*  
*Redes de citas en artículos científicos*

## No ponderados (no dirigidos)



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

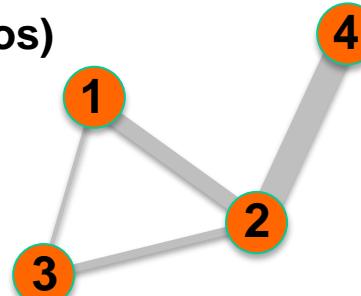
$$A_{ii} = 0$$

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

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

Redes de interacciones entre proteínas  
WWW

## Ponderados (no dirigidos)



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

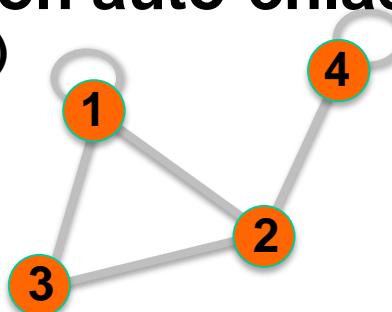
$$A_{ii} = 0$$

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

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

Redes de llamadas  
Redes metabólicas

## Grafos con auto-enlaces (no dirigidos)



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

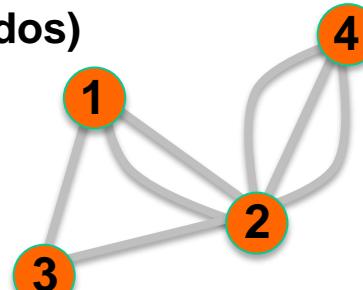
$$A_{ii} \neq 0$$

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

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

Redes de interacciones entre proteínas  
WWW

## Multigrafos (no dirigidos)



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

$$A_{ii} = 0$$

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

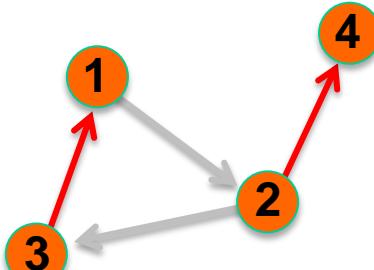
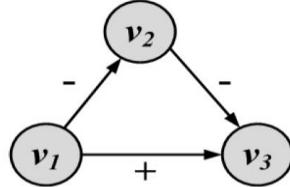
$$L = \frac{1}{2} \sum_{i,j=1}^N \text{nonzero}(A_{ij}) \quad \langle k \rangle = \frac{2L}{N}$$

Redes 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)

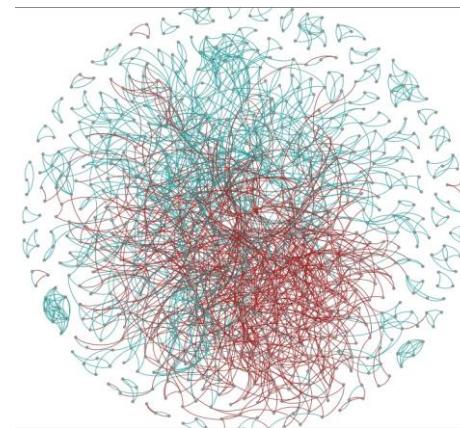


$$A_{ij} = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & -1 \\ -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

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

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

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



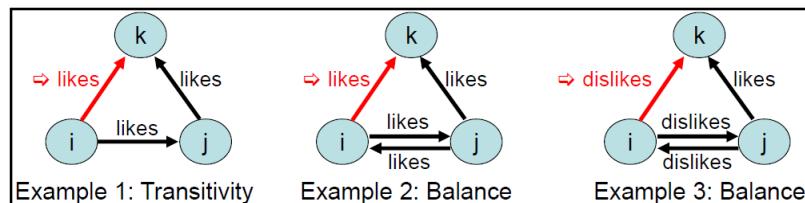
Muestreo de las opiniones positivas y negativas de la red Epinions

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

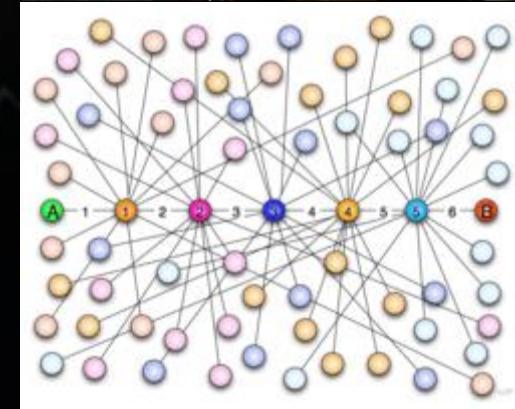
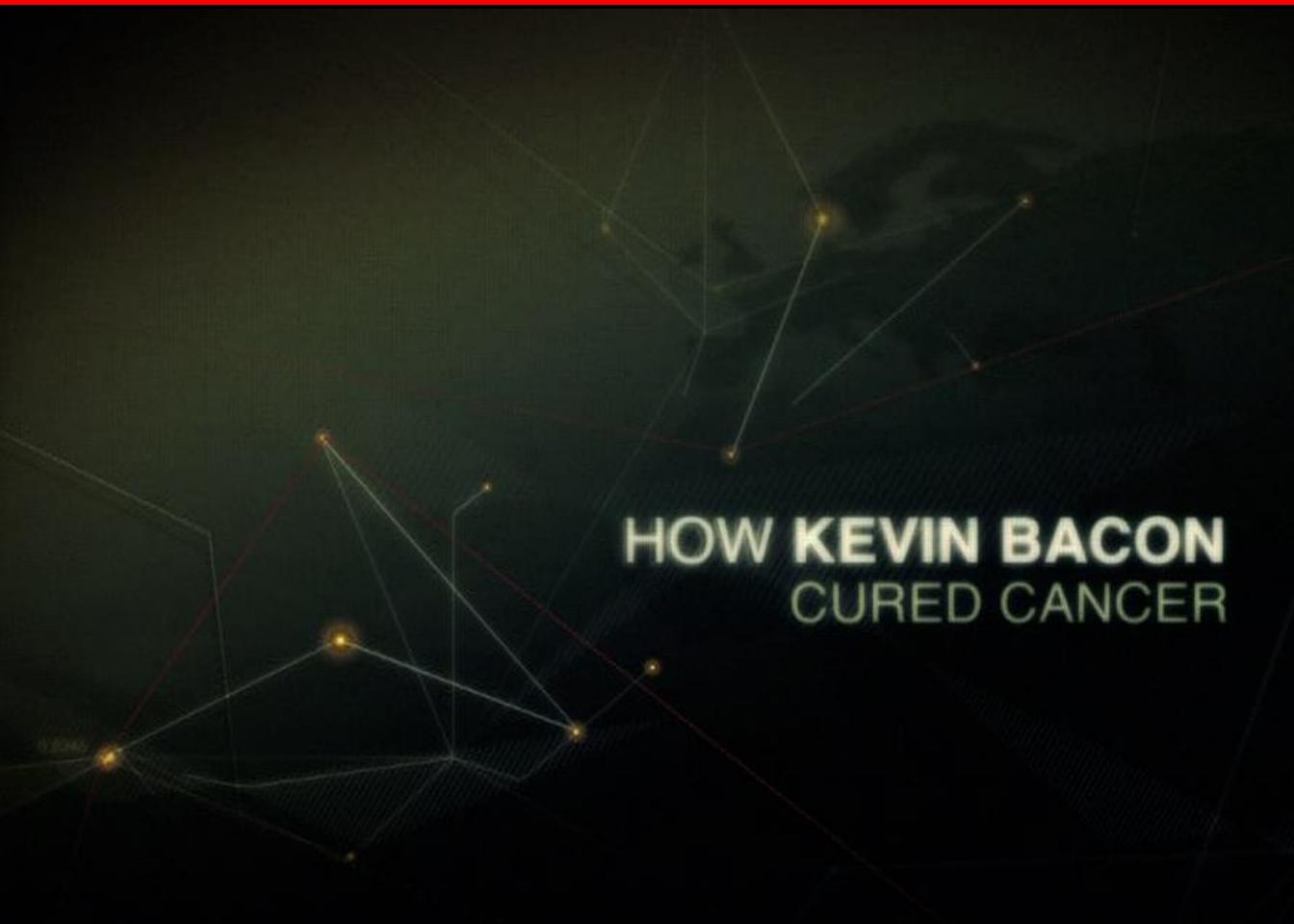
<http://snap.stanford.edu/data/soc-sign-epinions.html>

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

<http://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  $\langle k \rangle \ll N-1$

