

# Introduction to Computational Social Science

## Redes Complejas: Random and Scale Free Networks

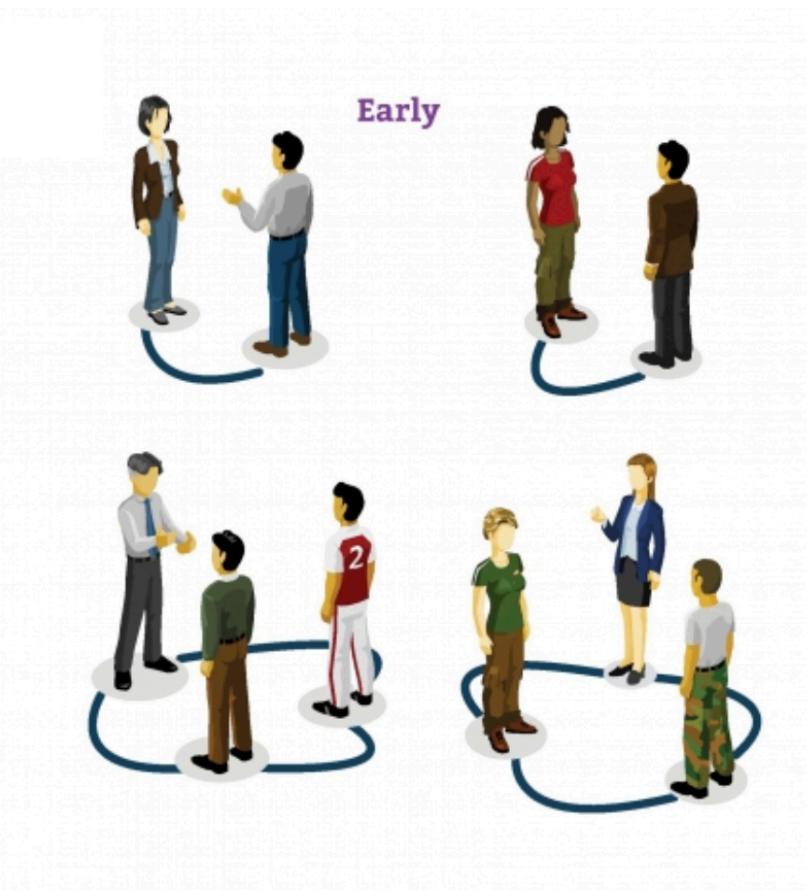
Cristian Candia-Castro Vallejos, Ph.D.

Assistant Professor, Facultad de Ingeniería, UDD  
Director of the Data Science Master's Degree, UDD  
External Faculty Northwestern Institute on Complex Systems, Northwestern University, USA  
Founder at CINS-Data

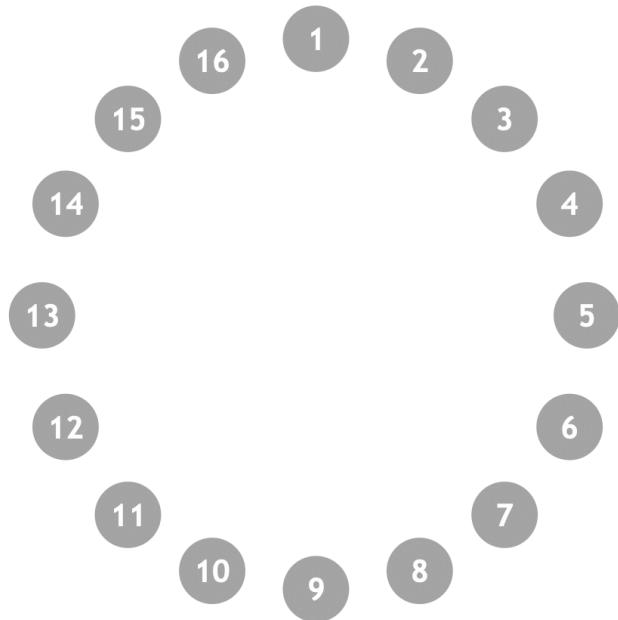
Ayudantes: Melanie Oyarzún, Ph.D.(c) y Victor Landaeta, Ph.D.

# Introducción

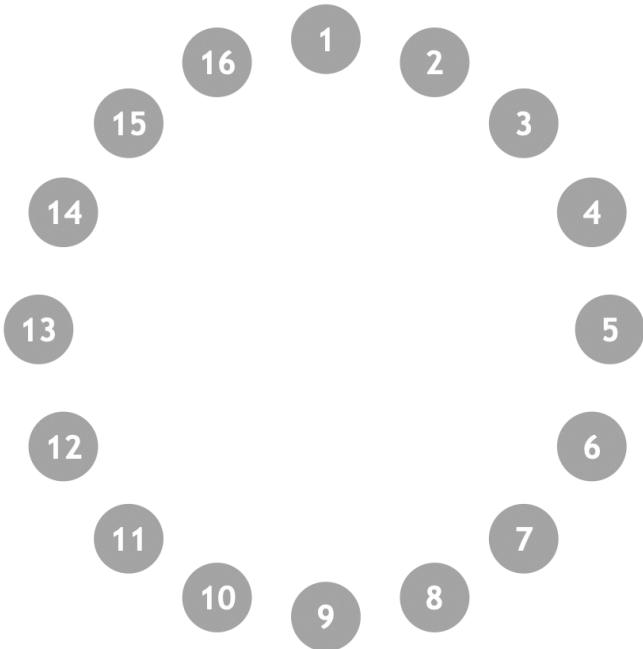
# MODELO DE REDES ALEATORIAS



# the party algorithm

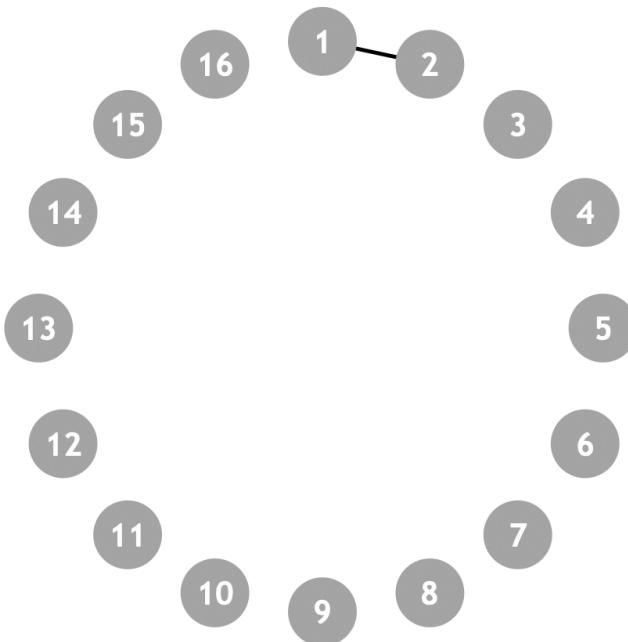


# the party algorithm



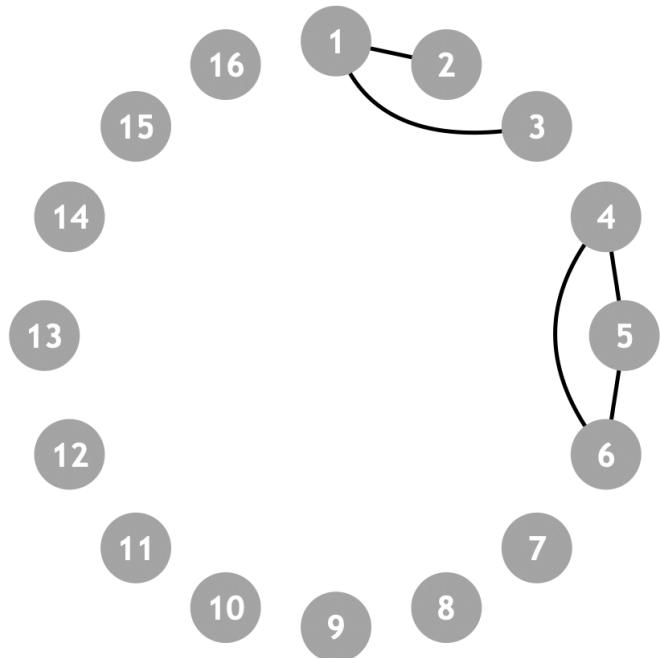
Hagamos un seguimiento de quién habla con quién

# the party algorithm



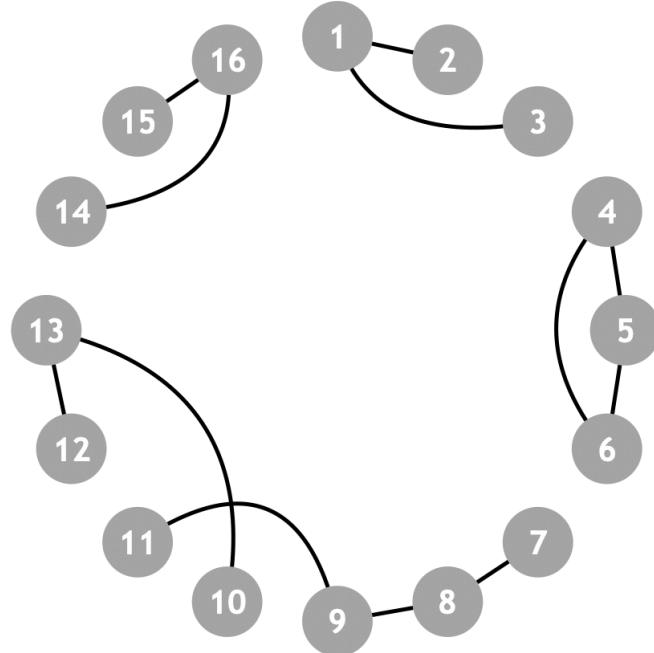
Hagamos un seguimiento de quién habla con quién

# the party algorithm



Hagamos un seguimiento de quién habla con quién

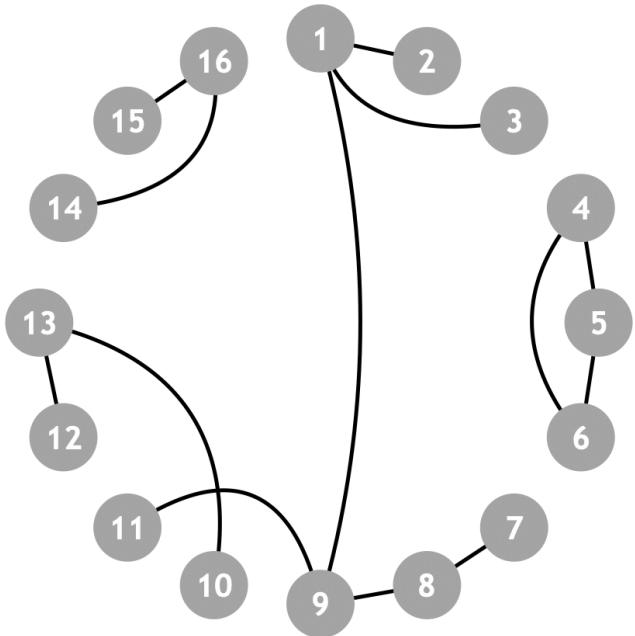
# the party algorithm



Hagamos un seguimiento de quién habla con quién

Al principio, la gente habla principalmente con otros que están sentados cerca.

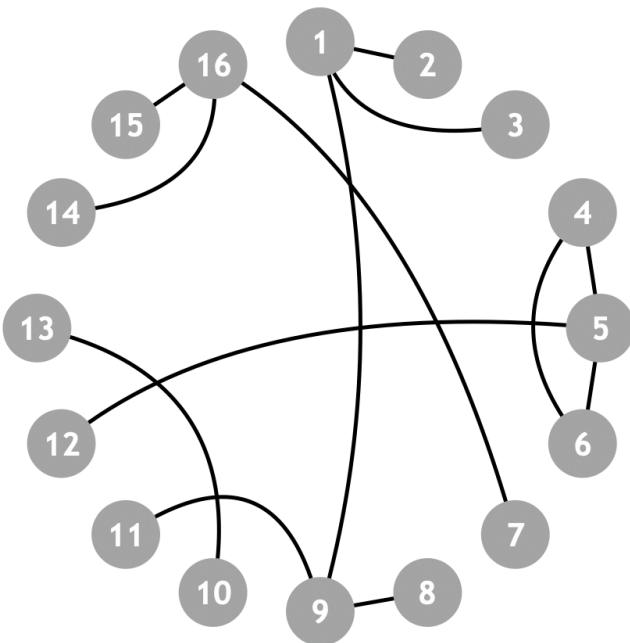
# the party algorithm



Hagamos un seguimiento de quién habla con quién

Sin embargo, como la cena avanza hacia las bebidas nocturnas y los invitados se ponen de pie, vemos la reorganización de sus relaciones y se forman nuevos vínculos

# the party algorithm

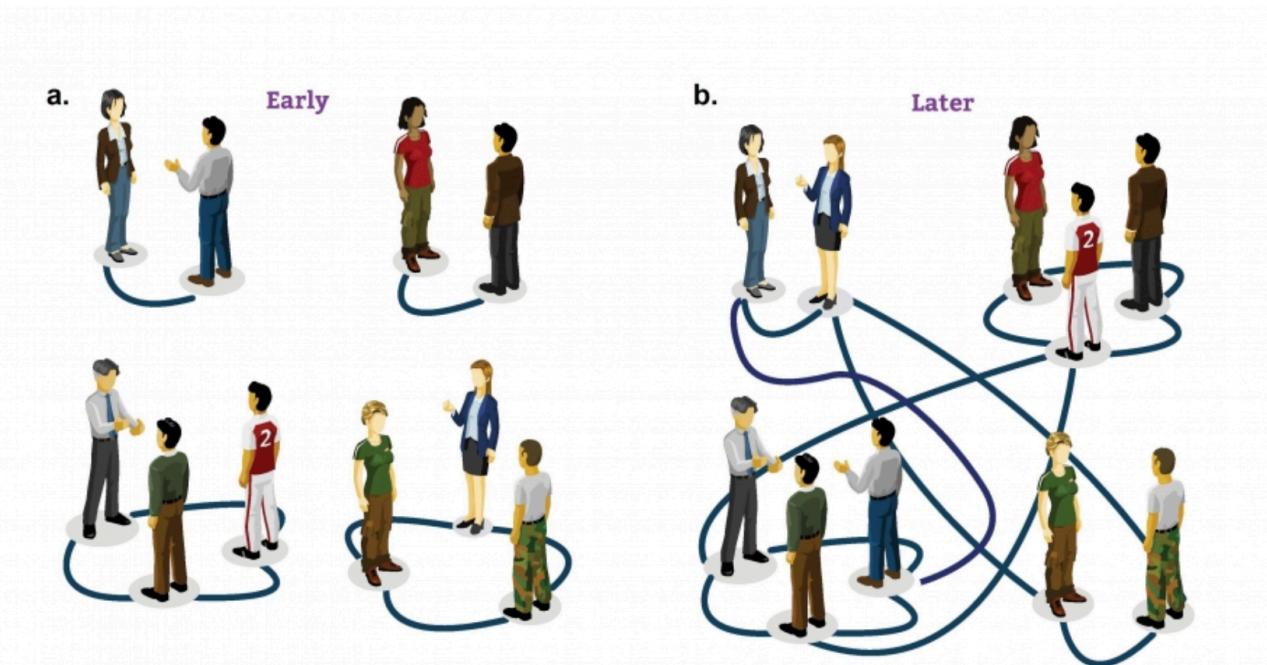


Estas redes no siguen ninguna regla física como en los ejemplos anteriores.

En cambio, son algo aleatorias.  
¿No te parece?

Pregunta: ¿Podemos construir un modelo que intente capturar las propiedades de estas redes?

# Un algoritmo simple para redes sociales

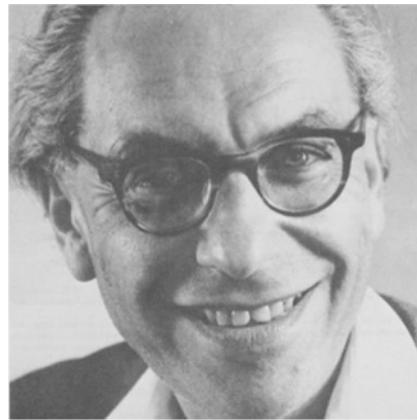


Invitados a una cena se reunen al azar y establecen relaciones sociales.

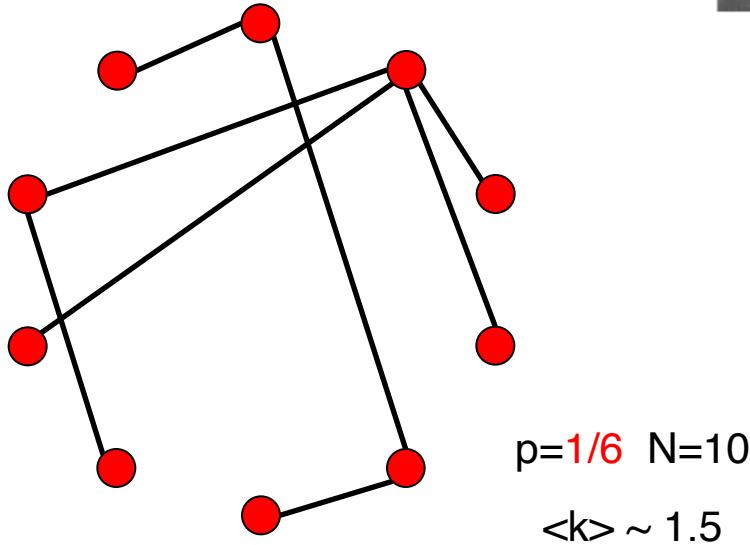
# El modelo de redes aleatorias

# MODELO DE REDES ALEATORIAS

Pál Erdős  
(1913-1996)



Alfréd Rényi  
(1921-1970)

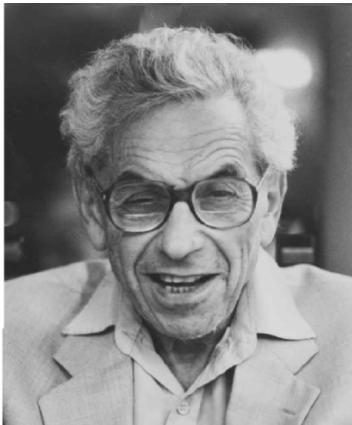


Erdős-Rényi model (1960)

Un **grafo aleatorio** es un grafo de  $N$  nodos donde cada par de nodos está conectado con una probabilidad  $p$ .

# Random Networks

a.

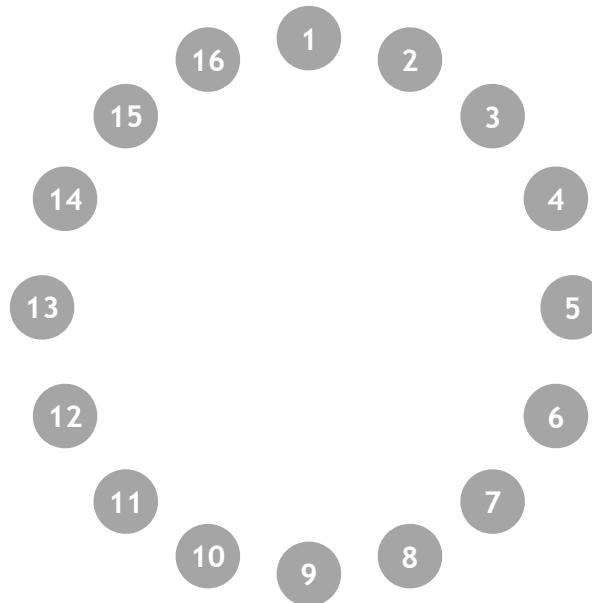


Pál Erdős

b.