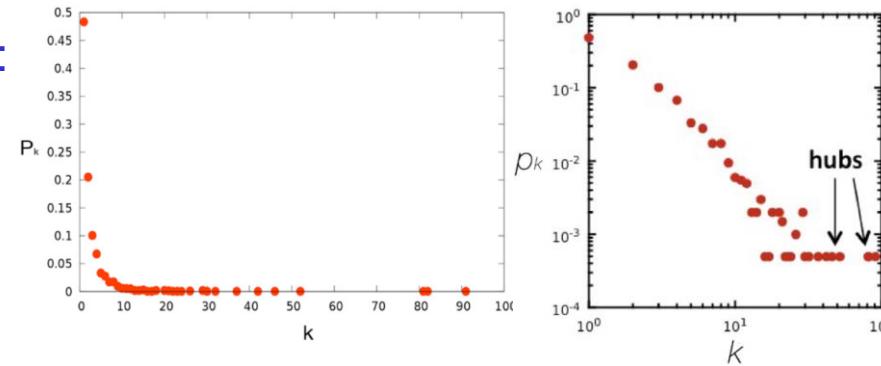
NETWORK NAME	NODES	LINKS	DIRECTED/ UNDIRECTED	N	L	$\langle k \rangle$
Internet	routers	Internet Connections	Undirected	192,244	609,066	2.67
WWW	webpages	links	Directed	325,729	1,497,134	4.60
Power Grid	power plants, transformers	cables	Undirected	4,941	6,594	2.67
Mobile-Phone Calls	subscribers	calls	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 7,000 millones ( $7 \cdot 10^9$ )

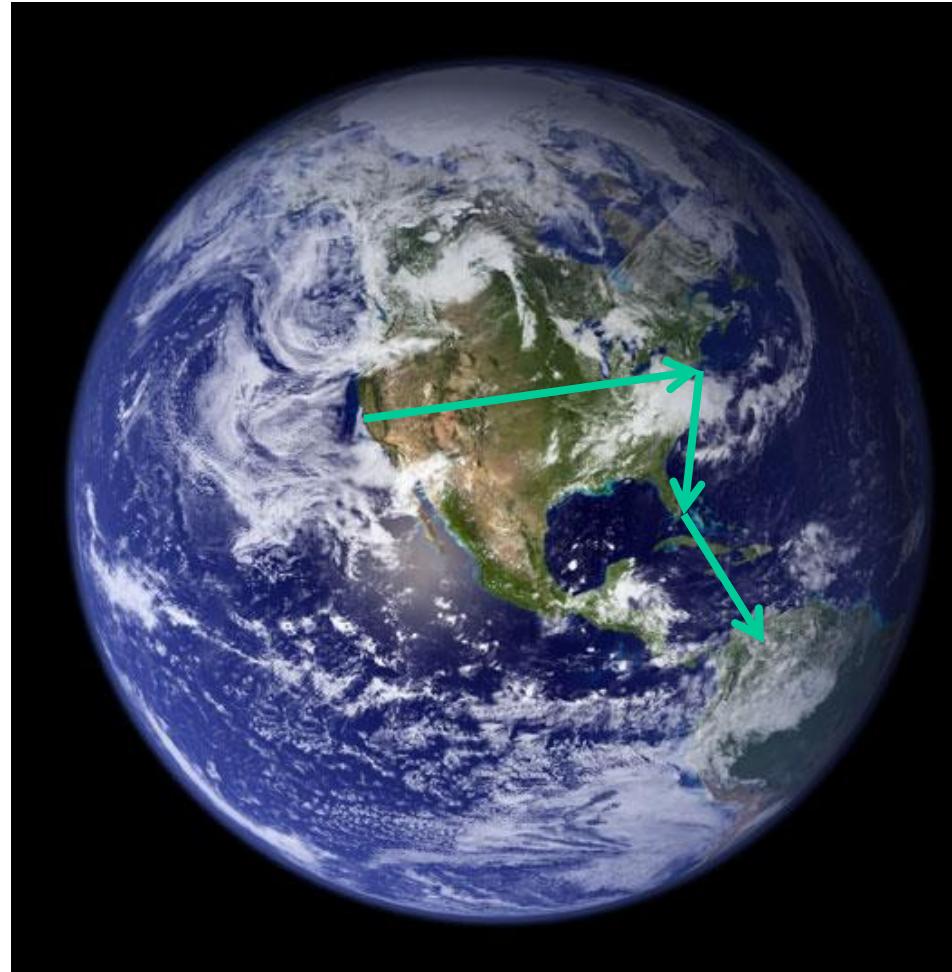
$$\langle d \rangle = \frac{\ln 7 \times 10^9}{\ln(10^3)} = 3.28$$

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)

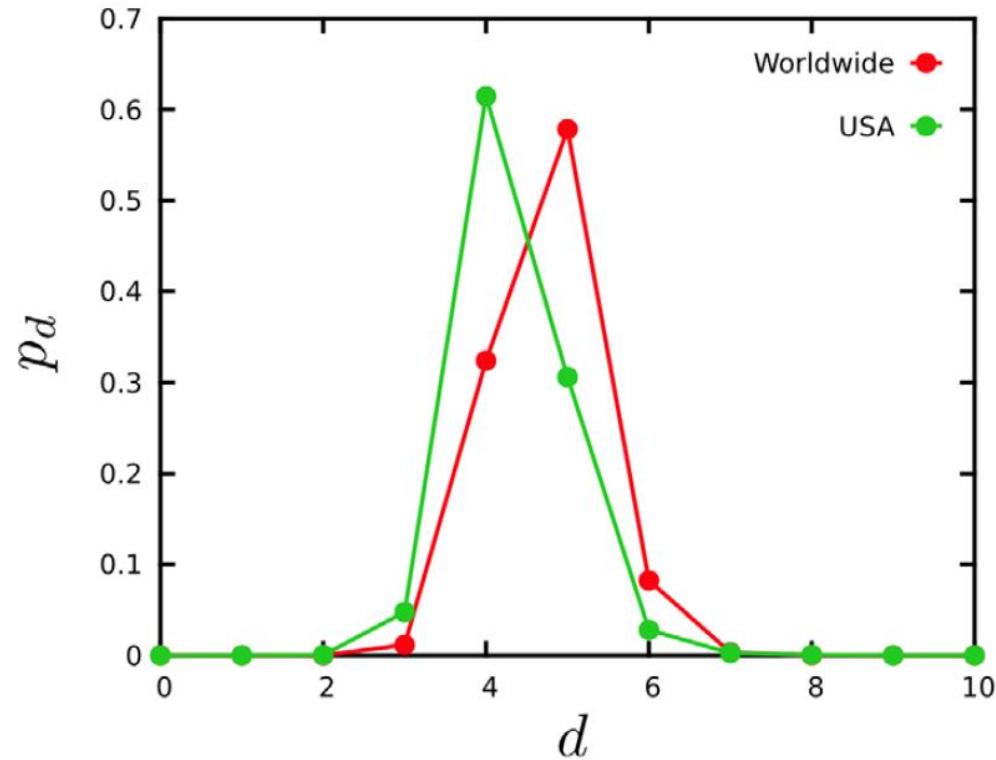
Experimento basado en Facebook:

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

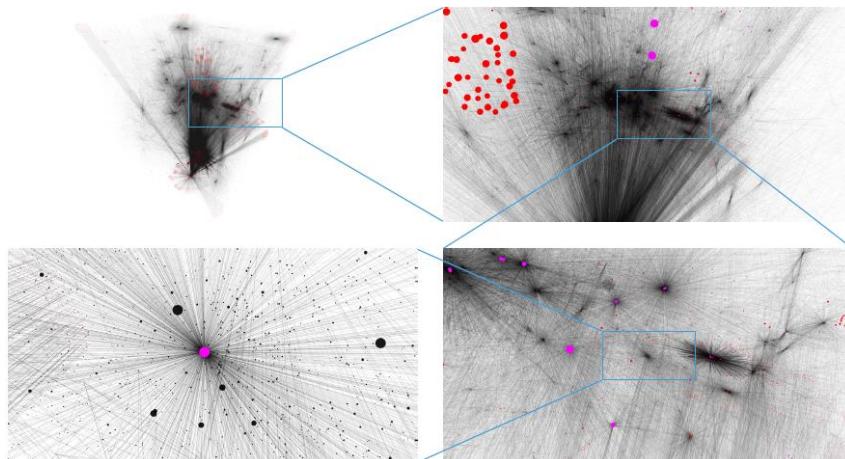
Nodos: **documentos WWW**

Enlaces: **hiperenlaces URL**

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

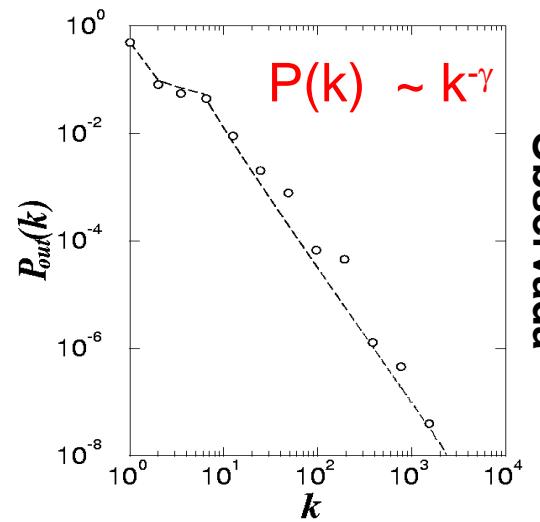
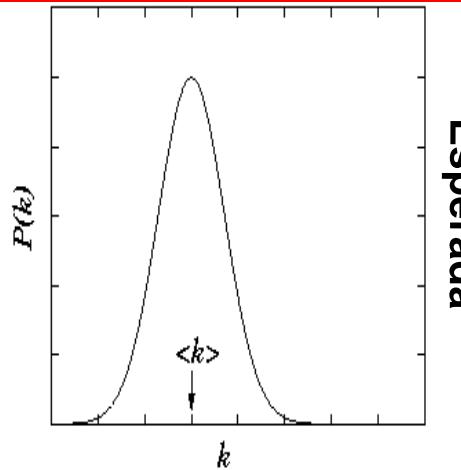
ROBOT: recopila todas las URLs encontradas en una página web y las sigue recursivamente