Alfréd Rényi



## Erdős-Rényi algorithm

**condición inicial** Comenzando con  $N$  nodos desconectados ( $L = 0$  enlaces)

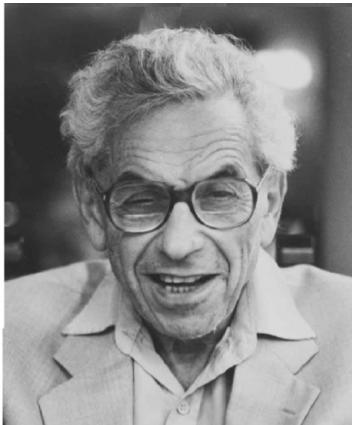
**paso iterativo** Para cada par de individuos, establezca un vínculo con probabilidad  $p$

# Random Networks

Conectar 1 con 2?

is Random() < p

a.

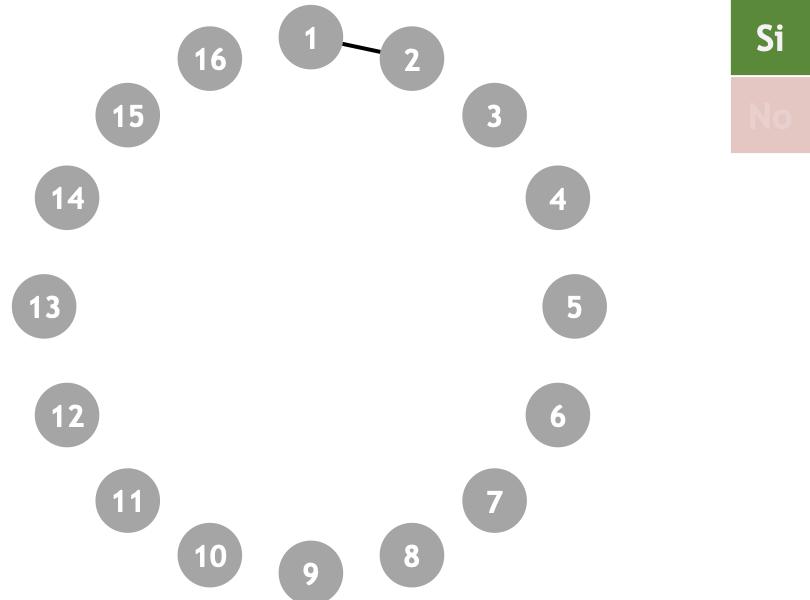


Pál Erdős

b.



Alfréd Rényi



## Erdős-Rényi algorithm

**condición inicial** Comenzando con N nodos desconectados ( $L = 0$  enlaces)

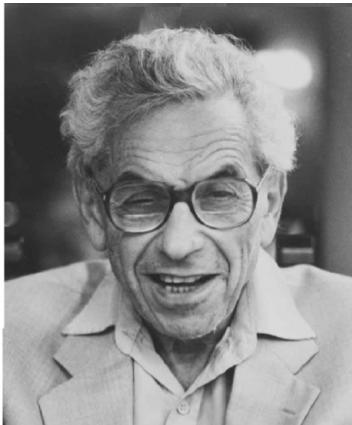
**paso iterativo** Para cada par de individuos, establezca un vínculo con probabilidad  $p$

# Random Networks

Conecitar 1 con 3?

is Random() < p

a.

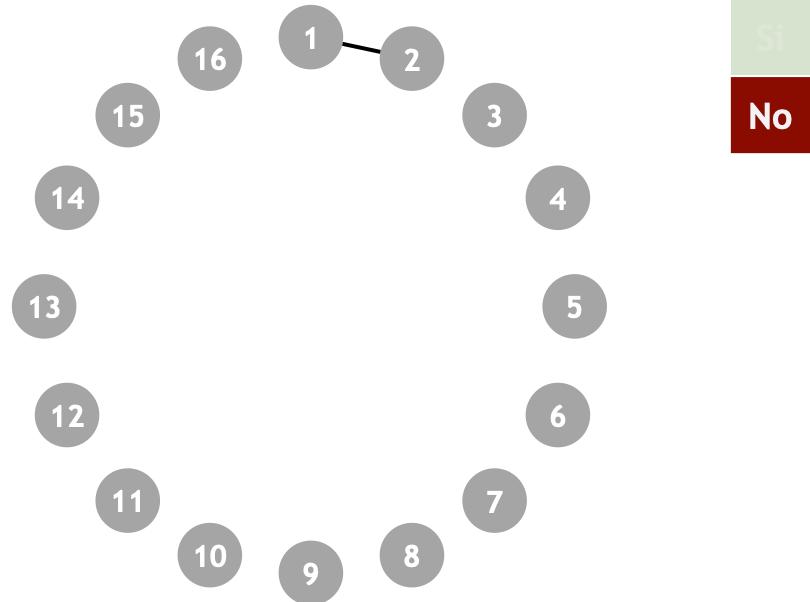


Pál Erdős

b.



Alfréd Rényi



## Erdős-Rényi algorithm

**condición inicial** Comenzando con N nodos desconectados ( $L = 0$  enlaces)

**paso iterativo** Para cada par de individuos, establezca un vínculo con probabilidad  $p$

# Random Networks

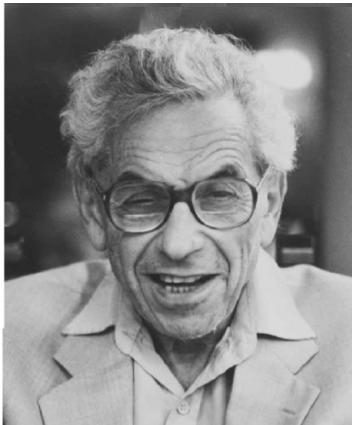
Conectar 1 con 4?

is Random() < p

Si

No

a.

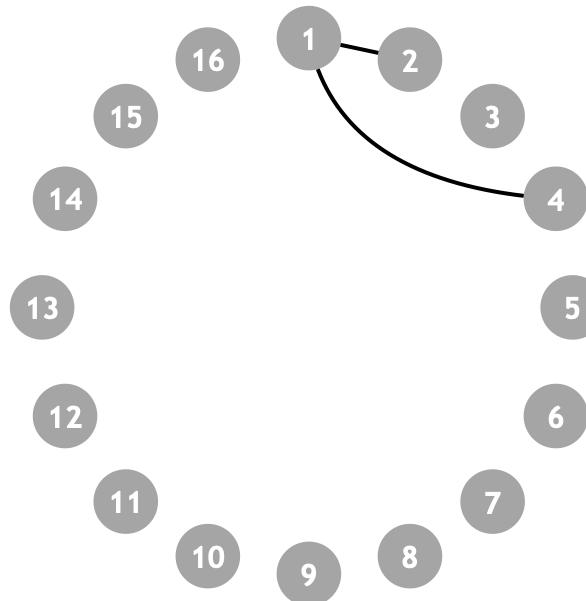


Pál Erdős

b.



Alfréd Rényi



## Erdős-Rényi algorithm

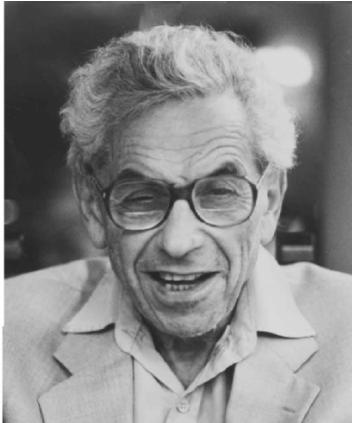
**condición inicial** Comenzando con N nodos desconectados ( $L = 0$  enlaces)

**paso iterativo** Para cada par de individuos, establezca un vínculo con probabilidad  $p$

# Random Networks

Repite para todas las Posibles conexiones de 1

a.

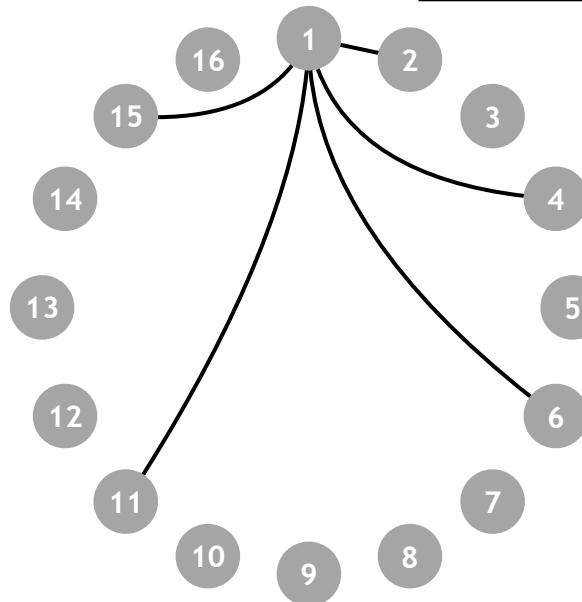


Pál Erdős

b.



Alfréd Rényi



## Erdős-Rényi algorithm

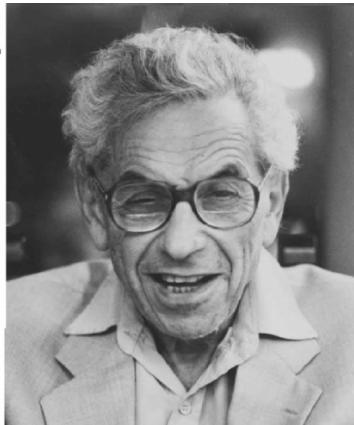
**condición inicial** Comenzando con  $N$  nodos desconectados ( $L = 0$  enlaces)

**paso iterativo** Para cada par de individuos, establezca un vínculo con probabilidad  $p$

# Random Networks

Repite para todos  
Los posibles enlaces

a.

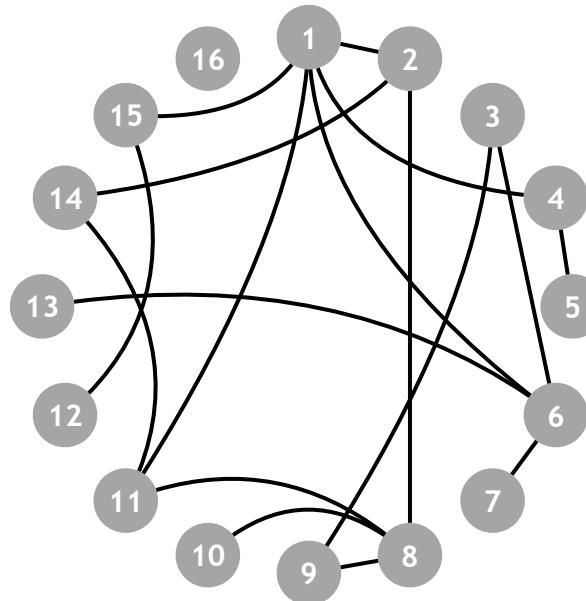


Pál Erdős

b.



Alfréd Rényi



## Erdős-Rényi algorithm

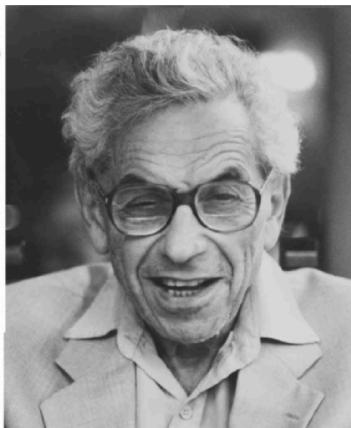
**condición inicial** Comenzando con  $N$  nodos desconectados ( $L = 0$  enlaces)

**paso iterativo** Para cada par de individuos, establezca un vínculo con probabilidad  $p$

El número de enlaces es variable.

# MODELO DE REDES ALEATORIAS

a.



Pál Erdős

b.



Alfréd Rényi

## Erdős-Rényi algorithm

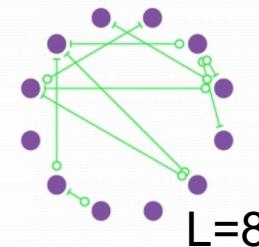
### condición inicial

Comenzando con  $N$  nodos desconectados ( $L = 0$  enlaces)

### paso iterativo

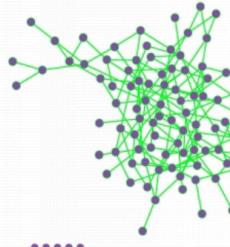
Para cada par de individuos, establezca un vínculo con probabilidad  $p$

$$N=12 \quad p=1/6$$

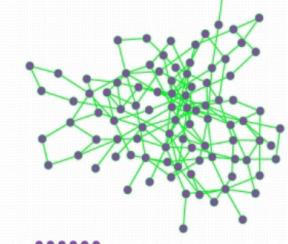


$$L=8$$

$$N=100 \quad p=0.03$$



$$L=7$$



Parece una red real, ¿no es así?

# Número de enlaces en una red aleatoria

**P(L)**: la probabilidad de tener exactamente **L** enlaces en una red de **N** nodos y probabilidad **p**:

$$P(L) = \binom{N}{L} p^L (1-p)^{\frac{N(N-1)}{2} - L}$$

El número máximo de enlaces  
en una red de **N** nodos.

Número de formas diferentes en que podemos  
elegir los enlaces **L** entre todos los enlaces  
potenciales.

Distribución binomial...

$$\binom{n}{k} = \frac{n(n-1)\cdots(n-k+1)}{k(k-1)\cdots 1},$$

$$\binom{n}{k} = \frac{n!}{k!(n-k)!},$$

# Tutorial Matemático. | Distribución Binomial: La línea base

$$p_x = \binom{N}{x} p^x (1-p)^{N-x}$$

Dist. Binomial de "x"

Ejemplo:  
N=Enlaces totales potenciales  
x=Enlaces de la red  
p= probabilidad

$$\langle x \rangle = \sum_{x=0}^N x p_x = Np$$

$$\langle x^2 \rangle = p(1-p)N + p^2N^2$$

$$\sigma_x = (\langle x^2 \rangle - \langle x \rangle^2)^{\frac{1}{2}} = [p(1-p)N]^{1/2}$$

# MODELO DE REDES ALEATORIAS

**P(L)**: la probabilidad de tener una red de L enlaces exactamente

$$P(L) = \binom{N}{L} p^L (1-p)^{\frac{N(N-1)}{2} - L}$$

- El número promedio de links  $\langle L \rangle$  en un grafo aleatorio

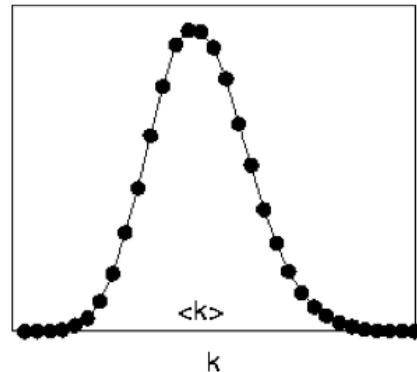
$$\langle L \rangle = \sum_{L=0}^{\frac{N(N-1)}{2}} LP(L) = p \frac{N(N-1)}{2} \longrightarrow \langle k \rangle = 2L/N = p(N-1)$$

- La varianza

$$\sigma^2 = p(1-p) \frac{N(N-1)}{2}$$

# Distribución de Grado

# DISTRIBUCION DE GRADO DE UNA RED ALEATORIA



Selecciona  $k$  nodos de un conjunto de  $N-1$  nodos

$$P(k) = \binom{N-1}{k} p^k (1-p)^{(N-1)-k}$$

Probabilidad de tener  $k$  enlaces

Probabilidad de perder  $N-1-k$  enlaces

$$\langle k \rangle = p(N-1)$$

$$\sigma_k^2 = p(1-p)(N-1)$$

$$\frac{\sigma_k}{\langle k \rangle} = \left[ \frac{1-p}{p} \frac{1}{(N-1)} \right]^{1/2} \approx \frac{1}{(N-1)^{1/2}}$$

A medida que la red aumenta de tamaño, la distribución se vuelve cada vez más estrecha: estamos cada vez más seguros de que el grado de un nodo se encuentra cerca de  $\langle k \rangle$  (valores homogéneos).

# DISTRIBUCION DE GRADO DE UNA RED ALEATORIA

$$P(k) = \binom{N-1}{k} p^k (1-p)^{(N-1)-k}$$
$$\langle k \rangle = p(N-1)$$
$$p = \frac{\langle k \rangle}{(N-1)}$$

Para  $N$  grandes y  $k$  pequeños, Podemos usar las siguientes aproximaciones ( $N \gg k$ ):

$$\binom{N-1}{k} = \frac{(N-1)!}{k!(N-1-k)!} = \frac{(N-1)(N-1-1)(N-1-2)\dots(N-1-k+1)(N-1-k)!}{k!(N-1-k)!} = \frac{(N-1)^k}{k!}$$

$$\ln[(1-p)^{(N-1)-k}] = (N-1-k) \ln\left(1 - \frac{\langle k \rangle}{N-1}\right) = -(N-1-k) \frac{\langle k \rangle}{N-1} = -\langle k \rangle \left(1 - \frac{k}{N-1}\right) \equiv -\langle k \rangle$$

$$(1-p)^{(N-1)-k} = e^{-\langle k \rangle}$$
$$\ln(1+x) = \sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n} x^n = x - \frac{x^2}{2} + \frac{x^3}{3} - \dots \quad \text{for } |x| \leq 1$$

$$P(k) = \binom{N-1}{k} p^k (1-p)^{(N-1)-k} = \frac{(N-1)^k}{k!} p^k e^{-\langle k \rangle} = \frac{(N-1)^k}{k!} \left(\frac{\langle k \rangle}{N-1}\right)^k e^{-\langle k \rangle} = e^{-\langle k \rangle} \frac{\langle k \rangle^k}{k!}$$

# DISTRIBUCIÓN DE GRADO DE POISSON

$$P(k) = \binom{N-1}{k} p^k (1-p)^{(N-1)-k}$$
$$\langle k \rangle = p(N-1)$$
$$p = \frac{\langle k \rangle}{(N-1)}$$

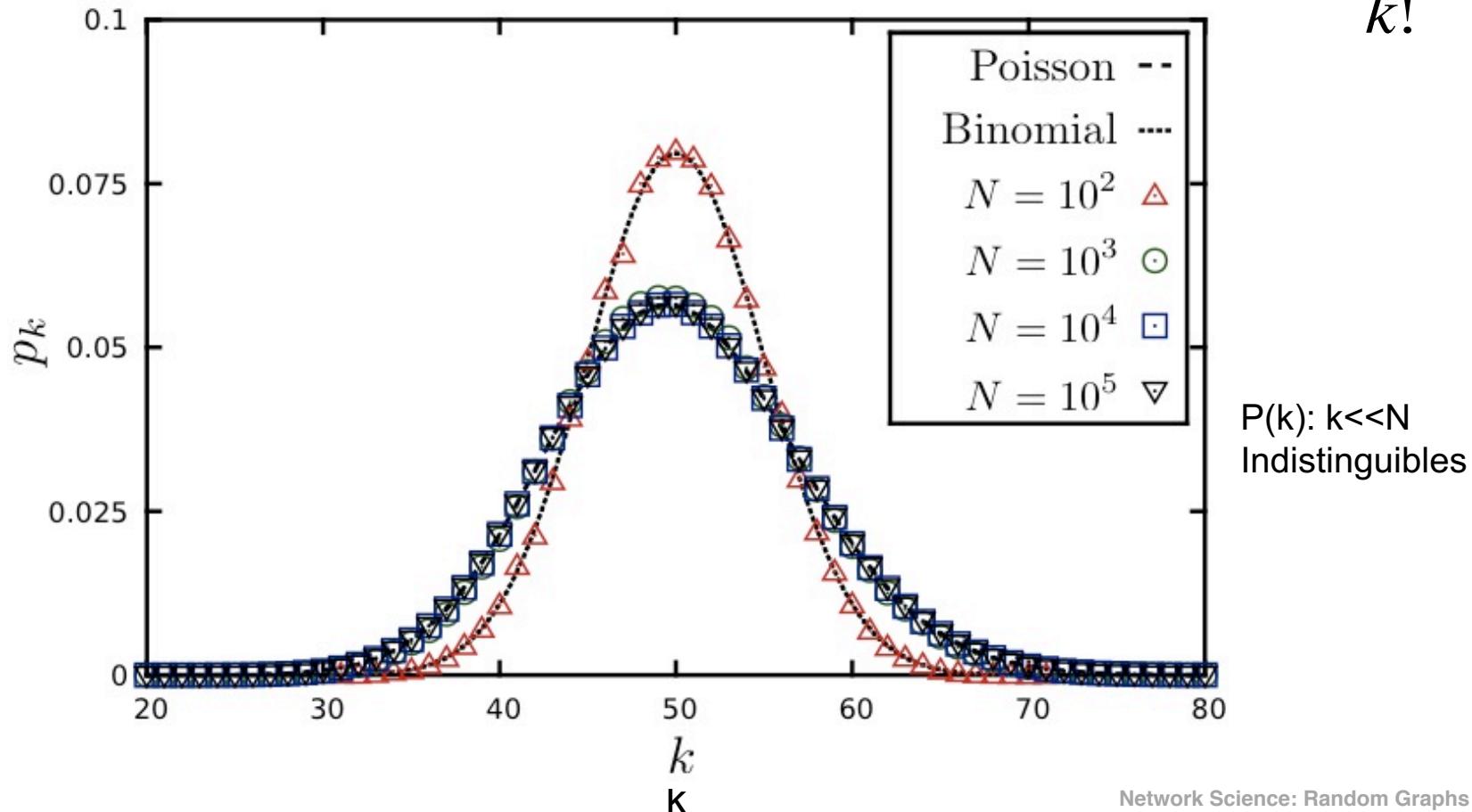
Para  $N$  grandes y  $k$  pequeños, llegamos a la distribución de Poisson:

$$P(k) = e^{-\langle k \rangle} \frac{\langle k \rangle^k}{k!}$$

# DISTRIBUCION DE GRADO DE UNA RED ALEATORIA

$$\langle k \rangle = 50$$

$$P(k) = e^{-\langle k \rangle} \frac{\langle k \rangle^k}{k!}$$



# DISTRIBUCION DE GRADO DE UNA RED ALEATORIA

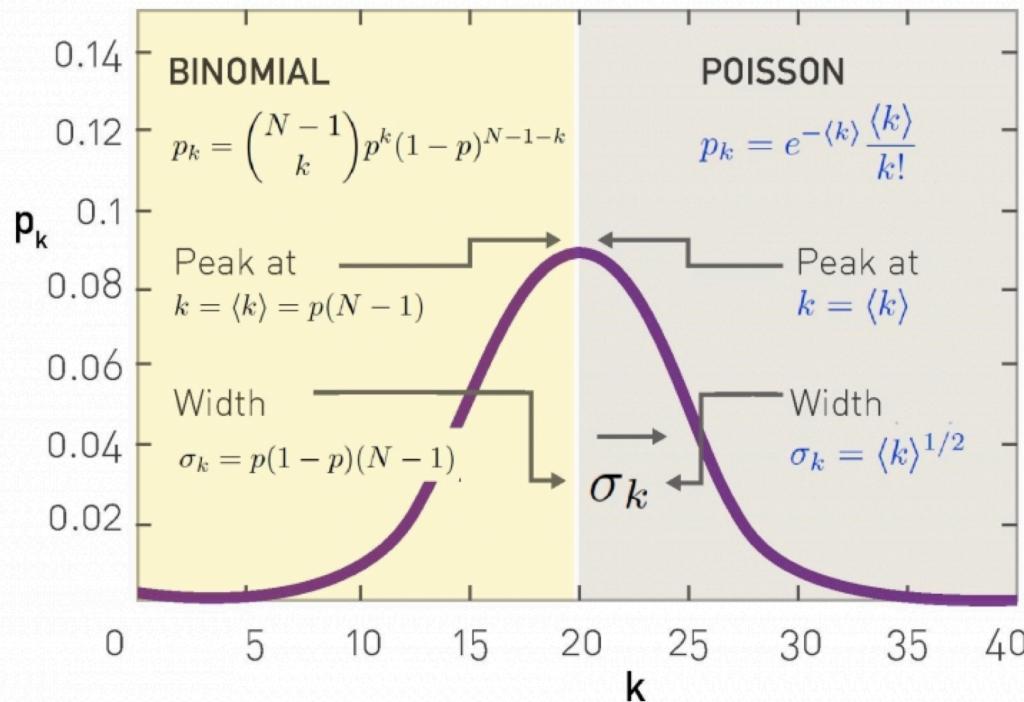
## Resultado exacto

- Distribución Binomial -

## Límite para N grande

- Distribución de Poisson -

Función de distribución de probabilidad  
(PDF)



Las redes reales **no** son Poisson

# NO HAY “OUTLIERS” EN UNA SOCIEDAD ALEATORIA

**Supongamos que una persona típica conoce aproximadamente 1,000 individuos por su nombre de pila**

$$P(k) = e^{-\langle k \rangle} \frac{\langle k \rangle^k}{k!}$$

→ El individuo más conectado tiene grado  $k_{max} \sim 1,185$

$$1 - P(k_{max}) = 1 - e^{-\langle k \rangle} \sum_{k=0}^{k_{max}} \frac{\langle k \rangle^k}{k!} = e^{-\langle k \rangle} \sum_{k=k_{max}+1}^{\infty} \frac{\langle k \rangle^k}{k!} \approx e^{-\langle k \rangle} \frac{\langle k \rangle^{k_{max}+1}}{(k_{max}+1)!},$$

→ El individuo menos conectado tiene grado  $k_{min} \sim 816$   $P(k_{min}) = e^{-\langle k \rangle} \sum_{k=0}^{k_{min}} \frac{\langle k \rangle^k}{k!}.$

**La probabilidad de encontrar un individuo con grado  $k > 2,000$  es  $10^{-27}$ .** Por lo tanto, la posibilidad de encontrar a un individuo con 2,000 conocidos es tan pequeña que tales nodos son virtualmente inexistentes en una sociedad aleatoria.

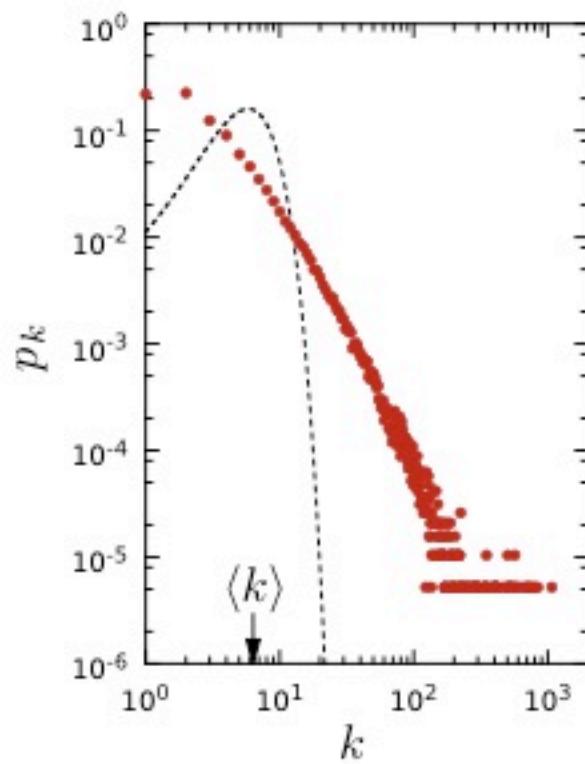
→ una sociedad aleatoria consistiría principalmente en individuos promedio, con todos con aproximadamente el mismo número de amigos.

→ Carecería de valores atípicos, individuos que son muy populares o solitarios.

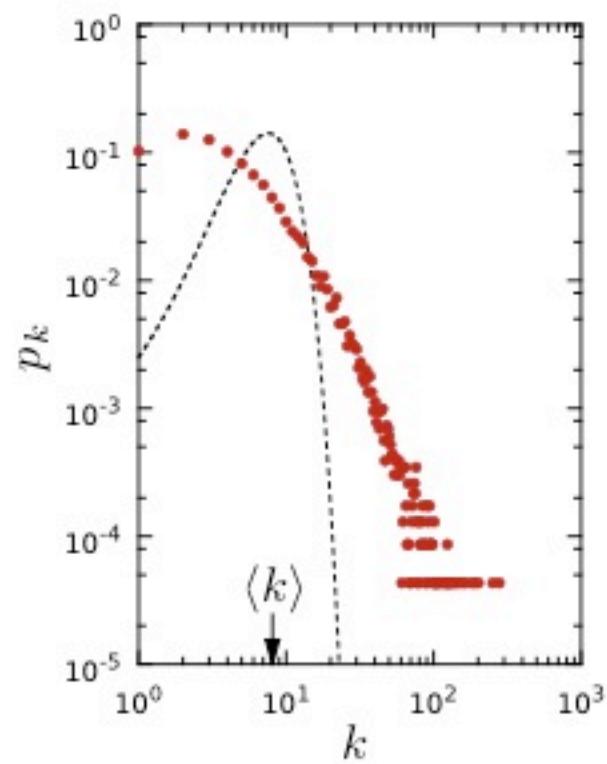
## ENFRENTANDO LA REALIDAD: Distribución en grado de redes reales.

$$P(k) = e^{-\langle k \rangle} \frac{\langle k \rangle^k}{k!}$$

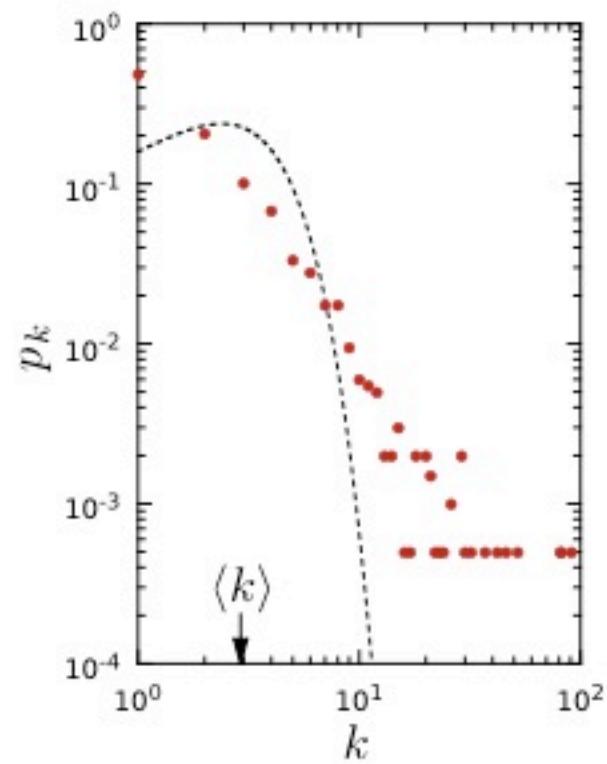
Internet



Science Collaboration



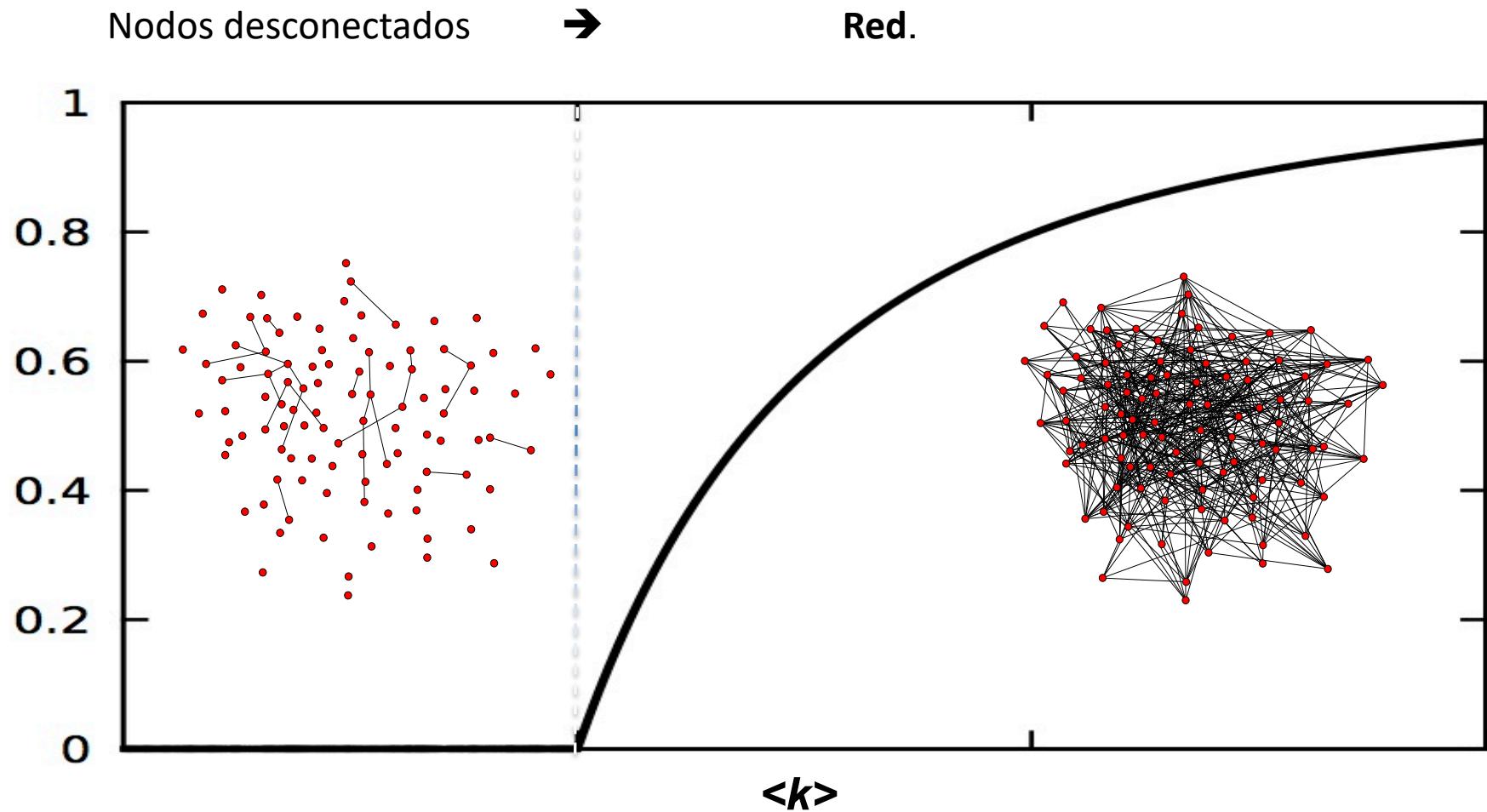
Protein Interactions

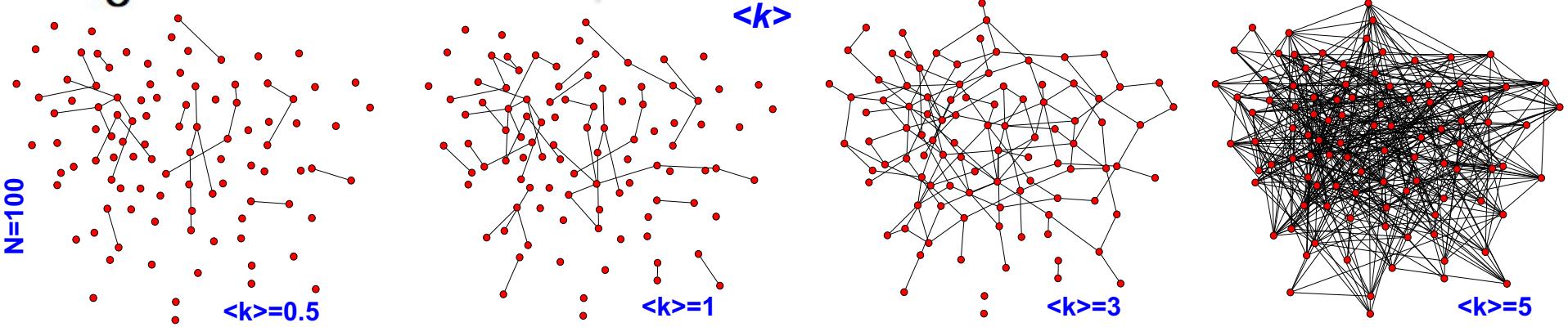
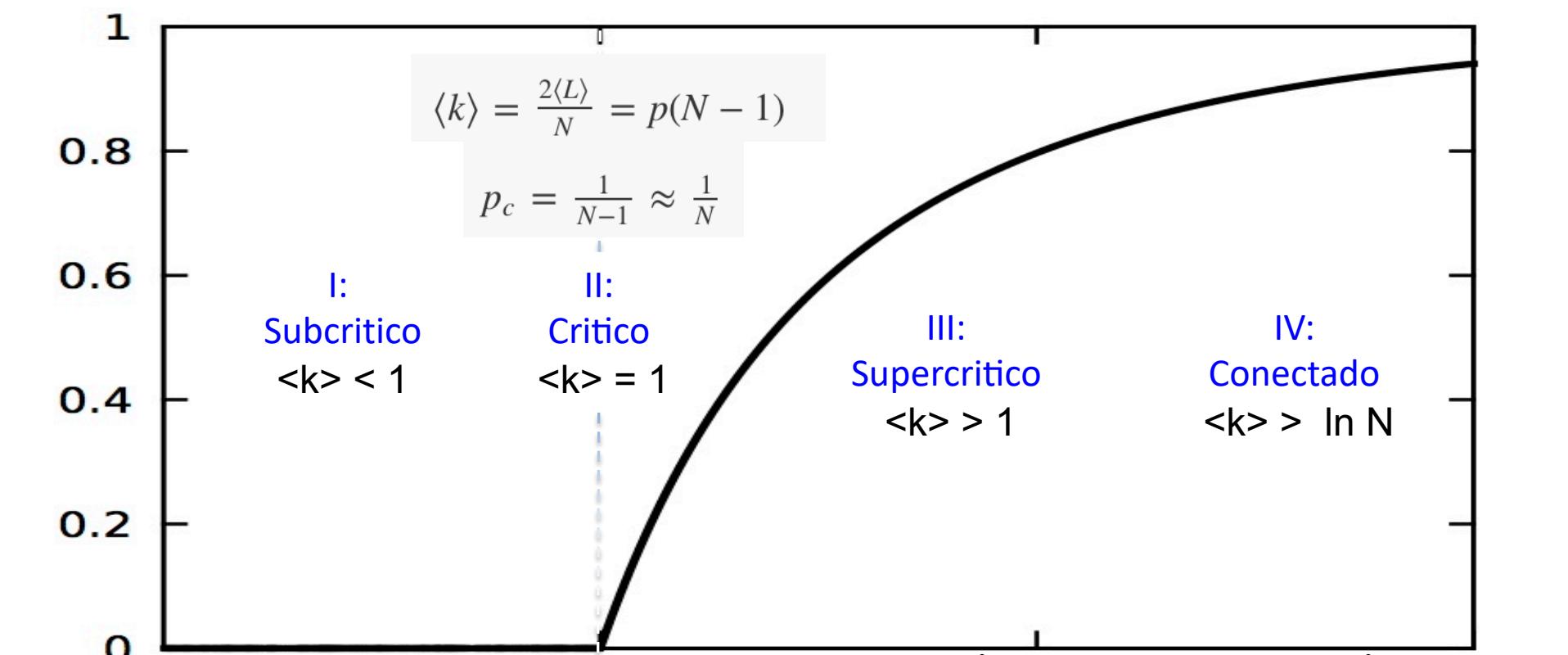


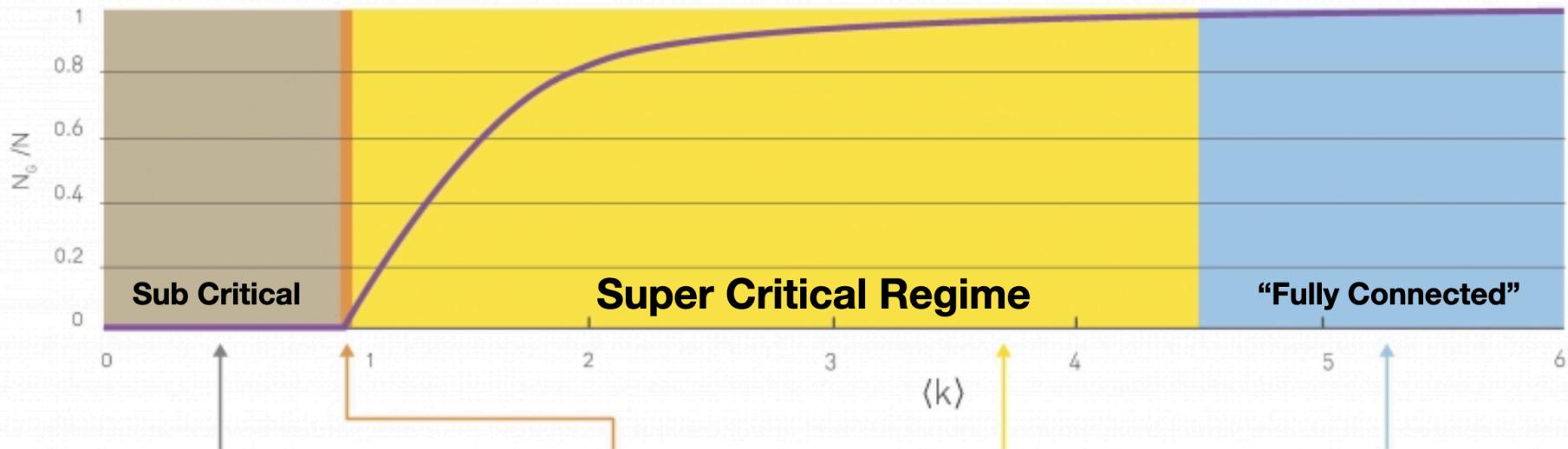
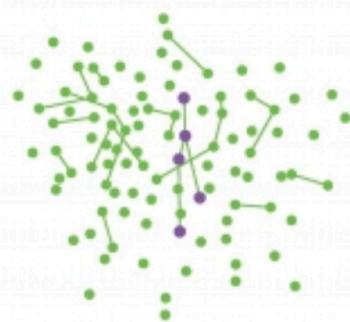
# La evolución de una red aleatoria.