En 1998 se pensaba que era una red aleatoria:

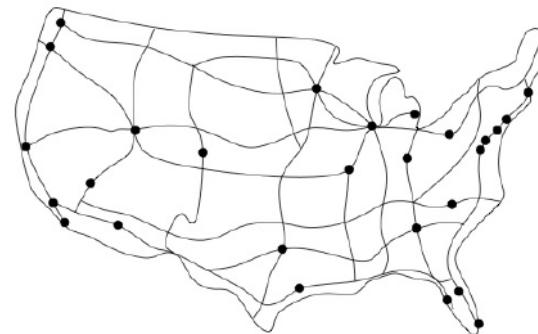
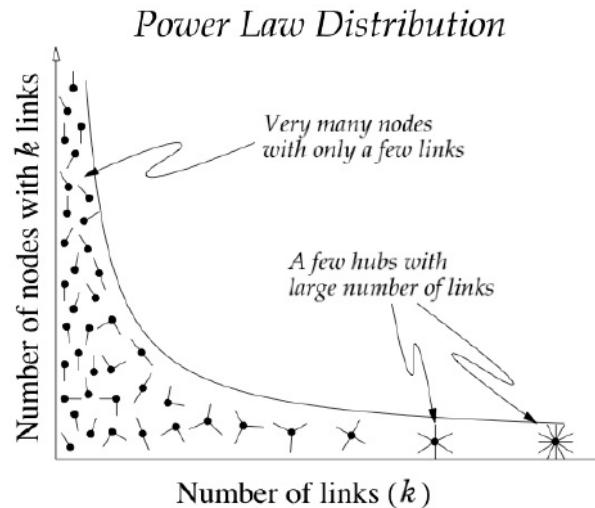
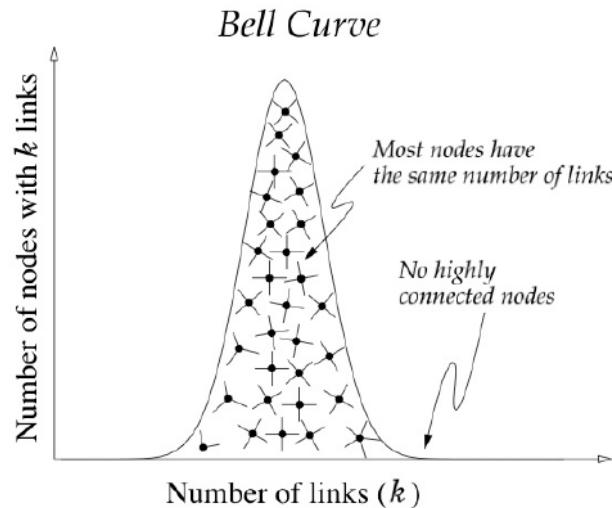


Mapa de la web (dominio nd.edu) con 325,725 páginas (1998)

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

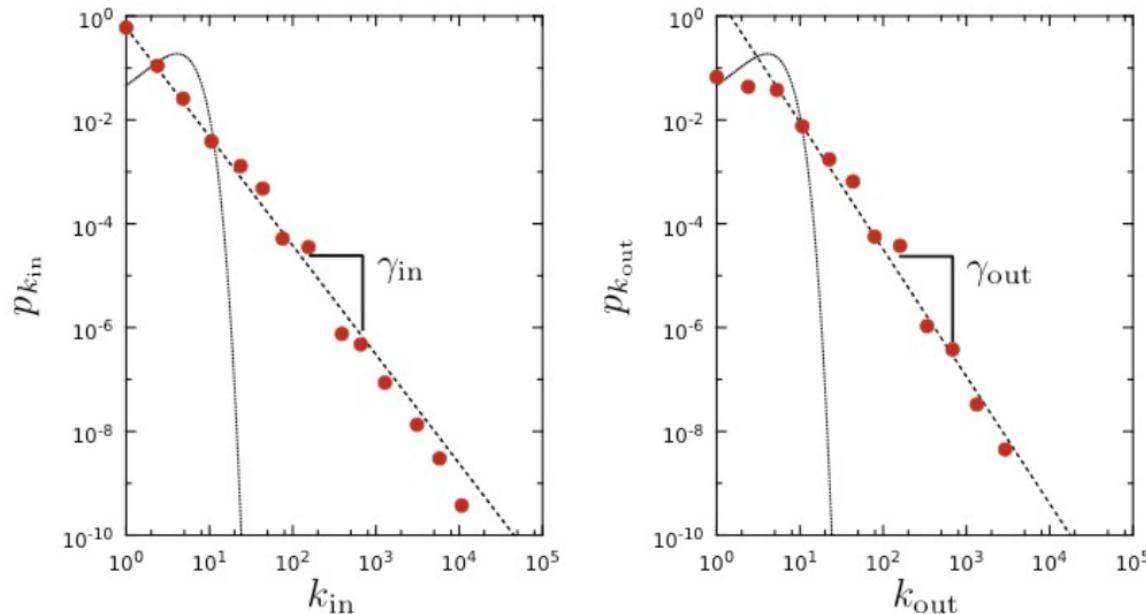


H. Jeong, R. Albert, A-L Barabasi, *Nature*, 401:130-131 (1999)



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.



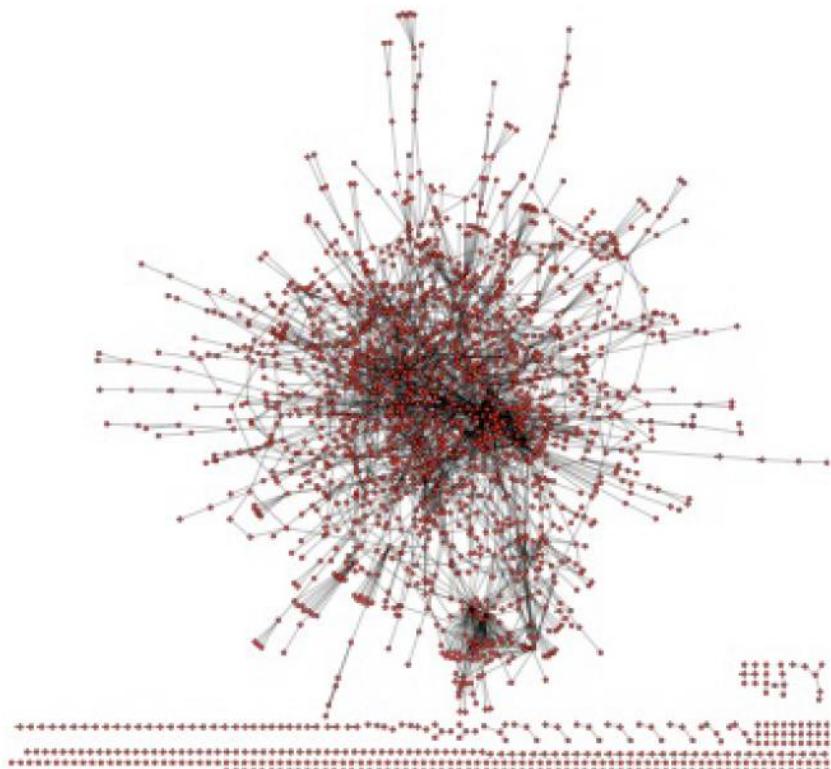
**Figure 4.2: The degree distribution of the WWW.**

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.