## Evolución de una red aleatoria.



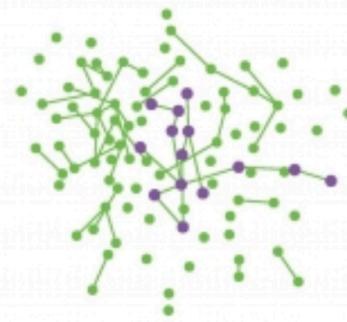


**a.****b.**

$$\langle k \rangle < 1$$

#### (b) Subcritical Regime

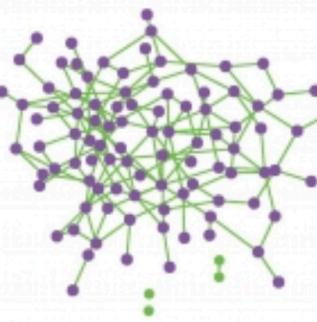
- \* No giant component
- \* Cluster size distribution:  $p_i \sim s^{-3/2} e^{-si}$
- \* Size of the largest cluster:  $N_G \sim \ln N$
- \* The clusters are trees

**c.**

$$\langle k \rangle = 1$$

#### (c) Critical Point

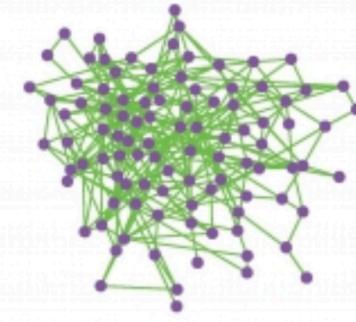
- \* No giant component
- \* Cluster size distribution:  $p_i \sim s^{-3/2}$
- \* Size of the largest cluster:  $N_G \sim N^{1/2}$
- \* The clusters may contain loops

**d.**

$$\langle k \rangle > 1$$

#### (d) Supercritical Regime

- \* Single giant component
- \* Cluster size distribution:  $p_i \sim s^{-3/2} e^{-si}$
- \* Size of the giant component:  $N_G \sim (p - p_c)N$
- \* The small clusters are trees
- \* Giant component has loops

**e.**

$$\langle k \rangle \geq \ln N$$

#### (e) Connected Regime

- \* Single giant component
- \* No isolated nodes or clusters
- \* Size of the giant component:  $N_G = N$
- \* Giant component has loops

# Evolución de una red aleatoria.

Nodos desconectados



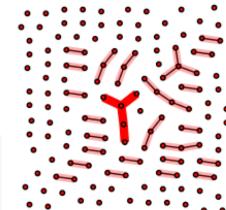
Red.

$$\langle k \rangle = \frac{2\langle L \rangle}{N} = p(N - 1)$$

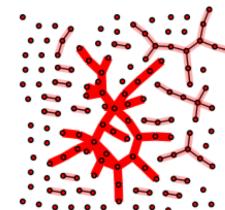
$$p_c = \frac{1}{N-1} \approx \frac{1}{N}$$

$\langle k_c \rangle = 1/N * (N-1) = 1$ , para  $N$  grande  
*(Erdos y Renyi, 1959)*

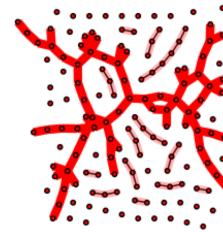
$p = 0.003$



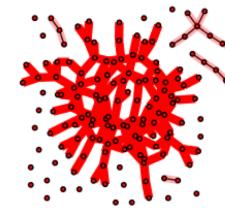
$p = 0.006$



$p = 0.008$



$p = 0.015$



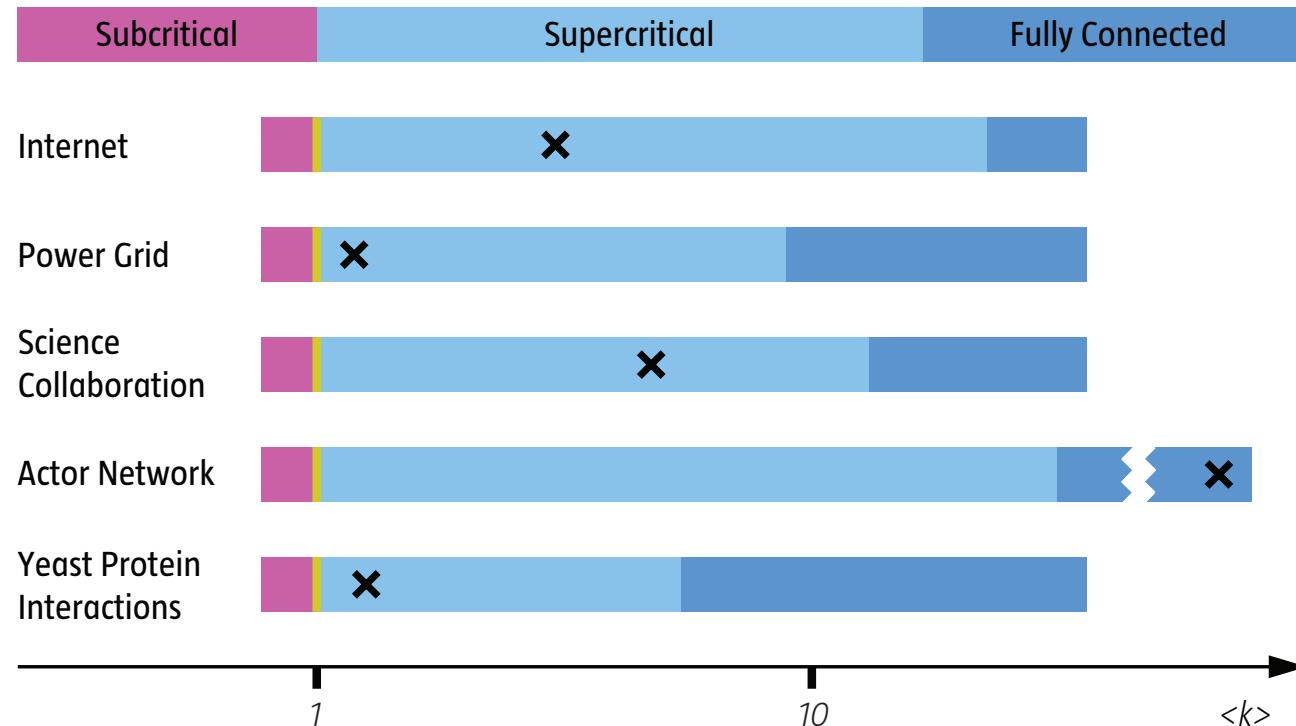
El hecho de que al menos un enlace por nodo sea **necesario** para tener un componente gigante no es inesperado. De hecho, para que exista un componente gigante, cada uno de sus nodos debe estar vinculado al menos a otro nodo.

Es algo inesperado, sin embargo, que un enlace es **suficiente** para el surgimiento de un componente gigante.

Es igualmente interesante que la aparición del componente gigante no sea gradual, pero sigue lo que los físicos llaman una **transición de fase** de segundo orden en  $\langle k \rangle = 1$ .

Las redes reales son supercríticas.

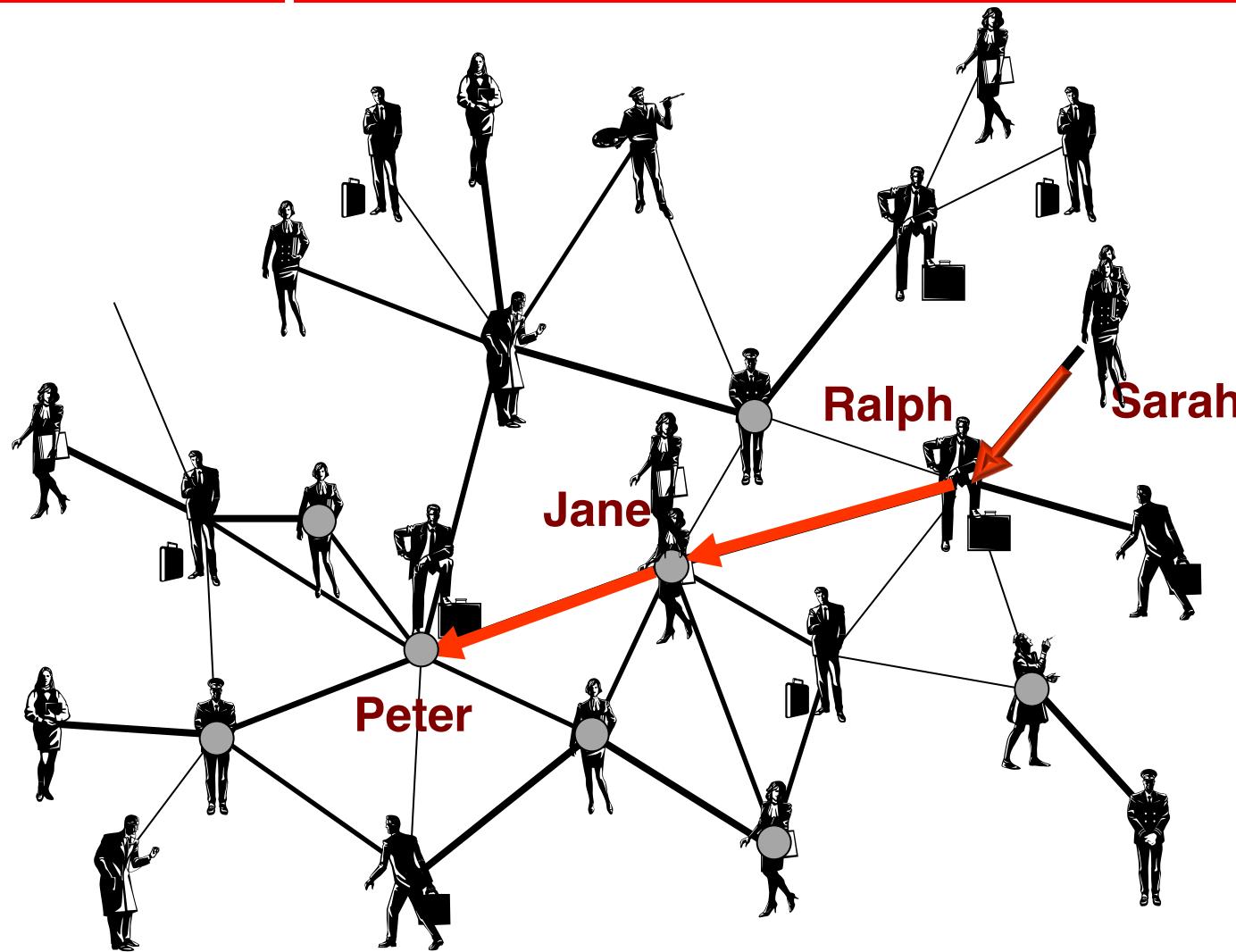
## Sección 7



Network	N	L	$\langle k \rangle$	$\ln N$
Internet	192,244	609,066	6.34	12.17
Power Grid	4,941	6,594	2.67	8.51
Science Collaboration	23,133	94,437	8.08	10.05
Actor Network	702,388	29,397,908	83.71	13.46
Protein Interactions	2,018	2,930	2.90	7.61

Red conectada  
 $\langle k \rangle > \ln N$

# Mundo Pequeño (Small World)



Frigyes Karinthy, 1929  
Stanley Milgram, 1967



Frigyes Karinthy (1887-1938)  
Hungarian Writer

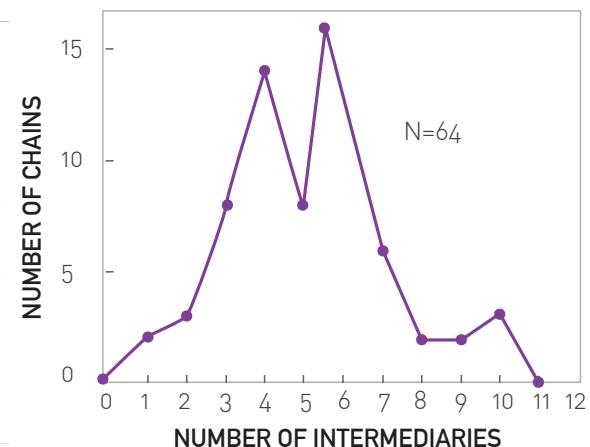
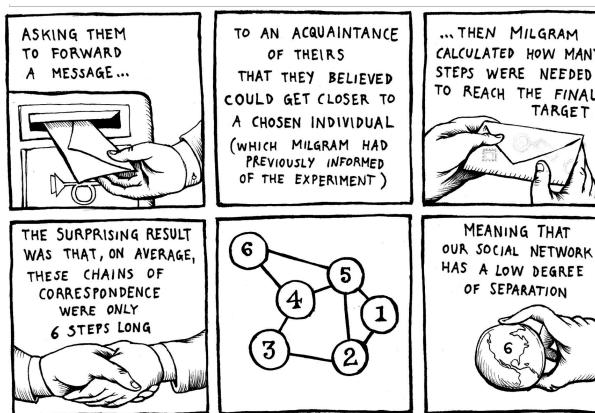
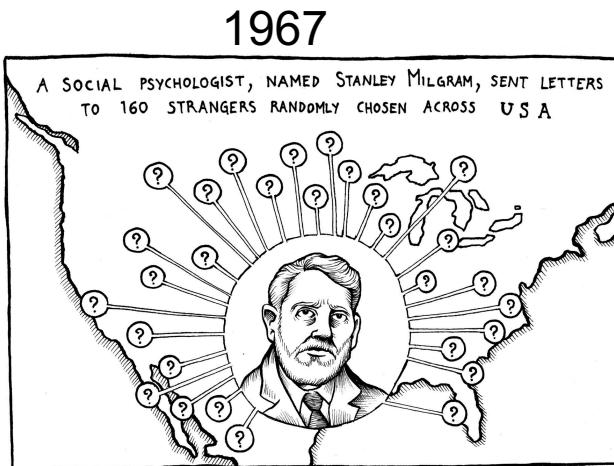
1929: *Minden másképpen van* (Everything is Different)  
*Láncszemek* (Chains)

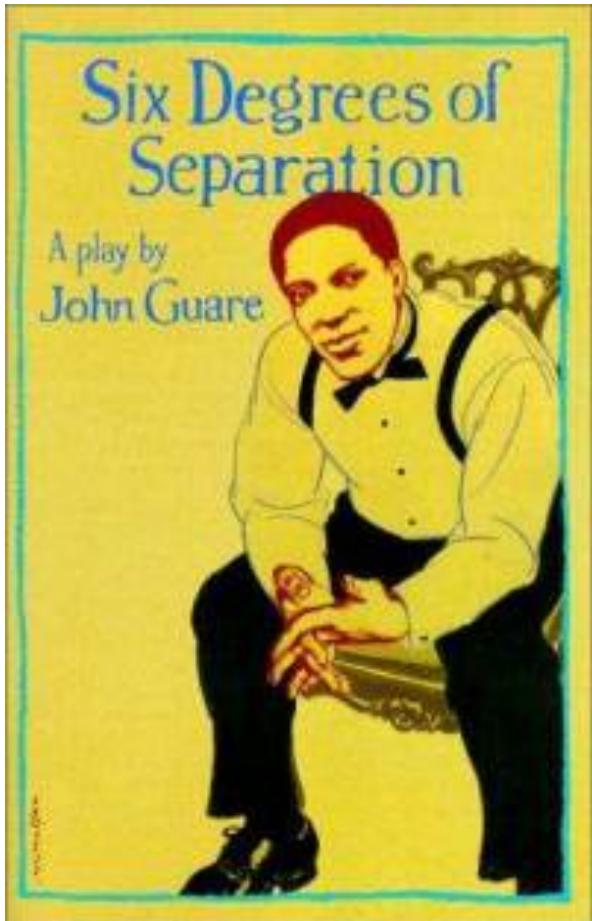
"Mire, Selma Lagerlöf acaba de ganar el Premio Nobel de Literatura, por lo que está obligada a conocer al Rey Gustav de Suecia, después de todo, él fue quien le entregó el Premio, como lo exige la tradición. El Rey Gustav, sin duda, es un apasionado jugador de tenis, que siempre participa en torneos internacionales. Se sabe que jugó con el Sr. Kehrling, a quien, por lo tanto, lo debe conocer con seguridad, y sucede que yo mismo conozco bastante bien al Sr. Kehrling".

"El trabajador conoce al gerente de la tienda, quién conoce a Ford; Ford tiene una relación amistosa con el director general de Hearst Publications, que el año pasado se hizo muy amigo de Arpad Pasztor, alguien que no solo conozco, sino que es un buen amigo mío. Así que fácilmente podría pedirle que envíe un telegrama a través del director general para que le diga a Ford que debe hablar con el gerente para que le informe al trabajador de la tienda que arme un auto rápidamente para mí, ya que necesito uno."

# The Small World Experiment

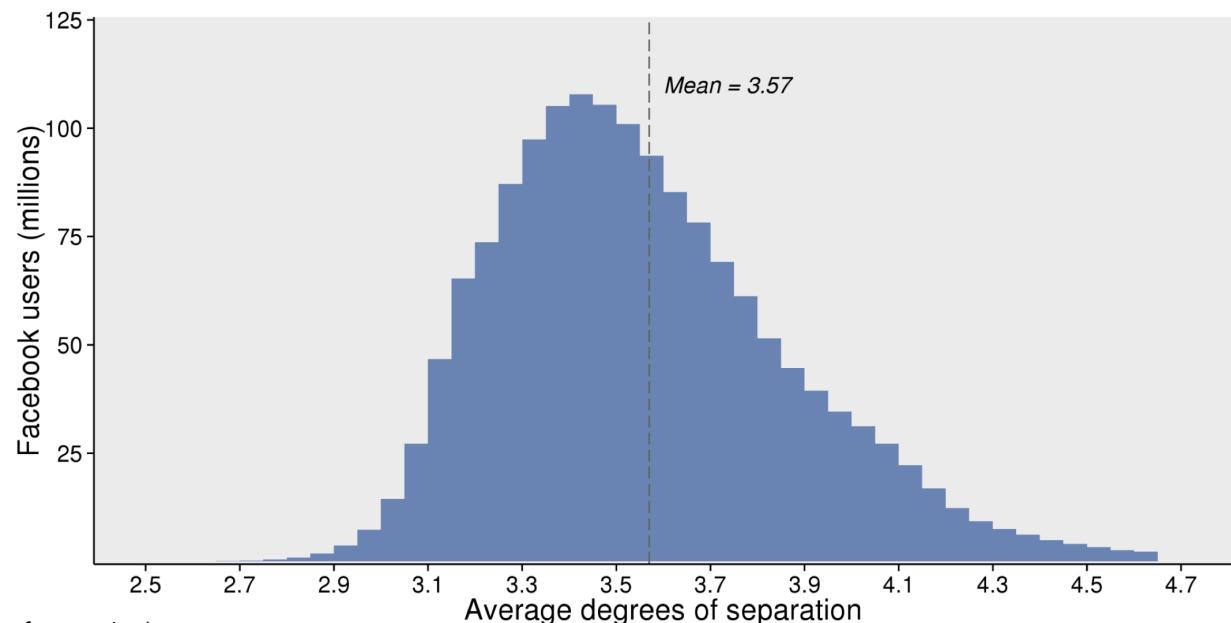
- Se entregaron cartas a personas en Nebraska para que las enviaran a una persona objetivo en Boston.
- Se instruyó a las personas para que pasaran las cartas a alguien que conocían por su nombre de pila.
- Las cartas que llegaron a destino siguieron caminos de una longitud de alrededor de 6 enlaces
- Seis grados de separación, ¿Cómo puede ser eso?



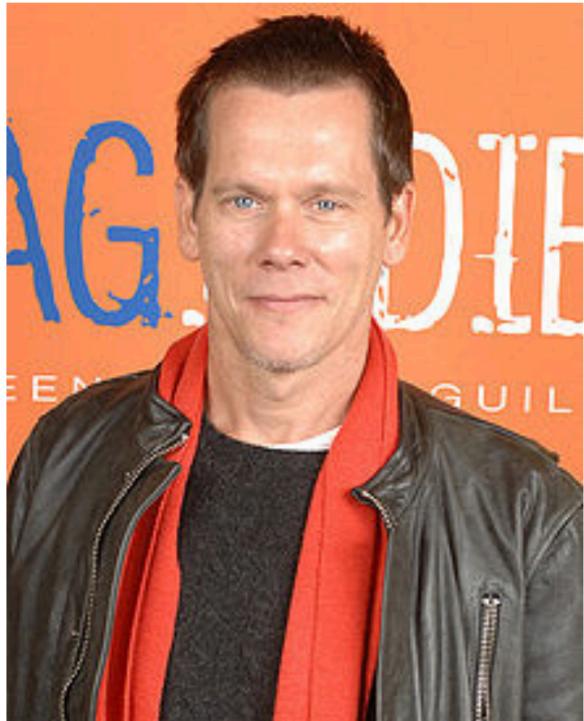


"Todos en este planeta están separados solo por otras seis personas. Seis grados de separación. Entre nosotros y todos los demás en este planeta. El presidente de los Estados Unidos. Un gondolero en Venecia ... No son solo los grandes nombres. Es cualquiera. Un nativo en una selva tropical. Un Tierra del Fuegano. Un esquimal. Estoy atado a todos en este planeta por un rastro de seis personas. Es un pensamiento profundo. Cómo cada persona es una puerta nueva, abriéndose a otros mundos".

- Informado en el blog de investigación de Facebook en 2016;
- La mayoría de las personas en Facebook tienen promedios de entre 2,9 y 4,2 grados de separación;
- La figura muestra los grados promedio estimados de separación entre todas las personas en Facebook. La persona promedio está conectada con todas las demás personas en un promedio de 3.57 pasos. La mayoría de personas tiene un promedio de entre 3 y 4 pasos;



# **The Six Degrees of Kevin Bacon**



**Paso 1. Nombra a un actor de cine**

**Paso 2. A través de los papeles en las películas, identifique una secuencia de actores que conecten al actor del Paso 1 con Kevin Bacon.**

**Paso 3. El ganador es la persona que encuentra el camino más corto**



Lorenza Izzo has a Bacon number of 2.

Find a di

Lorenza Izzo

was in

Once Upon a Time in Hollywood

with

Brad Pitt

was in

Moneyball

with

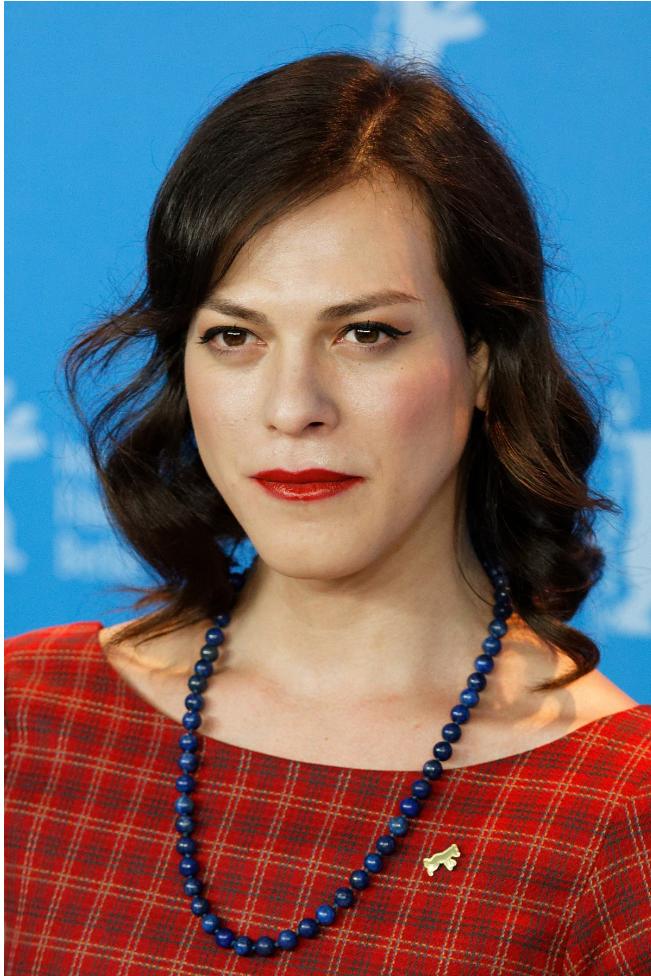
Kevin Bacon

Kevin Bacon

to Lorenza Izzo

Find link

More o



daniela vega has a Bacon number of 3.

Daniela Vega

was in

A Fantastic Woman

with

Alejandro Goic

was in

The 33

with

Bob Gunton

was in

JFK

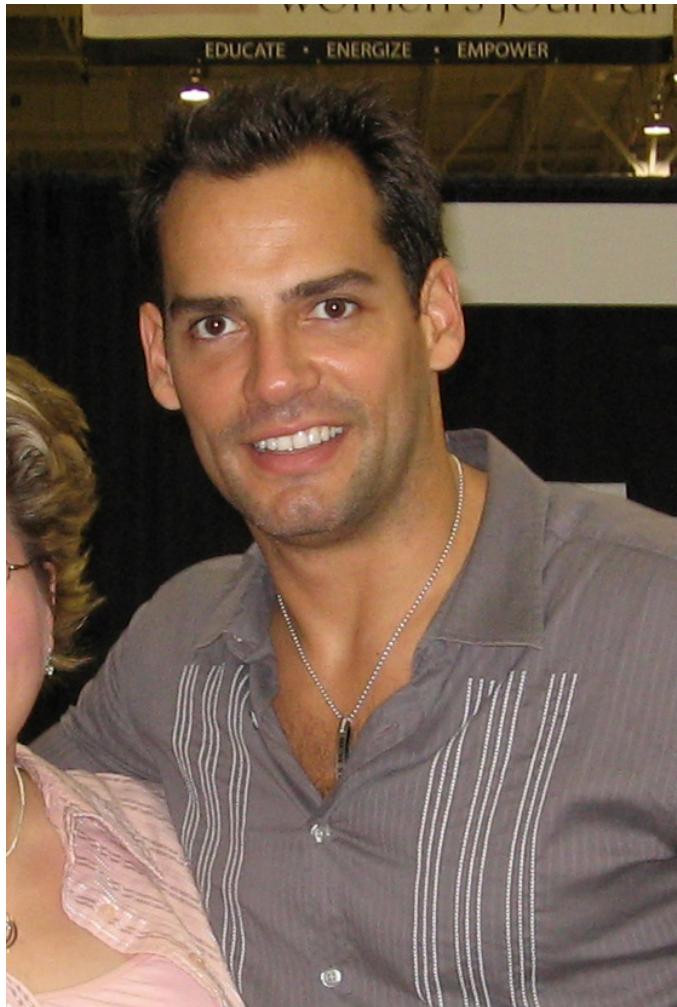
with

Kevin Bacon

Kevin Bacon

to daniela vega

Find link



cristian de la fuente has a Bacon number of 2.

Cristian de la Fuente

was in

Driven

with

Burt Reynolds

was in

Starting Over

with

Kevin Bacon

Kevin Bacon

to cristian de la fuente

Find link

More options



Bryan Cranston has a Bacon number of 2.

Bryan Cranston

was in

Little Miss Sunshine

with

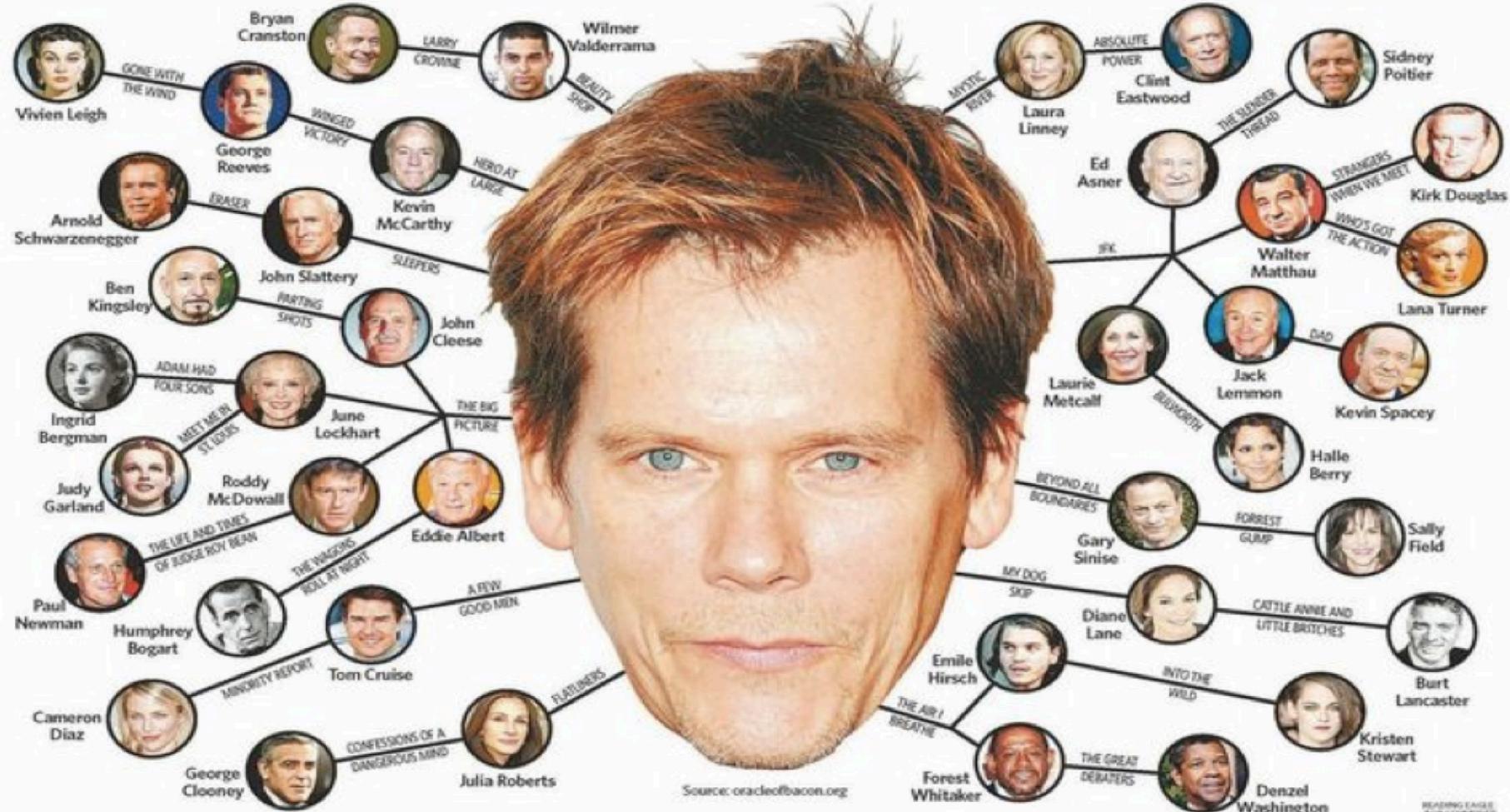
Steve Carell

was in

Crazy, Stupid, Love

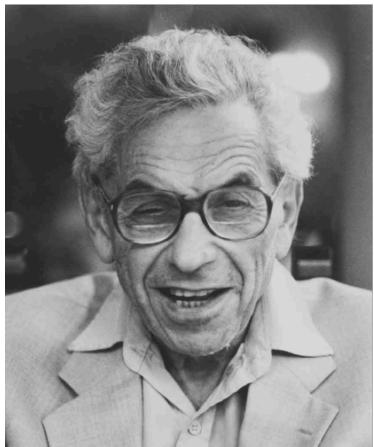
with

Kevin Bacon

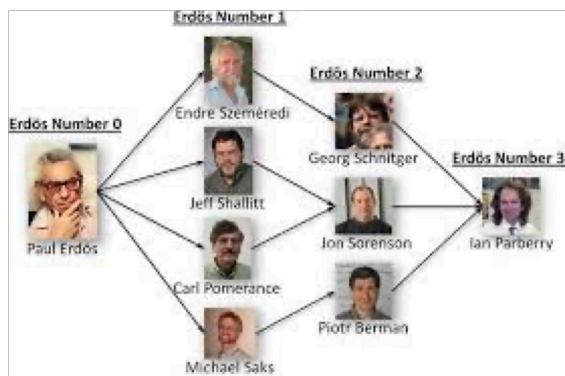


<https://oracleofbacon.org>

# Erdos Number

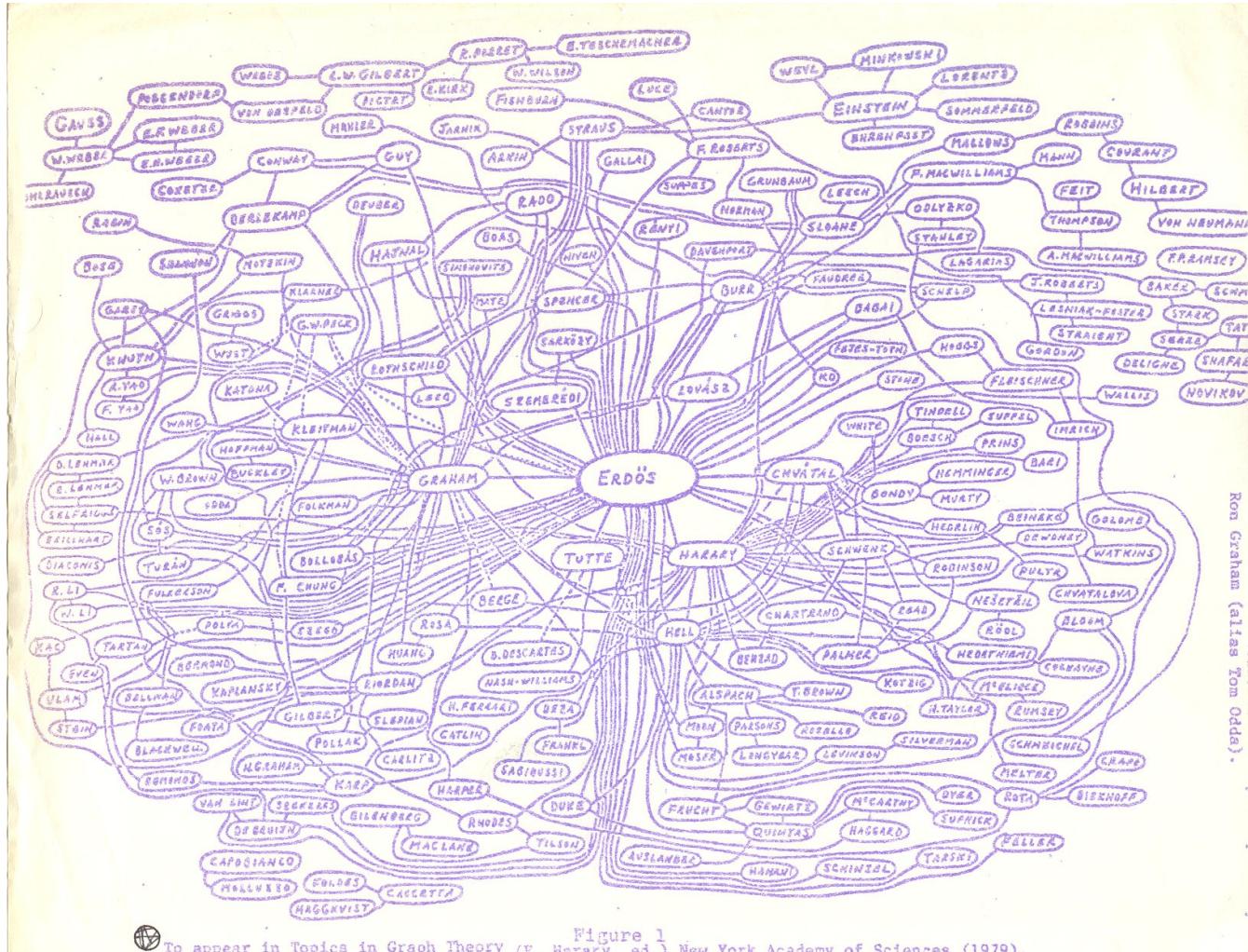


Paul Erdős



Erdős number	0	---	1 person
Erdős number	1	---	504 people
Erdős number	2	---	6593 people
Erdős number	3	---	33605 people
Erdős number	4	---	83642 people
Erdős number	5	---	87760 people
Erdős number	6	---	40014 people
Erdős number	7	---	11591 people
Erdős number	8	---	3146 people
Erdős number	9	---	819 people
Erdős number	10	---	244 people
Erdős number	11	---	68 people
Erdős number	12	---	23 people
Erdős number	13	---	5 people

# DATOS DE RED: REDES DE COLABORACIÓN DE CIENCIAS



Erdos:  
1,400 papers  
507 coauthors

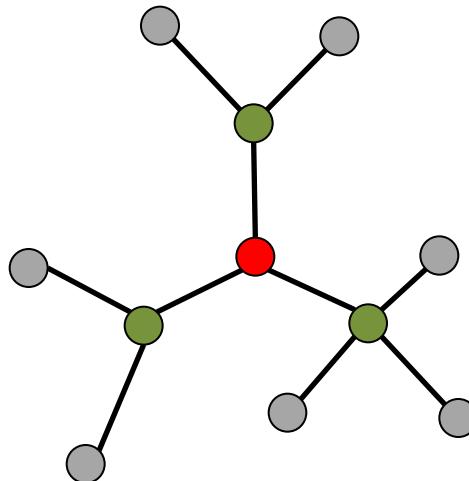
Einstein: EN=2  
Paul Samuelson EN=5

ALB: EN= 3  
CCV: EN=4



# DISTANCIA EN GRAFOS ALEATORIOS

Los grafos aleatorios tienden a tener una topología en forma de árbol con grados de nodo casi constantes.