# EJEMPLO: ANÁLISIS DE LA RED DE PROTEÍNAS DE LA LEVADURA (1)

Red de interacción entre proteínas (PPI) de la levadura. Los nodos son proteínas y están conectados si hay evidencia empírica de que interactúan

**¡Demasiado compleja para analizarla visualmente o con la matriz de adyacencia!**



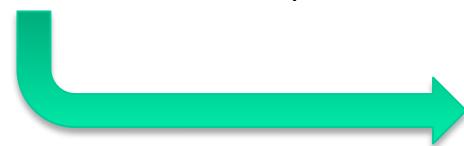
## EJEMPLO: ANÁLISIS DE LA RED DE PROTEÍNAS DE LA LEVADURA (2)

Valores de las medidas:  $N=2018$  nodos,  $L=2930$  enlaces,  $Densidad=0.00144$

Grado medio:  $\langle k \rangle = \frac{2L}{N} = 2.90 \rightarrow$

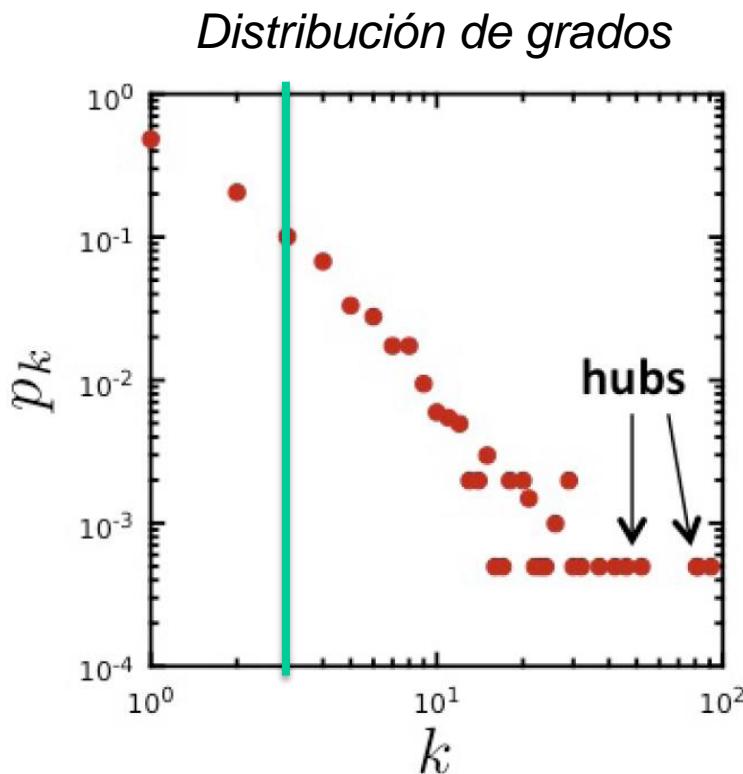
una proteína típica interactúa con otras 2-3

Además, la distribución de grados indica que no hay un nivel de interacción alto. Un 69% de los nodos tienen grado menor que 3 ( $k_i < \langle k \rangle$ )



Existen unos pocos nodos fuertemente conectados (*hubs*), el mayor con grado 91

Consecuencia de la **propiedad libre de escala (scale-free)**, muy común en redes reales (que estudiaremos en temas posteriores)



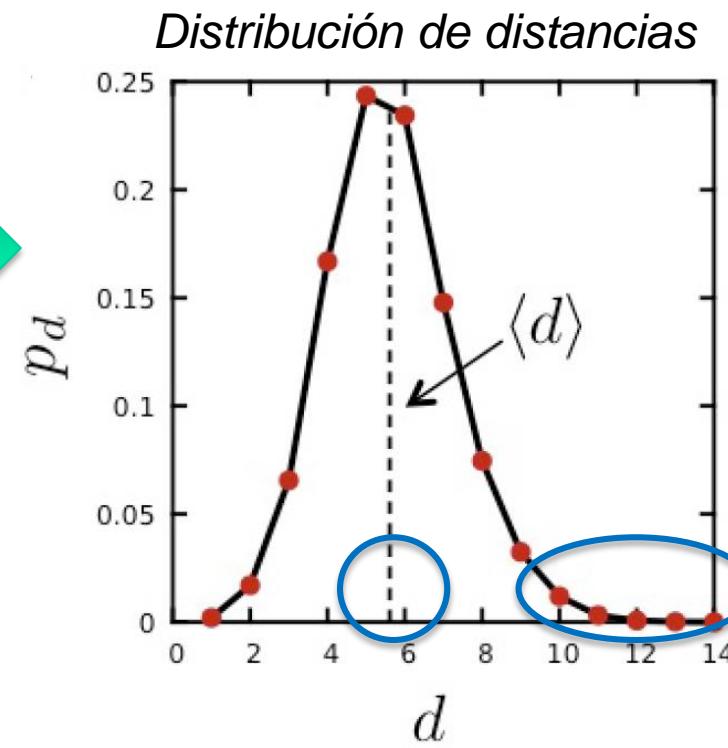
## EJEMPLO: ANÁLISIS DE LA RED DE PROTEÍNAS DE LA LEVADURA (3)

**Diámetro:**  $d_{max}=14$ . Viendo la red, pensaríamos que hay variaciones grandes en las distancias entre nodos pero la distribución tiene una media ( $\langle d \rangle=5.61$ ) y una desviación típica ( $\sigma_d=1.64$ ) bajas (muchos nodos con distancia cercana a la media)

$p_d$  cae fuertemente para las distancias grandes, indicando que no existen distancias grandes



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



## EJEMPLO: ANÁLISIS DE LA RED DE PROTEÍNAS DE LA LEVADURA (4)

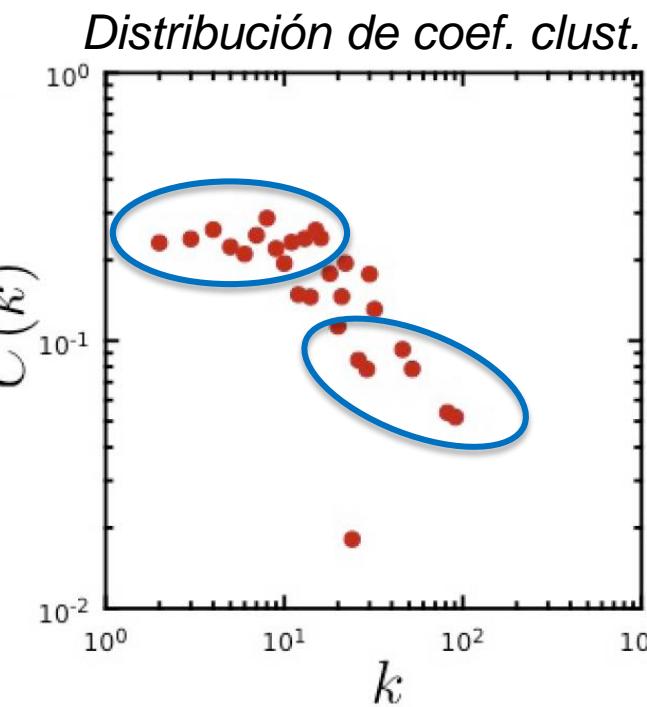
**Conectividad:** la red presenta 185 componentes conexas, muchas de ellas formados por una sola proteína aislada que no interactúa con el resto

Vemos una componente gigante que agrupa el 81% de los nodos: 1647 de los 2018. Hay también algunas componentes aisladas muy pequeñas. Esta **fragmentación** también es común en redes reales

**Coeficiente de clustering medio:**  $\langle C \rangle = 0.12$ . **Aunque no lo parezca, es bastante alto**, indicando un grado significativo de clustering local

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**



# 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

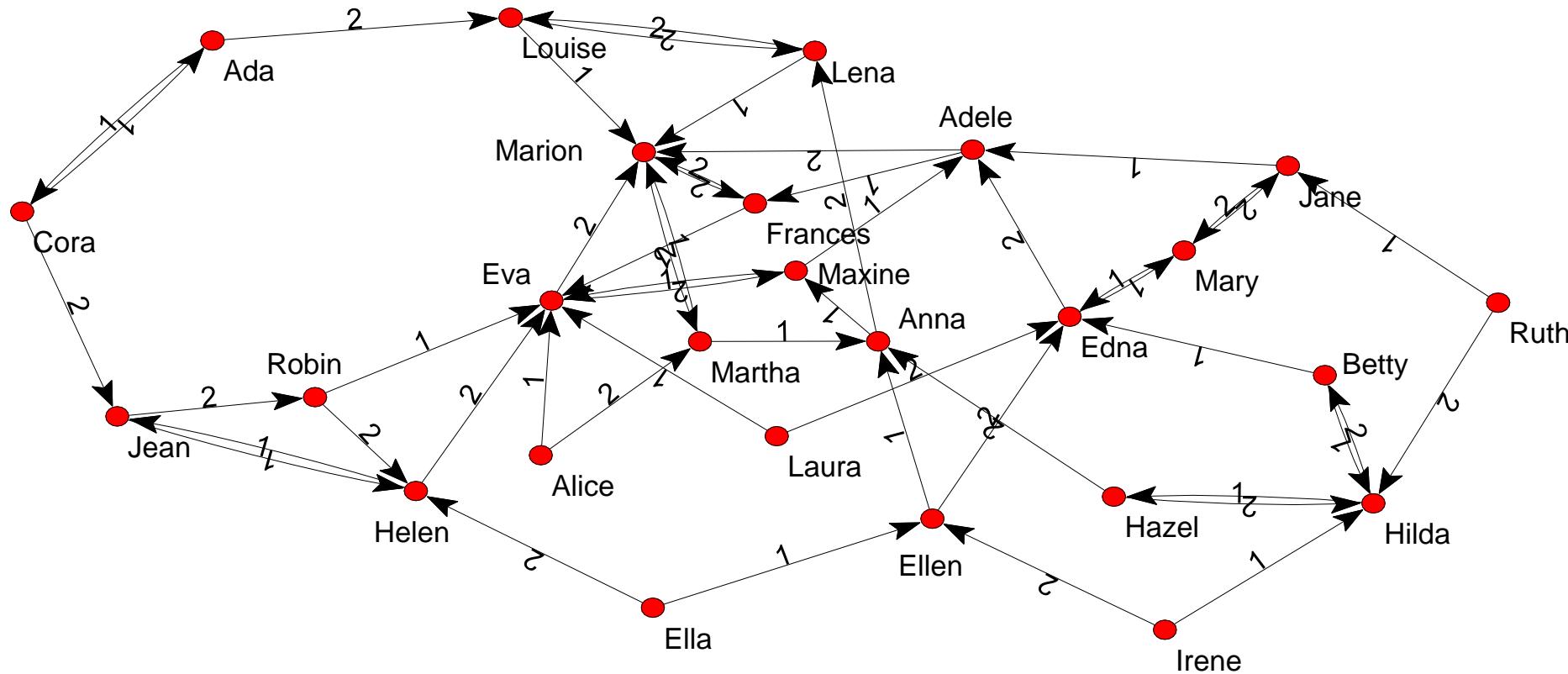
Según Newman, las redes reales se pueden categorizar en cuatro tipos principales:

- 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

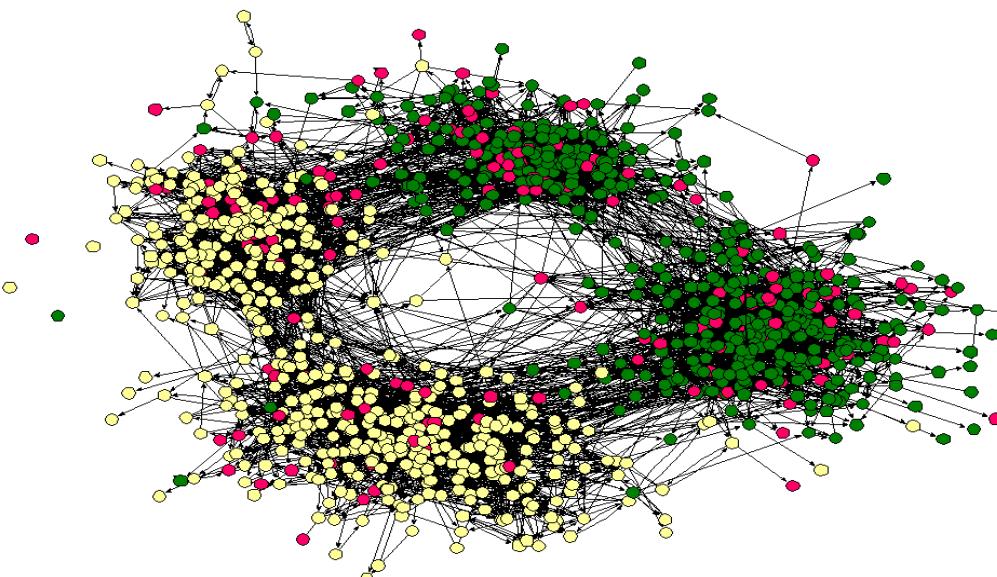
Las redes sociales son las que resultan de interacciones humanas y sociales, como las redes de amistad, las redes informales de comunicación entre empresas y las redes de colaboración (p.ej., la red de actores de Hollywood y las redes de coautoría científica), entre otras

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

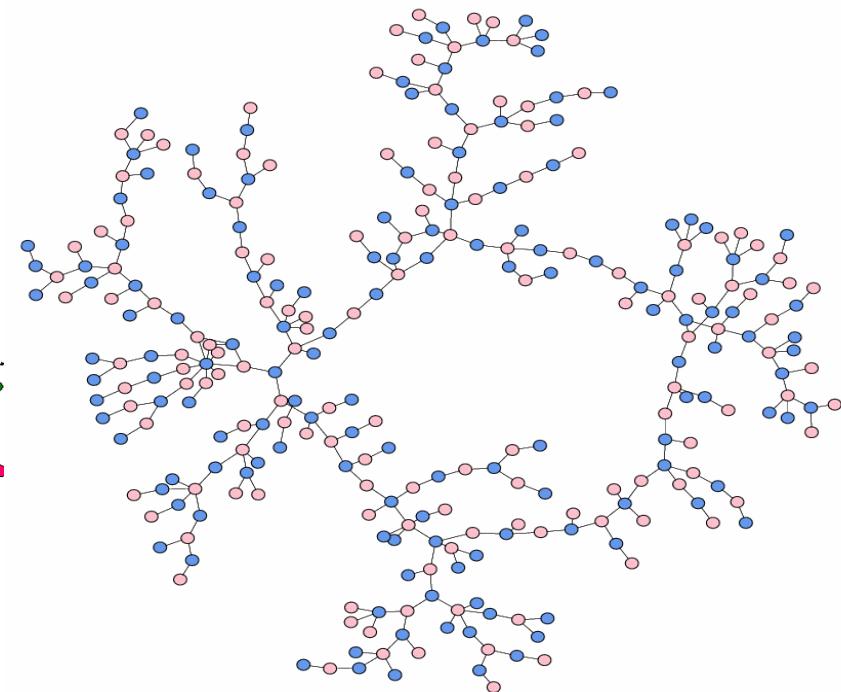


## EJEMPLOS (2):

## Redes de Amistad y Redes Sexuales



Relaciones de amistad en  
un instituto Americano



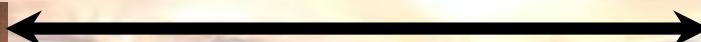
Relaciones sexuales en un  
instituto Americano

## EJEMPLOS (3):

## Red de Actores de Hollywood

Nodos: actores

Enlaces: actuaciones conjuntas



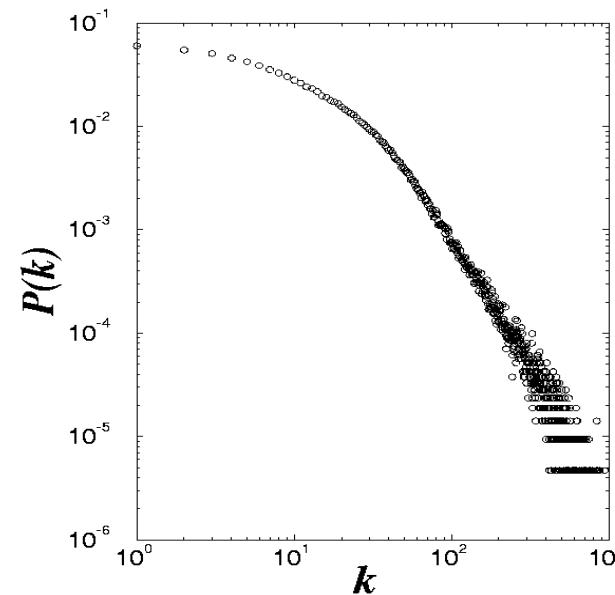
*Days of Thunder* (1990)  
*Far and Away* (1992)  
*Eyes Wide Shut* (1999)

$N = 212,250$  actores

$\langle k \rangle = 28.78$

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

$\gamma = 2.3$



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

## Facebook

Screenshot of a Facebook profile page for Oscar Cordón García.

**Header:** Busca personas, lugares y cosas | | Oscar | Inicio | | |

**Left Sidebar (Menu):**

- 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](#)
  - [Reunion dia 25 de ...](#)
  - [Crear grupo](#)
  - [Ver grupos nuevos](#)

**Central Content Area:**

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

¿Qué estás pensando?

**Post by 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  
**Reactions:** Ati, Carmen Zarco Fernández, Caty Rguez Cordon, Eva McCormack y 3 personas más os gusta esto.

**Comments:**

- 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...