$\langle k \rangle$  nodes at distance one ( $d=1$ ).

$\langle k \rangle^2$  nodes at distance two ( $d=2$ ).

$\langle k \rangle^3$  nodes at distance three ( $d =3$ ).

...

$\langle k \rangle^d$  nodes at distance  $d$ .

Podemos escribir la cantidad total de nodos de la red como:

$$N = 1 + \langle k \rangle + \langle k \rangle^2 + \dots + \langle k \rangle^{d_{\max}} = \frac{\langle k \rangle^{d_{\max}+1} - 1}{\langle k \rangle - 1} \approx \langle k \rangle^{d_{\max}}$$

$$d_{\max} = \frac{\log N}{\log \langle k \rangle}$$

# DISTANCIA EN GRAFOS ALEATORIOS

$$d_{\max} = \frac{\log N}{\log \langle k \rangle}$$

En la mayoría de las redes, esto ofrece una mejor aproximación a la distancia promedio entre dos nodos elegidos al azar,  $\langle d \rangle$ , que al diámetro  $d_{\max}$ .

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

Llamaremos a este fenómeno de mundo pequeño la propiedad de que la **longitud promedio de la trayectoria o el diámetro depende logarítmicamente del tamaño del sistema**. Por lo tanto, "pequeño" significa que  $\langle d \rangle$  es proporcional al logaritmo de  $N$ , en lugar de a  $N$ .

El término  $1/\log \langle k \rangle$  implica cuanto más densa sea la red, menor será la distancia entre los nodos.

# DISTANCIAS EN GRÁFOS ALEATORIOS

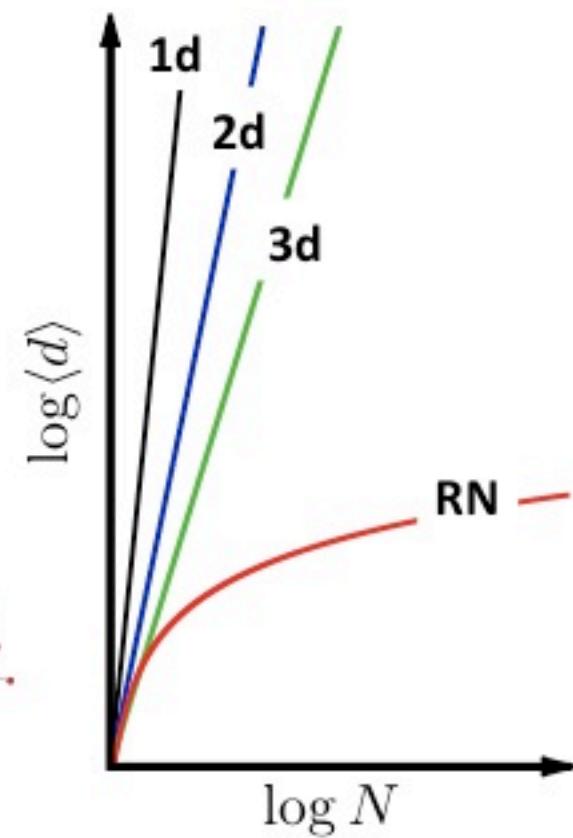
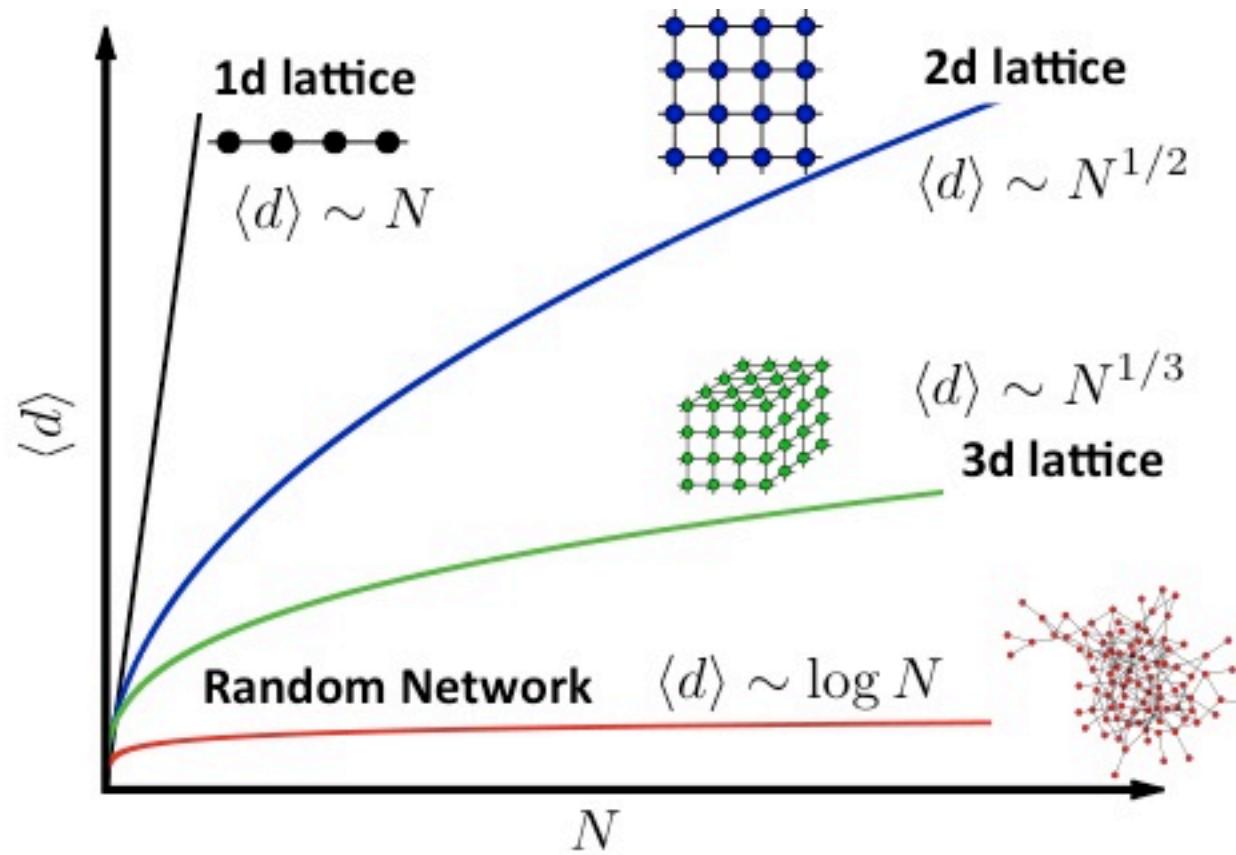
comparados con datos reales

Network	N	L	$\langle k \rangle$	$\langle d \rangle$	$d_{max}$	$\ln N / \ln \langle k \rangle$
Internet	192,244	609,066	6.34	6.98	26	6.58
WWW	325,729	1,497,134	4.60	11.27	93	8.31
Power Grid	4,941	6,594	2.67	18.99	46	8.66
Mobile-Phone Calls	36,595	91,826	2.51	11.72	39	11.42
Email	57,194	103,731	1.81	5.88	18	18.4
Science Collaboration	23,133	93,437	8.08	5.35	15	4.81
Actor Network	702,388	29,397,908	83.71	3.91	14	3.04
Citation Network	449,673	4,707,958	10.43	11.21	42	5.55
E. Coli Metabolism	1,039	5,802	5.58	2.98	8	4.04
Protein Interactions	2,018	2,930	2.90	5.61	14	7.14

T-1-1-1-1-1

Dadas las enormes diferencias en diámetro, tamaño y grado promedio, la aproximación es excelente.

¿Por qué sorprenden los pequeños mundos? Sorprendente en comparación con qué?



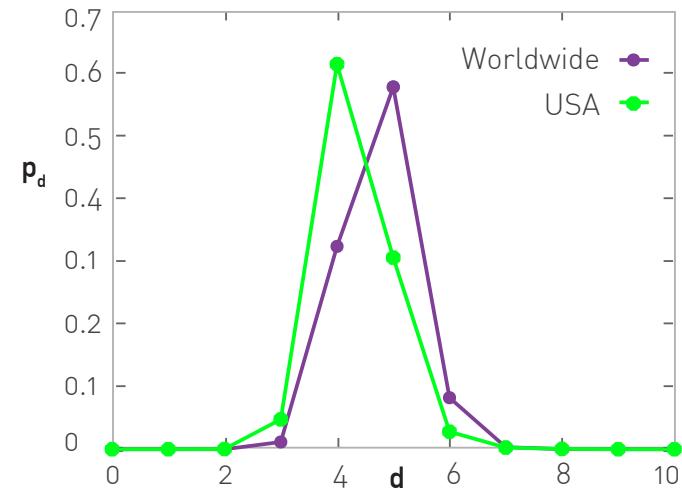
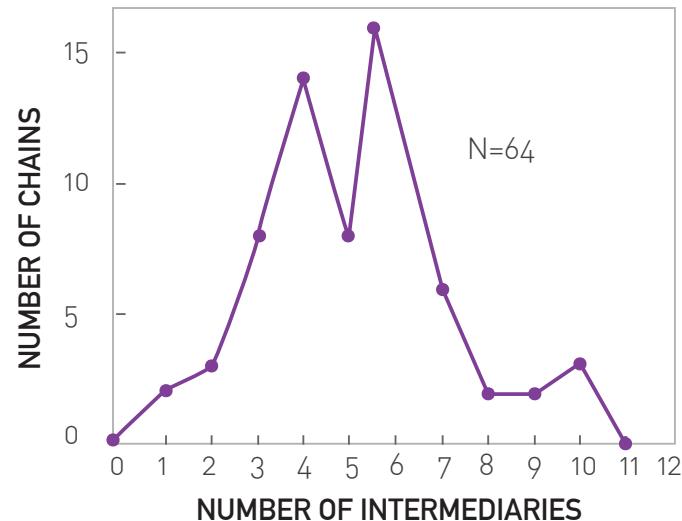
# Tres, cuatro o seis grados?

Para las redes sociales del mundo:

$$\langle k \rangle \approx 10^3$$

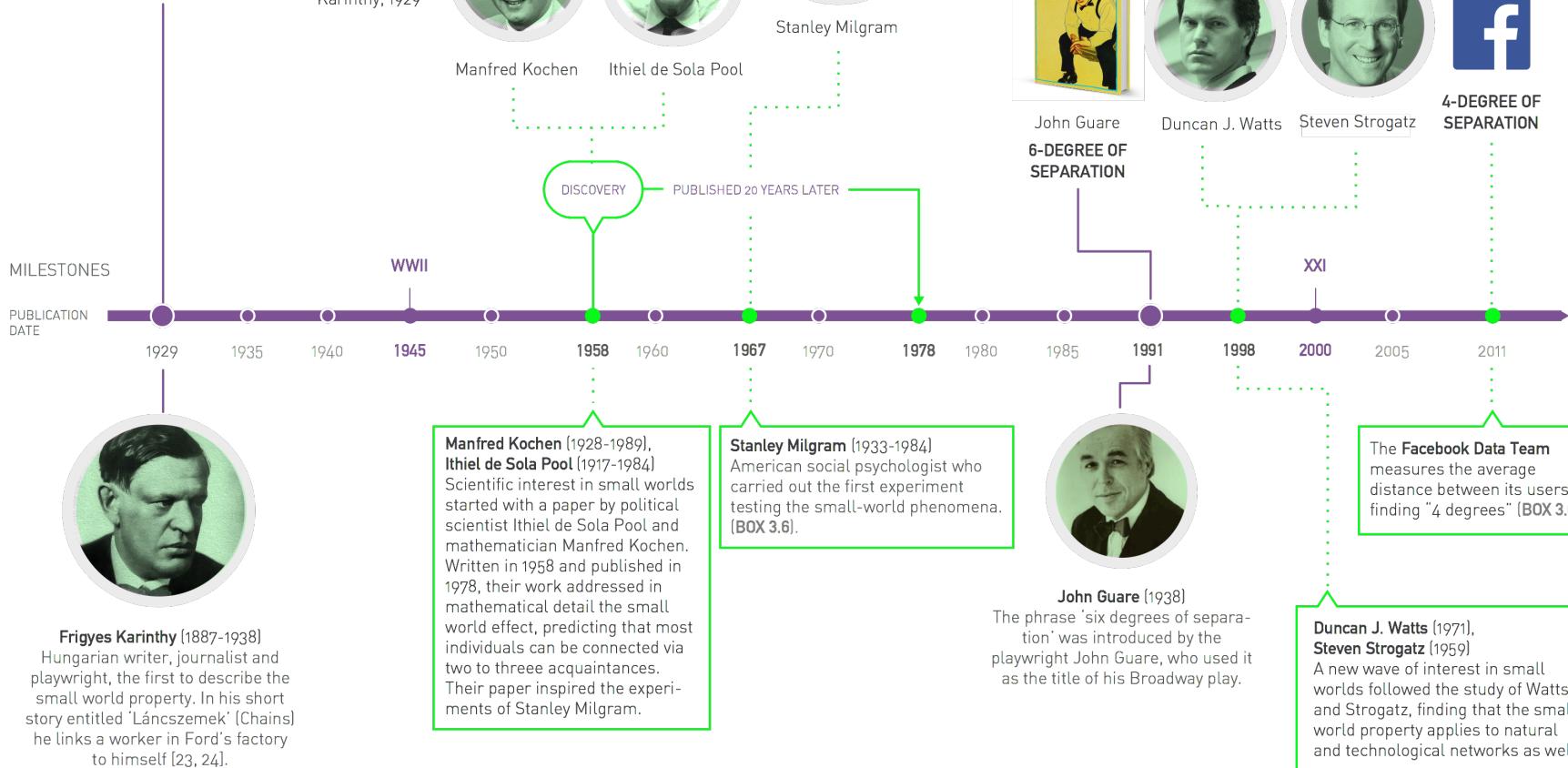
$N \approx 7 \times 10^9$  para la población mundial.

$$\langle d \rangle = \frac{\ln(N)}{\ln \langle k \rangle} = 3.28$$



"The worker knows the manager in the shop, who knows Ford; Ford is on friendly terms with the general director of Hearst Publications, who last year became good friends with Árpád Pásztor, someone I not only know, but to the best of my knowledge a good friend of mine."

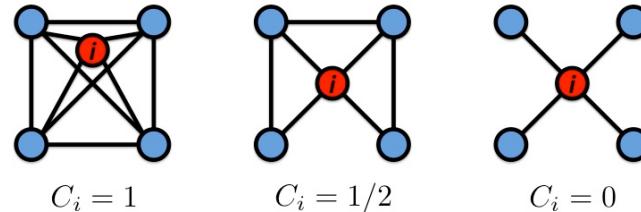
Karinthy, 1929



# Coeficiente de Agrupamiento (Clustering)

# COEFICIENTE DE AGRUPAMIENTO

$$C_i \equiv \frac{2 \langle L_i \rangle}{k_i(k_i - 1)}$$



Dado que los enlaces son independientes y tienen la misma probabilidad  $p$ ,

$$\langle L_i \rangle \approx p \frac{k_i(k_i - 1)}{2}$$

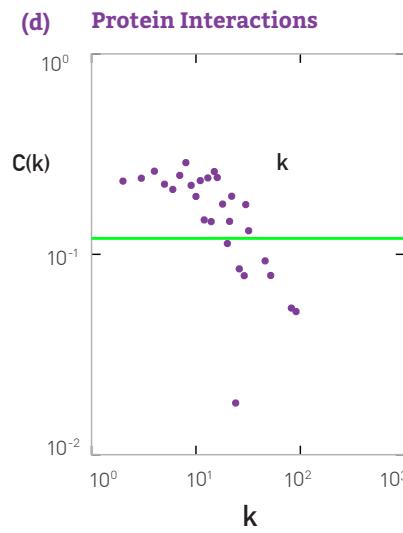
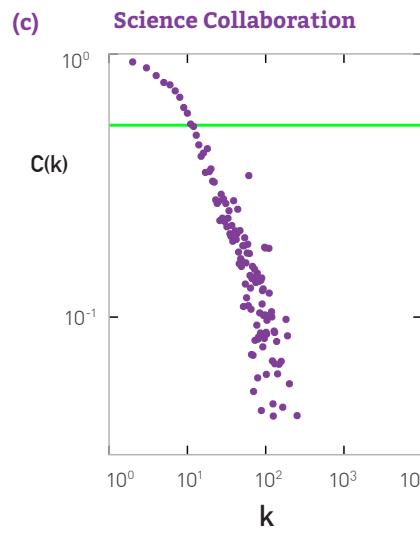
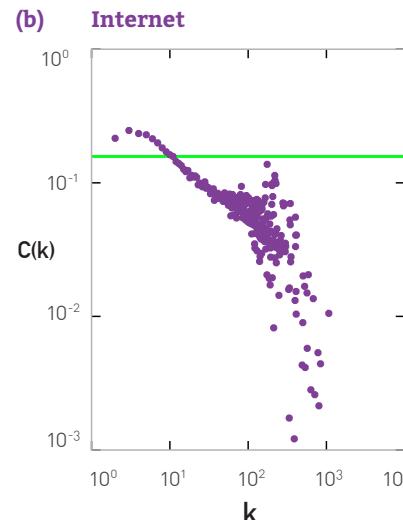
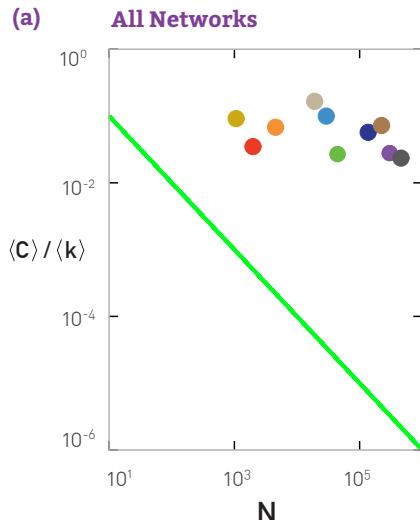


$$C_i = \frac{2\langle L_i \rangle}{k_i(k_i - 1)} = p = \frac{\langle k \rangle}{N}.$$

$k_i$ =número de vecinos del nodo i  
 $L_i$ =Número de enlaces que pasan por el nodo i

- El coeficiente de agrupamiento de los grafos aleatorios es pequeño.
- Para grados fijos,  $C$  disminuye con el tamaño del sistema  $N$ .
- $C$  es independiente del grado  $k$  de nodos individuales.