**Recent Activity:** Pedro López Gallardo ha añadido 11 fotos nuevas. 2 h ·

**Right Sidebar:**

16 2 invitaciones a eventos

Hoy es el cumpleaños de [Mike Nachtegael](#)

PERSONAS QUE QUIZÁ CONOZCAS Ver todas

Maribel Garcia Arenas 14 amigos en común [Añadir a mis amigos](#)

Javier Álvarez 27 amigos en común [Añadir a mis amigos](#)

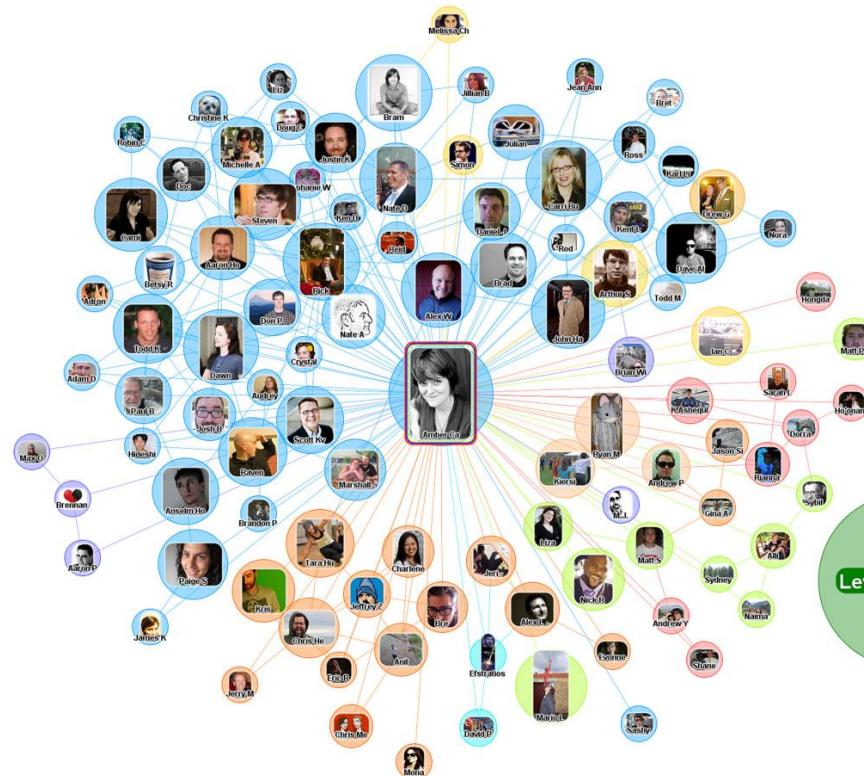
Luz Garcia 5 amigos en común [Añadir a mis amigos](#)

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):

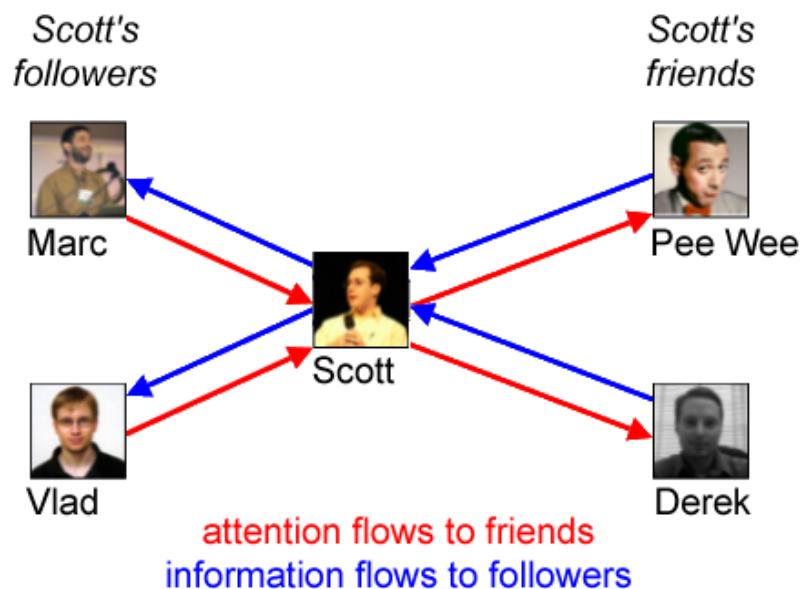
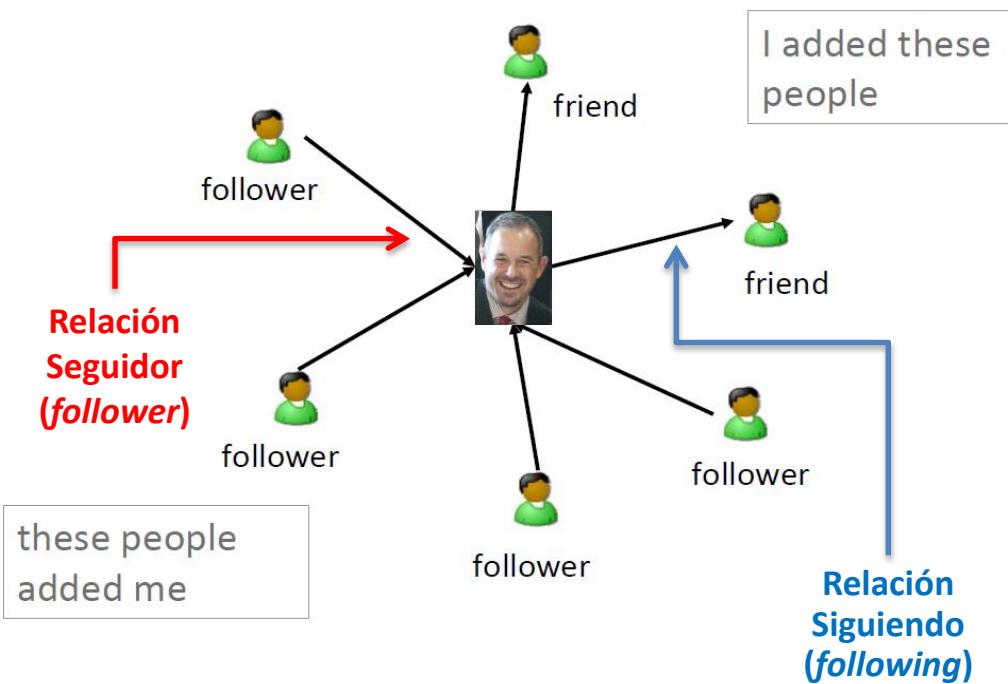
## Twitter

The screenshot shows the Twitter interface for the 'Who to Follow' section. At the top, there's a navigation bar with links like Home, Profile, Messages, and Who To Follow. Below that, a search bar and a 'Who to Follow' button are visible. The main content area is titled 'Who to follow' and includes sections for 'View Suggestions', 'Browse Interests', and 'Find Friends'. A sidebar on the left lists interests under 'View Technology (60)' including Art & Design, Books, Business, Charity, Deals, Entertainment, Family, Fashion, Food & Drink, Funny, Health, Music, News, Politics, Science, Sports, Staff Picks, and Staff Picks: NFL Playoffs, Technology, Travel, and Twitter. The main content area displays profiles for several users:

- guardiantech**: Guardian Tech. News and comment from the Guardian's technology team. Follow button.
- kaifulee**: Kai-Fu Lee. I am the CEO of Innovation Works (www.innovation-works.com). I used to work at Microsoft, SGI, Apple, and Google. I have written several Chinese books. Follow button.
- sacca**: Chris Sacca. I'm an investor in Twitter & other startups. Learn more at www.lowercaselc.com. Just finished bicycling across the USA and am thinking about my next adventure. Follow button.
- kevinrose**: Kevin Rose. Tech angel investor. Founder of digg.com, wefollow.com. Random ideas, entrepreneur, climber of rocks, video blogger, & tea drinker. Follow button.
- ars**: arstechnica. Ars Technica. The Art of Technology. Follow button.
- lessig**: Lessig. law prof, reformer. Follow button.
- leolaporte**: Leo Laporte. Podcaster, broadcaster and tech pundit. The Tech. Follow button.

On the right side, there are sections for 'You recently followed' (with a link to 'View All') and 'Similar to Michelle Cale'. It also shows a list of recently viewed profiles and provides an 'Invite Friends' feature where you can enter an email address and click 'Invite'. The bottom of the page features a 'Related Services' section with links to Microsoft and LinkedIn, along with a footer containing standard links like About, Help, and Privacy.

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

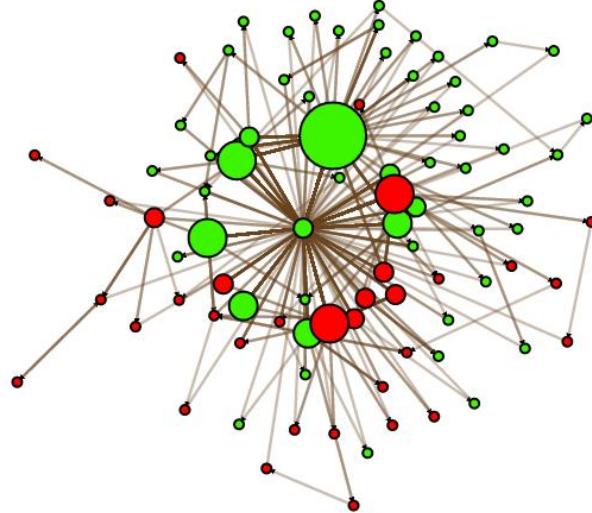


Ejemplo de red de amistad de Twitter

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

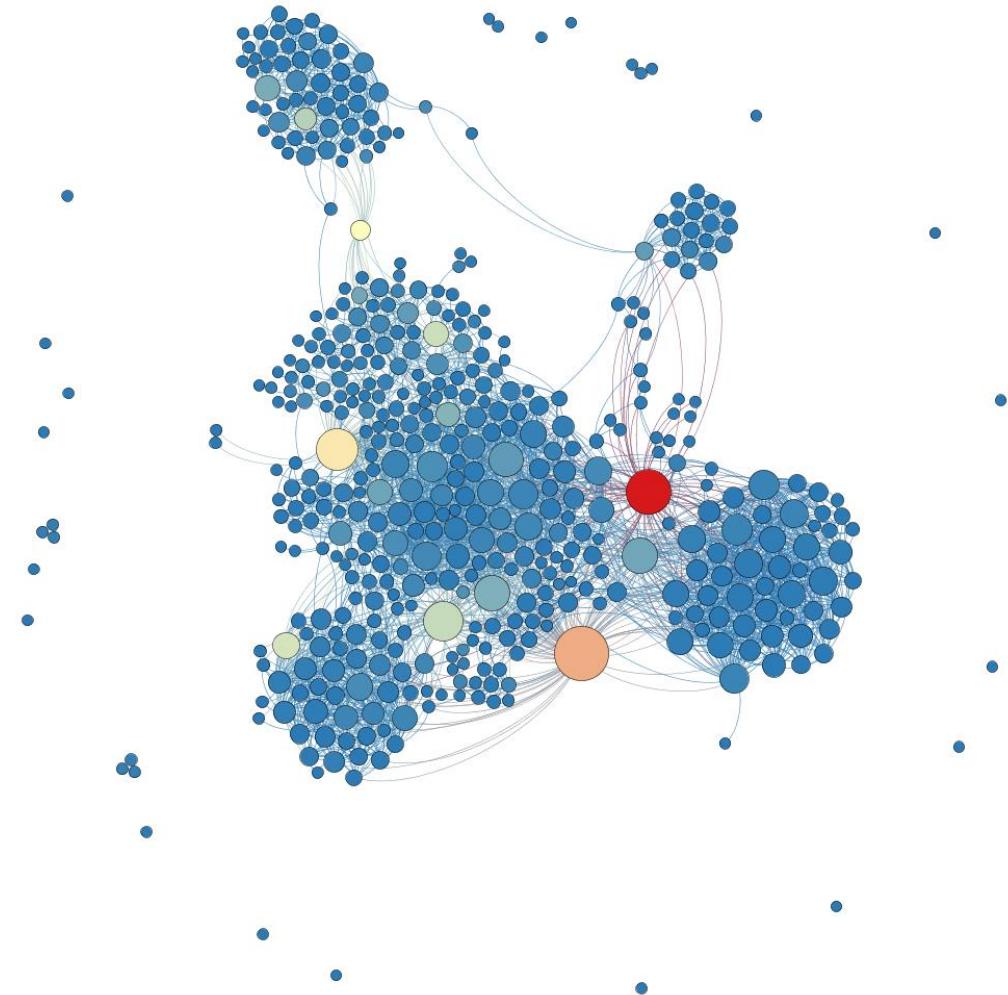
- **Redes de flujo de información** (*information flow*): Los nodos son usuarios y los enlaces (dirigidos) indican el camino seguido por el tweet (“quién ha retuiteado a quién”). La visualización de los *retweets* refleja el flujo de información en Twitter
- **Redes de co-ocurrencia de hashtags**: Los 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)

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

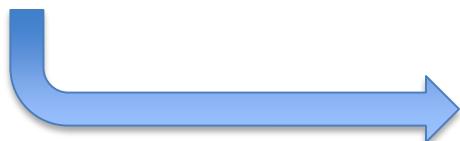


# ANÁLISIS BÁSICO DE LA RED DE FACEBOOK DE OSCAR CORDÓN (2)

**Valores de las medidas:**  $N=483$  nodos,  $L=3323$  enlaces,  $Densidad=0.029$

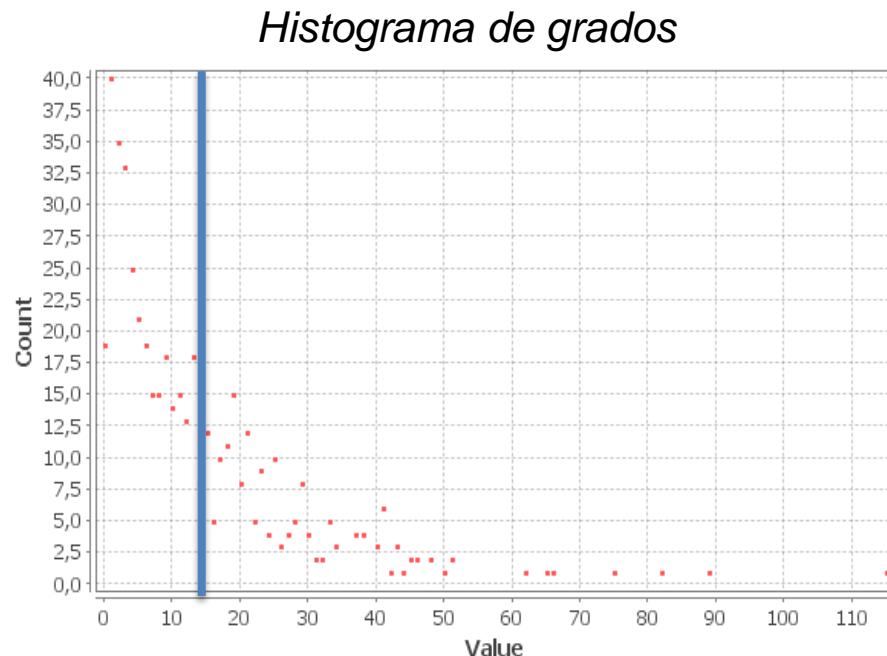
Grado medio:  $\langle k \rangle = \frac{2L}{N} = 13.76 \rightarrow$  cada nodo de la red tiene casi 14 amigos en media

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



Existen unos pocos nodos fuertemente conectados (**hubs**), el mayor con grado 115. Sólo el 1% de los nodos (5 nodos) tienen más de 66 amigos (66, 75, 62, 89 y 115)

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



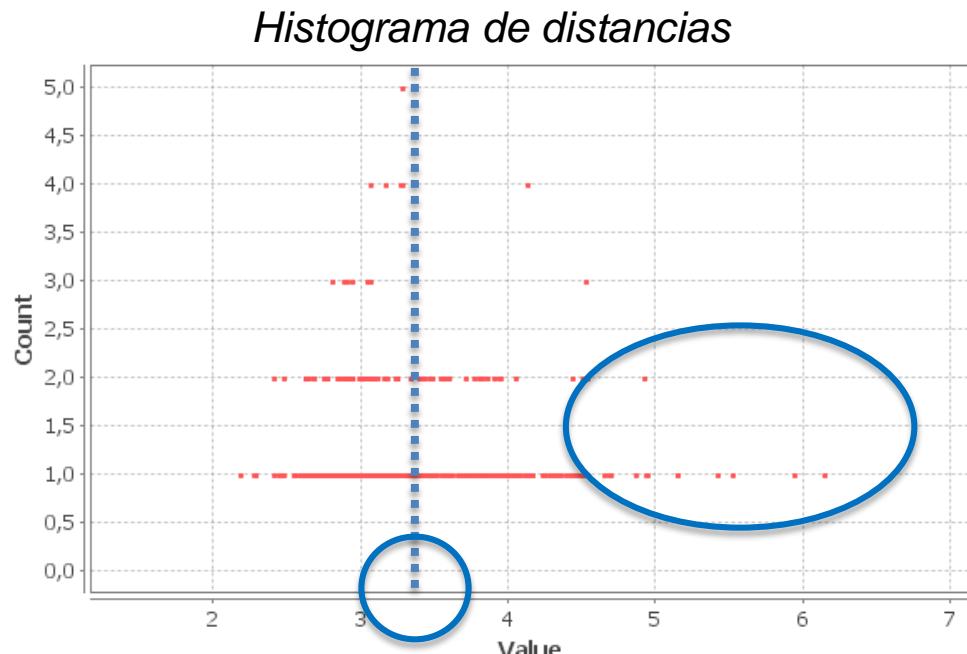
# ANÁLISIS BÁSICO DE LA RED DE FACEBOOK DE OSCAR CORDÓN (3)

**Diámetro:**  $d_{max}=9$ . Viendo la red, pensaríamos que hay variaciones grandes en las distancias entre nodos pero la distribución tiene una media ( $\langle d \rangle=3.35$ ) y una desviación típica bajas (muchos nodos con distancia cercana a la media)

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 (“**teoría de los seis grados**”)