# COEFICIENTE DE AGRUPAMIENTO



$$C_i = \frac{2\langle L_i \rangle}{k_i(k_i - 1)} = p = \frac{\langle k \rangle}{N}.$$

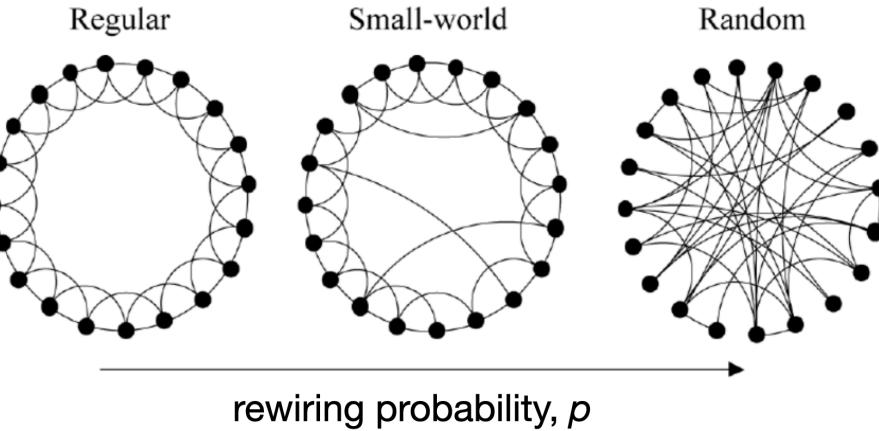
C disminuye con el tamaño del sistema N.

C es independiente del grado k de los nodos individuales.

# Small World Networks

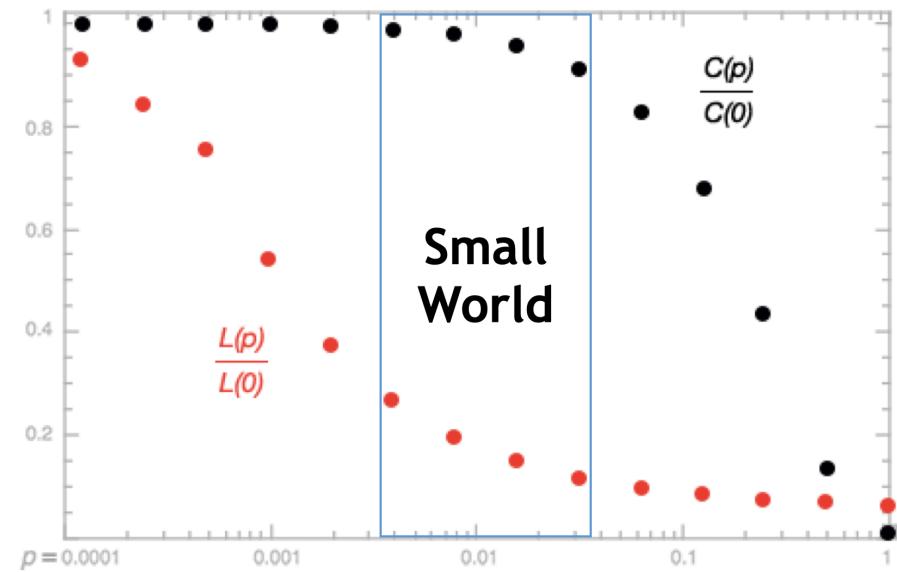


Duncan J Watts    Steven Strogatz



## Watts & Strogatz algorithm

A partir de una red de anillo regular, destruya cada enlace con probabilidad  $p$ , creando un nuevo enlace a un nodo aleatorio.



Las redes reales no son aleatorias.

# ¿Las redes reales son como los grafos aleatorios?

A medida que se dispone de datos cuantitativos sobre redes reales, podemos comparar su topología con las predicciones de la teoría de grafos aleatorios.

Tenga en cuenta que una vez que tenemos  $N$  y  $\langle k \rangle$  para una red aleatoria, de ella podemos derivar cada propiedad medible. De hecho, tenemos:

Longitud media del camino:

$$\langle l_{rand} \rangle \approx \frac{\log N}{\log \langle k \rangle}$$

Coeficiente de agrupamiento:

$$C_i = \frac{2\langle L_i \rangle}{k_i(k_i - 1)} = p = \frac{\langle k \rangle}{N}.$$

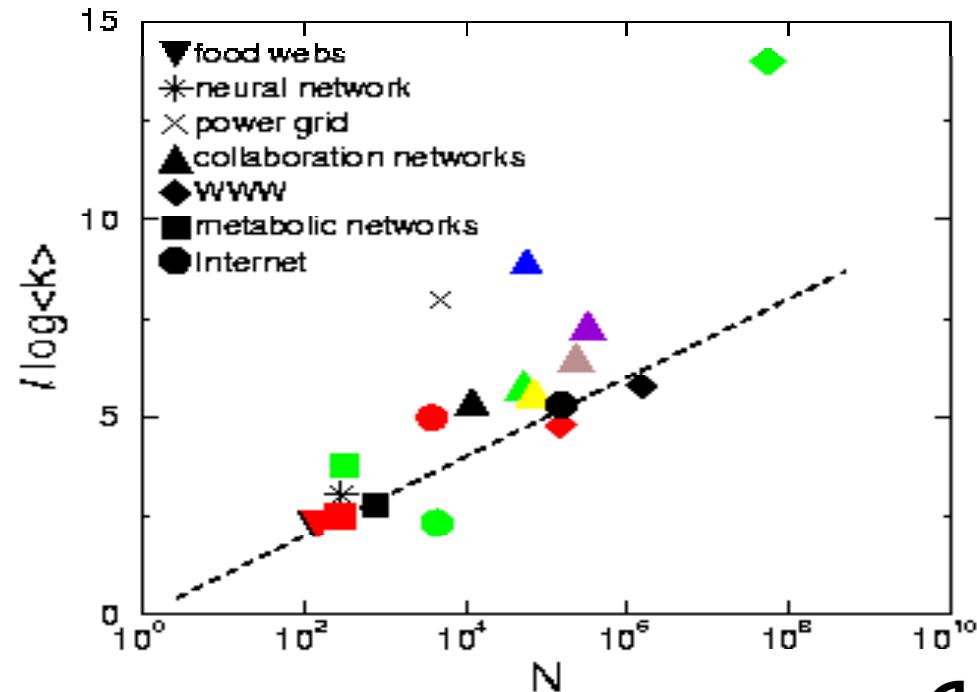
Distribución de Grados:

$$P(k) = e^{-\langle k \rangle} \frac{\langle k \rangle^k}{k!}$$

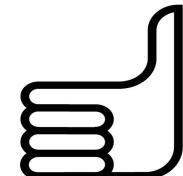
# LONGITUD DE CAMINOS EN REDES REALES

## Predicción:

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



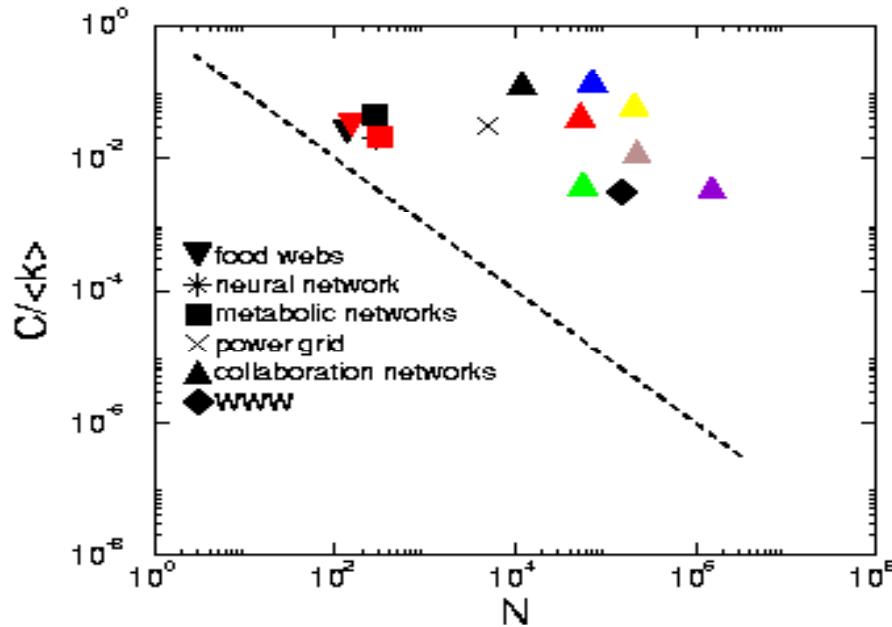
Las redes reales tienen distancias cortas como los grafos aleatorios.



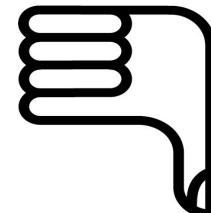
# COEFICIENTE DE CLUSTERING

## Predicción:

$$C_i = \frac{2\langle L_i \rangle}{k_i(k_i - 1)} = p = \frac{\langle k \rangle}{N}.$$



$C_{rand}$  Subestima en órdenes de magnitudes el coeficiente de agrupamiento de las redes reales.



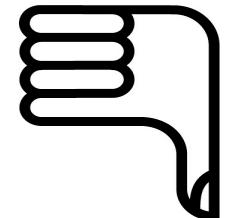
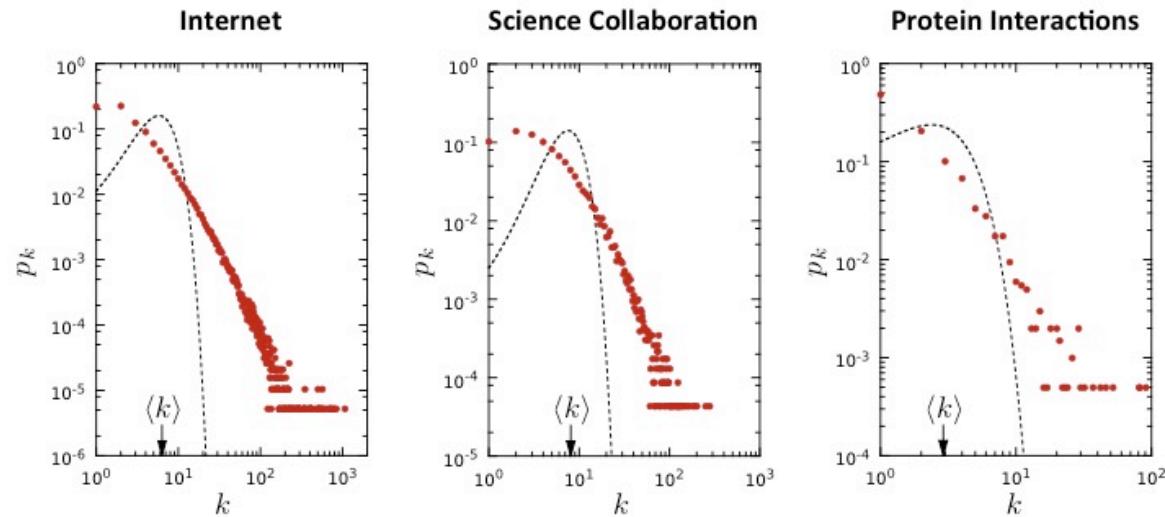
# LA DISTRIBUCIÓN DE GRADO

## Predicción:

$$P(k) = e^{-\langle k \rangle} \frac{\langle k \rangle^k}{k!}$$

## Data:

$$P(k) \approx k^{-\gamma}$$



# ¿Las redes reales son como los grafos aleatorios?

A medida que se dispone de datos cuantitativos sobre redes reales, podemos comparar su topología con las predicciones de la teoría de grafos aleatorios.

Tenga en cuenta que una vez que tenemos  $N$  y  $\langle k \rangle$  para una red aleatoria, de ella podemos derivar cada propiedad medible. De hecho, tenemos:

Longitud media del camino:

$$\langle l_{rand} \rangle \approx \frac{\log N}{\log \langle k \rangle}$$



Coeficiente de agrupamiento:

$$C_i = \frac{2\langle L_i \rangle}{k_i(k_i - 1)} = p = \frac{\langle k \rangle}{N}.$$



Distribución de Grados:

$$P(k) = e^{-\langle k \rangle} \frac{\langle k \rangle^k}{k!}$$

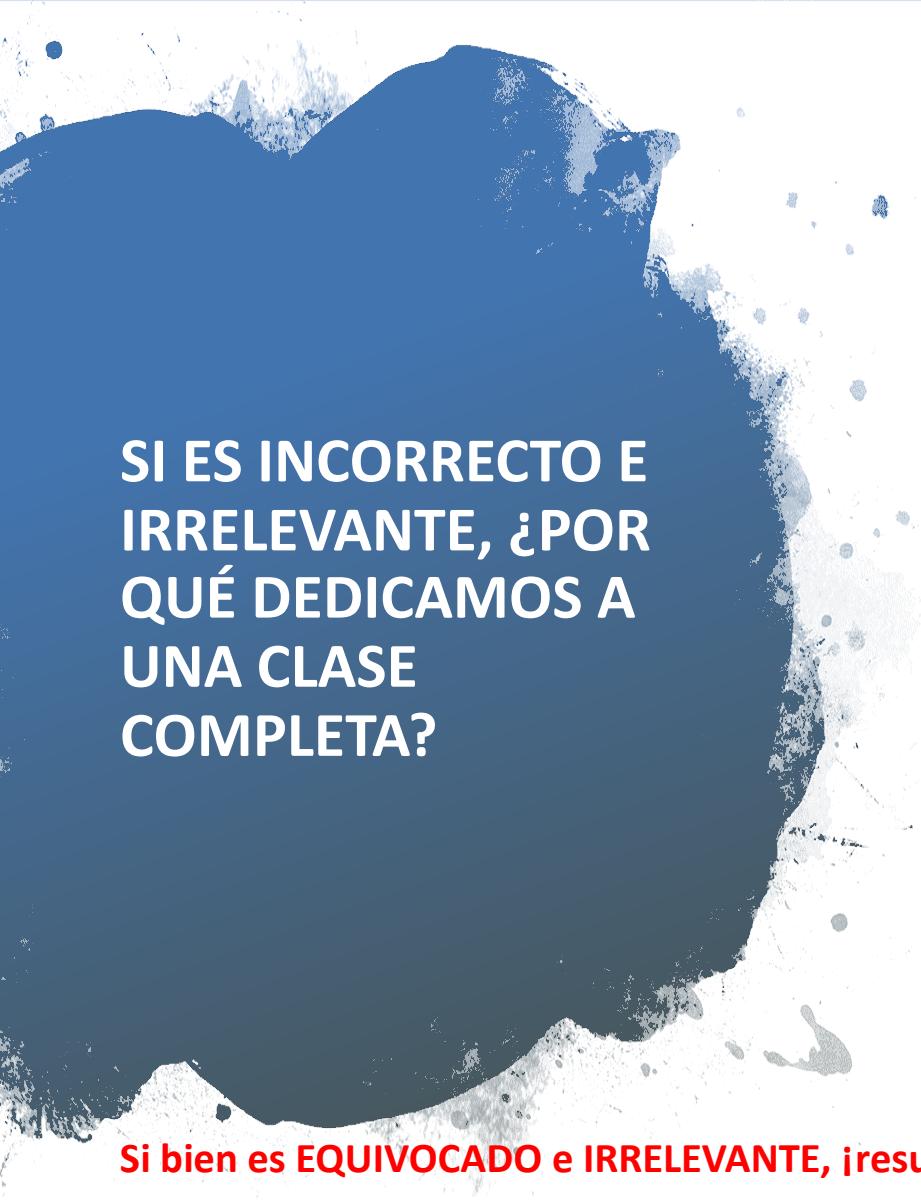


## ¿ES EL MODELO ALEATORIO RELEVANTE PARA LOS SISTEMAS REALES?

(B) Lo más importante: debemos preguntarnos, ¿son las redes reales aleatorias?

La respuesta simplemente es no

**No hay ninguna red en la naturaleza que sepamos que sería descrita por el modelo de red aleatorio.**



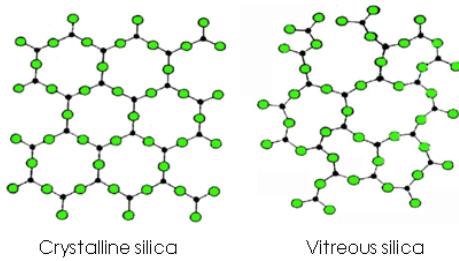
# SI ES INCORRECTO E IRRELEVANTE, ¿POR QUÉ DEDICAMOS A UNA CLASE COMPLETA?

- Es el modelo de referencia para el resto de la clase.
- Nos ayudará a calcular muchas cantidades, que luego se pueden comparar con los datos reales, entendiendo hasta qué punto es una propiedad particular el resultado de algún proceso aleatorio.
- Patrones en redes reales, que son compartidos por una gran cantidad de redes reales, pero que se desvían de las predicciones del modelo de red aleatorio.
- Para identificarlos, debemos entender cómo se vería una propiedad en particular si está totalmente controlada por procesos aleatorios.

Si bien es EQUIVOCADO e IRRELEVANTE, ¡resultará extremadamente ÚTIL!

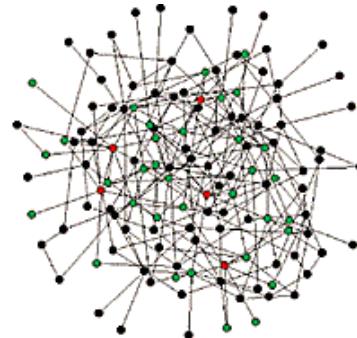
# Scale-free Networks

# Desde sistemas físicos a sociales



Crystalline silica

Vitreous silica

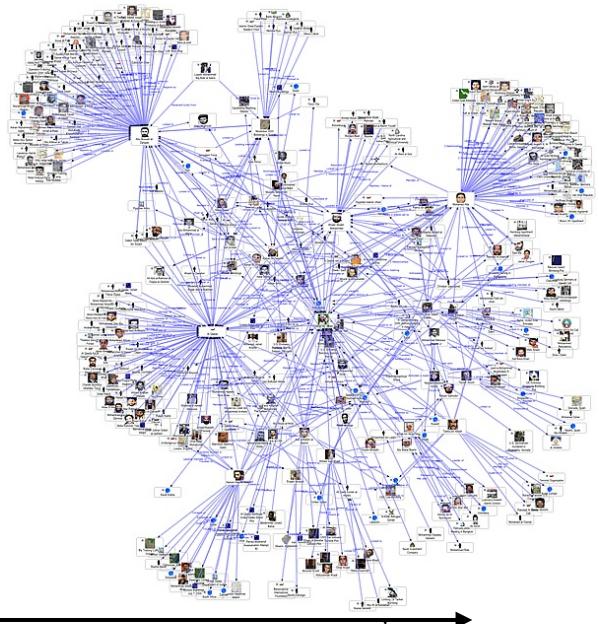


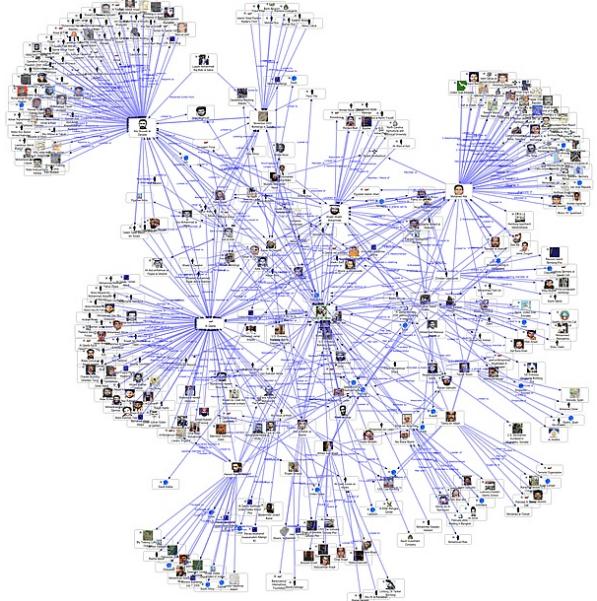
increasing network heterogeneity/complexity