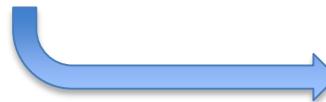


# ANÁLISIS BÁSICO DE LA RED DE FACEBOOK DE OSCAR CORDÓN (4)

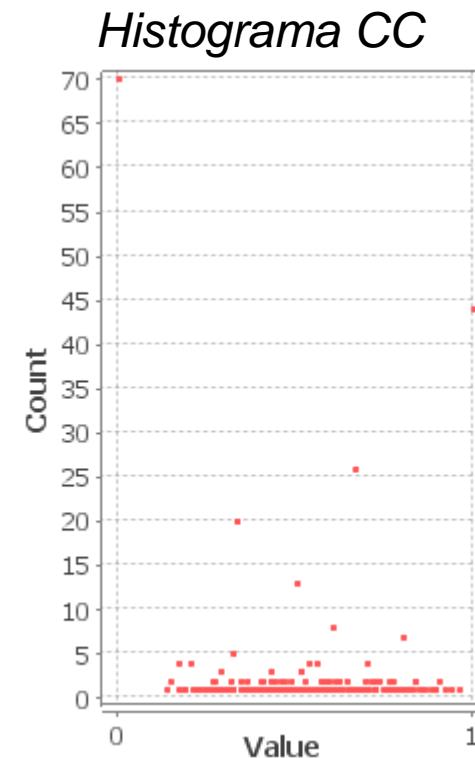
**Conectividad:** la red presenta 24 componentes conexas, muchas de ellas (19) formadas por una sola persona aislada que no tiene relación con el resto

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 bastante alto** (habitual en redes sociales), indicando un grado significativo de clustering local



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**



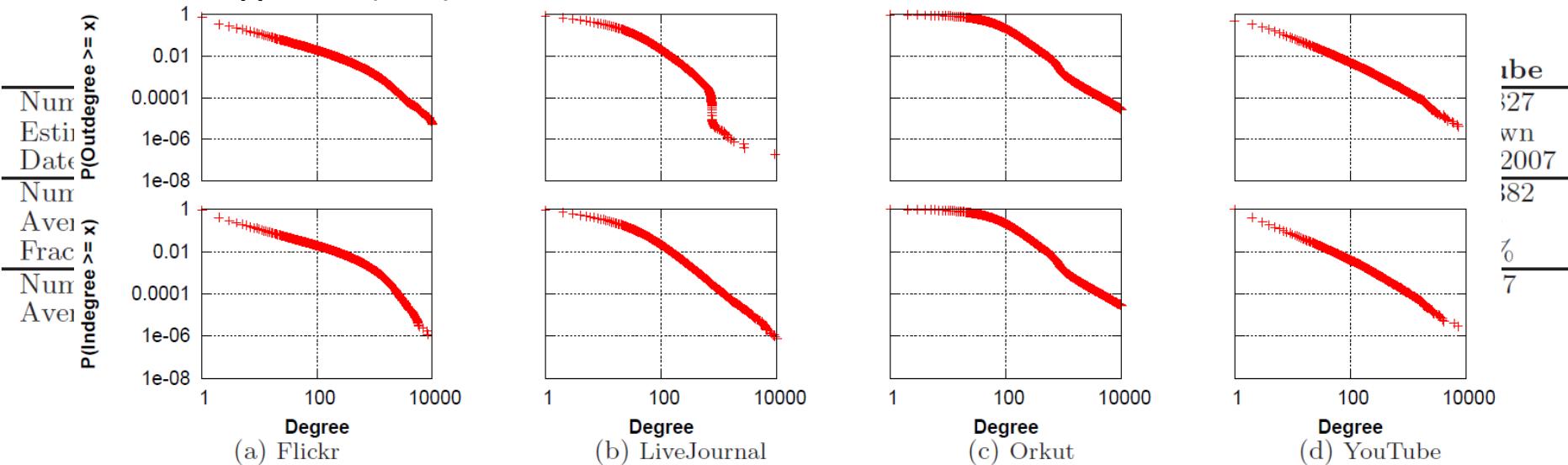
# ANÁLISIS DE REDES SOCIALES ON-LINE (1)

**Nodos:** usuarios

**Enlaces:** relaciones

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	
	$\alpha$	$D$	$\alpha$	$D$
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 ( $\alpha$ ) 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 (2)

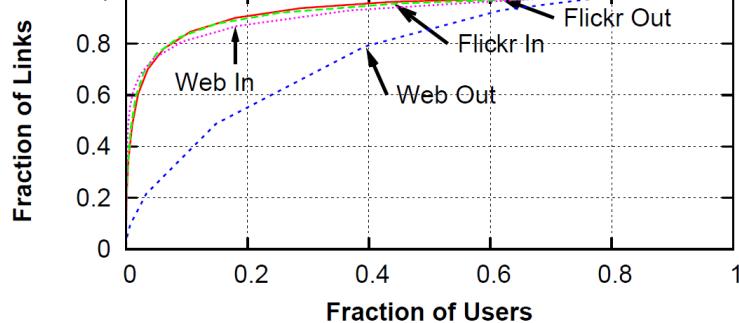


Figure 3: Plot of the distribution of links across nodes. Social networks show similar distributions for outgoing and incoming links, whereas the Web links shows different distributions.

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.

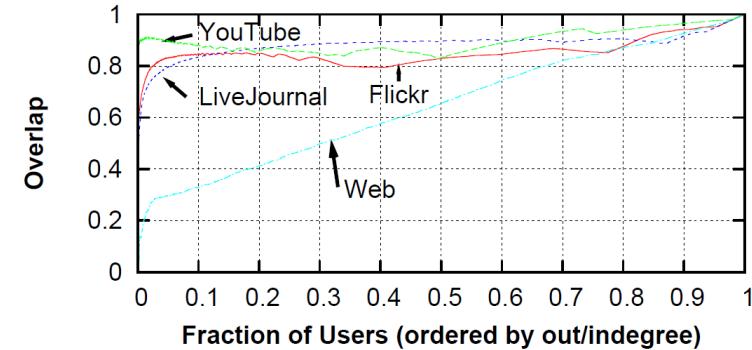


Figure 4: Plot of the overlap between top  $x\%$  of nodes ranked by outdegree and indegree. The high-indegree and high-outdegree nodes are often the same in social networks, but not in the Web.

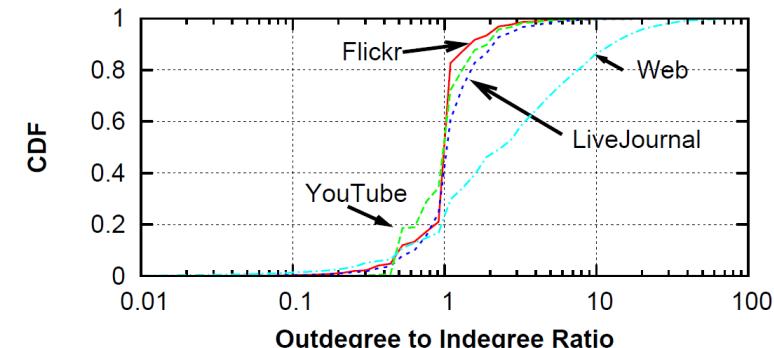


Figure 5: CDF of outdegree to indegree ratio. Social networks show much stronger correlation between indegree and outdegree than the Web.

# 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)

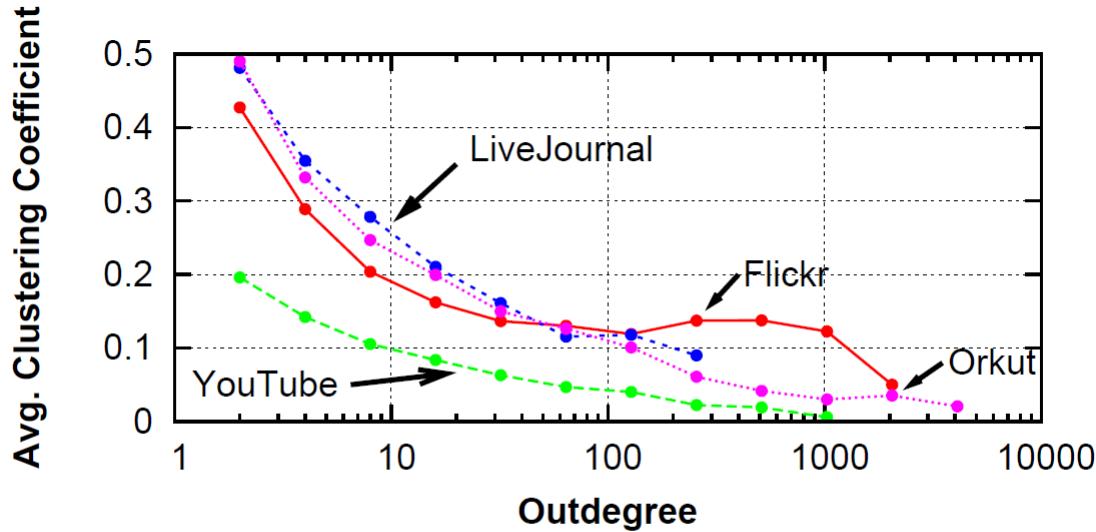


Figure 9: Clustering coefficient of users with different outdegrees. The users with few “friends” are tightly clustered.

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.

# 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>