regular, lattices  
physical systems

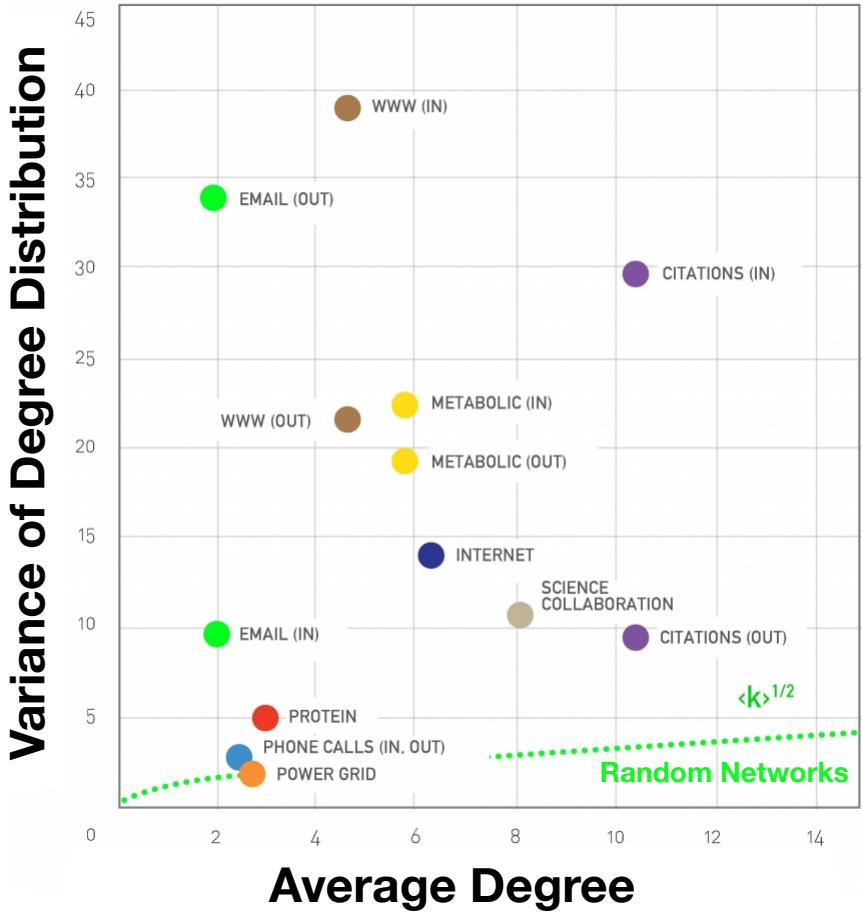
random structures  
mathematics

complex structures  
social systems





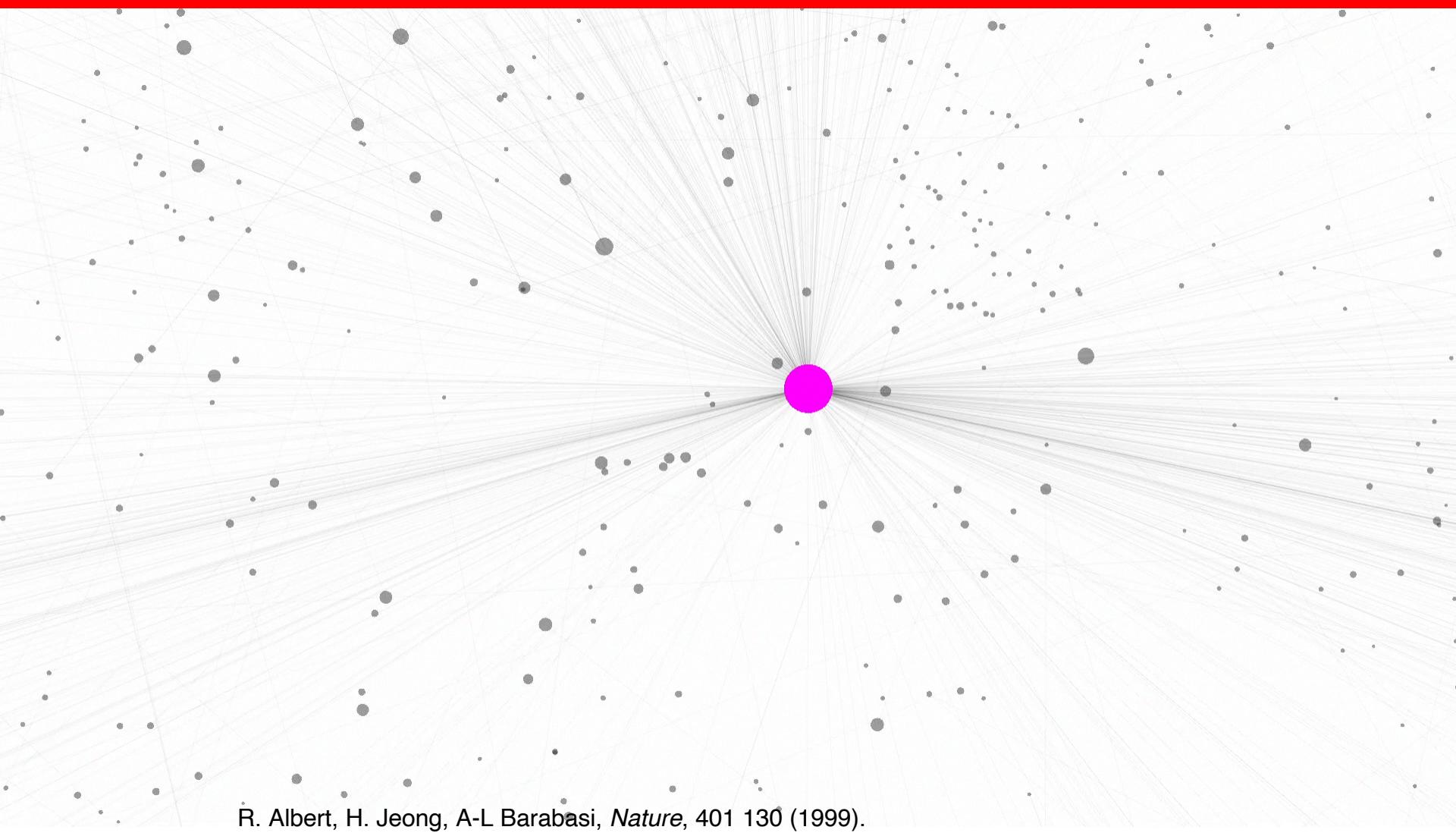
**Las redes reales no son random son libre de escala!**



# Leyes de potencia y redes libre de escala

El nombre libre de escala captura la falta de una escala interna, una consecuencia del hecho de que los nodos con grados muy diferentes coexisten en la misma red. Esta característica distingue las redes libre de escala de las grillas, en las que todos los nodos tienen exactamente el mismo grado ( $\sigma = 0$ ), o de las redes aleatorias, cuyos grados varían en un rango estrecho ( $\sigma = \langle k \rangle^{1/2}$ )

# WORLD WIDE WEB



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

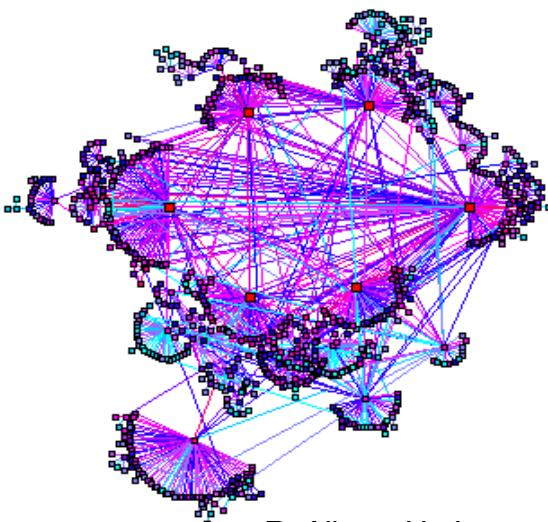
# WORLD WIDE WEB

Nodos: **documentos WWW**

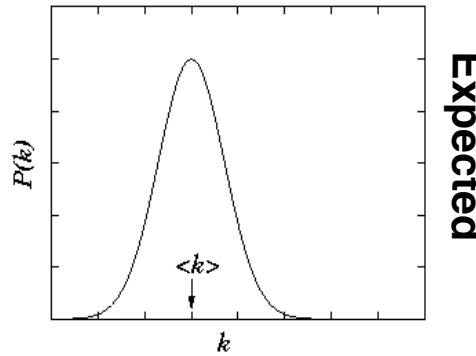
Enlaces: **links URL**

Más de 3 mil millones de documentos.

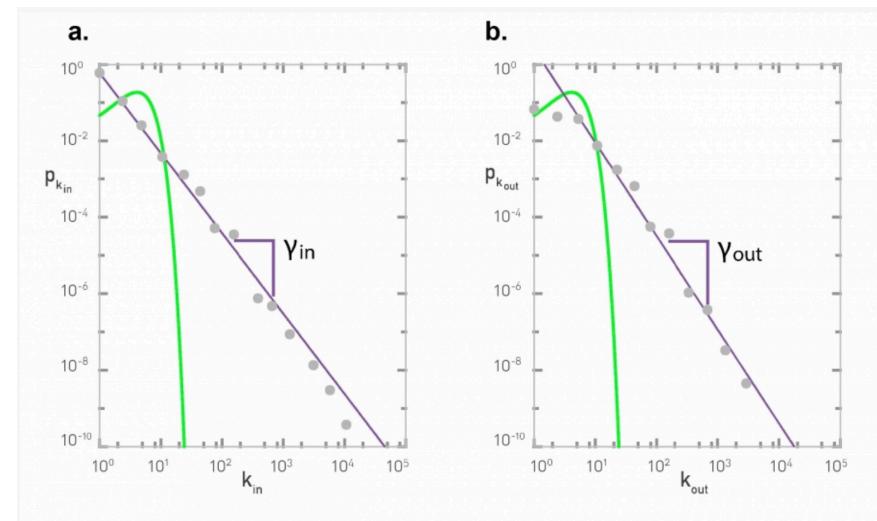
ROBOT: recopila todas las URL encontradas en un documento y las sigue de manera recursiva

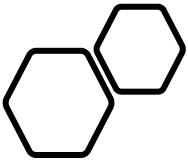


R. Albert, H. Jeong, A-L Barabasi, *Nature*, 401 130 (1999).



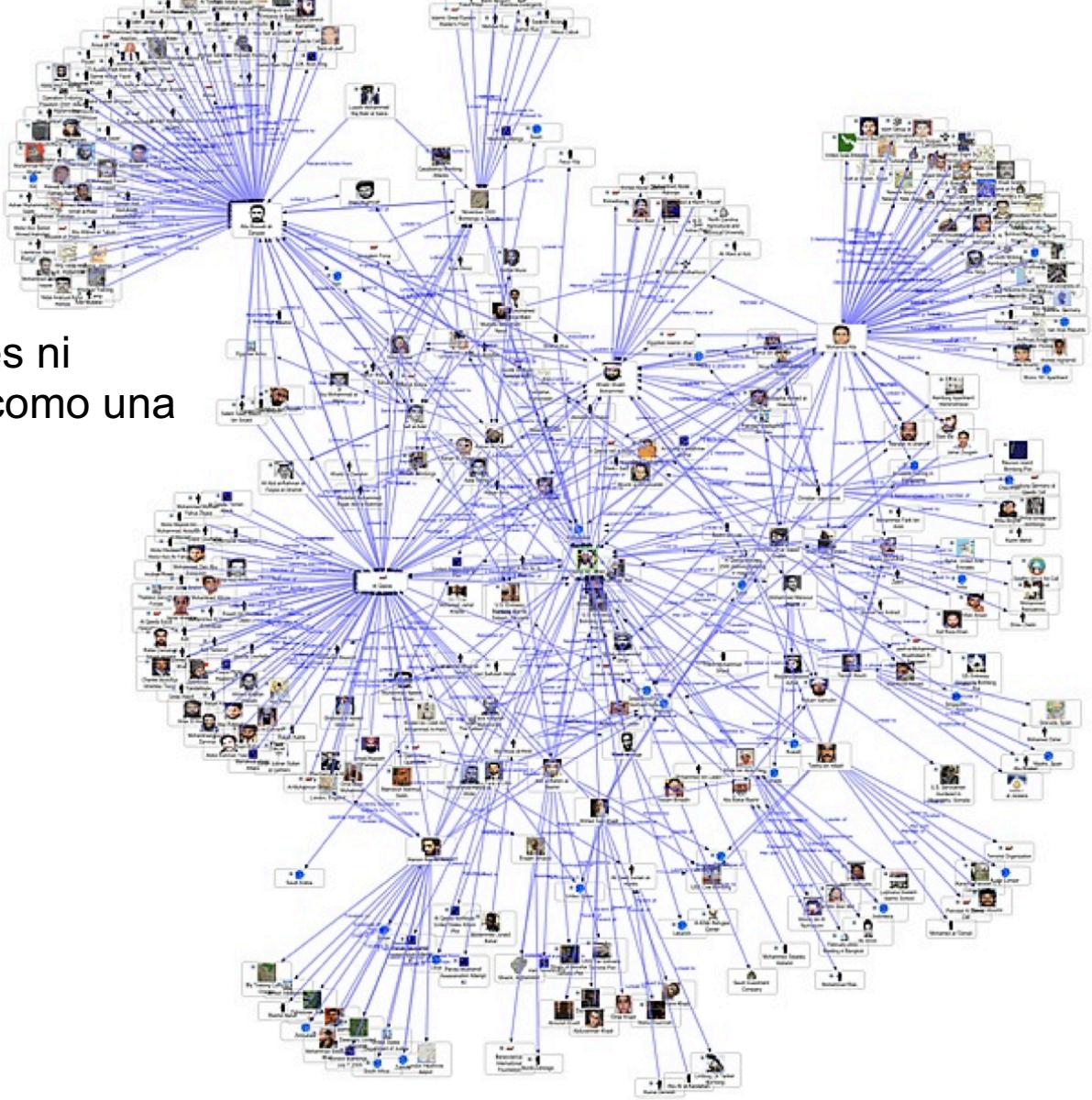
Expected





Las redes reales no son regulares ni aleatorias, ¡son major descritas como una red libre de escala!

- El nombre sin escala capture la falta de una escala interna, una consecuencia del hecho de que nodos con grados muy diferentes coexisten en la misma red. Esta característica distingue las redes sin escala de las redes, en las que todos los nodos tienen exactamente el mismo grado ( $\sigma = 0$ ), o de las redes aleatorias, cuyos grados varían en un rango estrecho ( $\sigma = \langle k \rangle^{1/2}$ )



# Formalismo discreto vs. continuo.

## Formalismo discreto

Como los grados de nodo son siempre enteros positivos, el formalismo discreto captura la probabilidad de que un nodo tenga exactamente  $k$  enlaces:

$$p_k = Ck^{-\gamma}.$$

$$\sum_{k=1}^{\infty} p_k = 1.$$

$$C \sum_{k=1}^{\infty} k^{-\gamma} = 1 \quad C = \frac{1}{\sum_{k=1}^{\infty} k^{-\gamma}} = \frac{1}{\zeta(\gamma)},$$

$$p_k = \frac{k^{-\gamma}}{\zeta(\gamma)}$$

INTERPRETATION:

$$p_k$$

## Formalismo continuo

En los cálculos analíticos a menudo es conveniente suponer que los grados pueden tomar cualquier valor real positivo:

$$p(k) = Ck^{-\gamma}.$$

$$\int_{k_{\min}}^{\infty} p(k)dk = 1$$

$$C = \frac{1}{\int_{k_{\min}}^{\infty} k^{-\gamma} dk} = (\gamma - 1)k_{\min}^{\gamma-1}$$

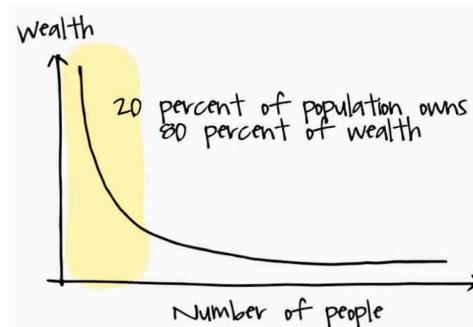
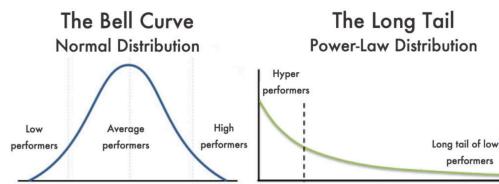
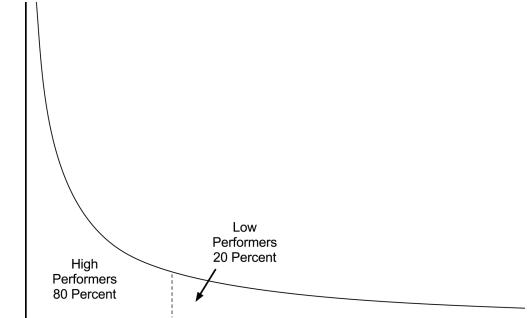
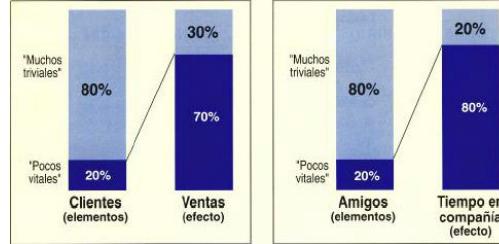
$$p(k) = (\gamma - 1)k_{\min}^{\gamma-1}k^{-\gamma}.$$

$$\int_{k_1}^{k_2} p(k)dk$$

# REGLA 80/20

- Vilfredo Federico Damaso

Pareto (1848 – 1923), Economista, politólogo y filósofo italiano, que contribuyó de manera importante a nuestra comprensión de la distribución del ingreso y al análisis de las elecciones individuales. Algunos de sus principios fundamentales llevan su nombre, como la eficiencia de Pareto, la distribución de Pareto (otro nombre para una distribución de ley de poder), el principio de Pareto (o ley 80/20).



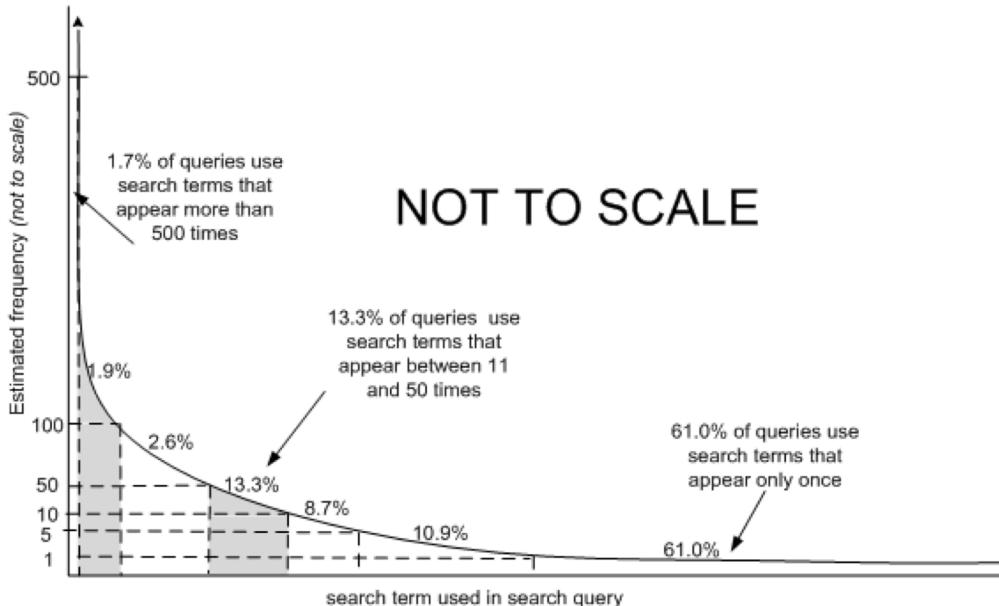
# The Long Tail



Why the Future of Business  
Is Selling Less of More

CHRIS ANDERSON

"Anderson's insights influence Google's strategic thinking in a profound way.  
READ THIS BRILLIANT AND TIMELY BOOK."  
—ERIC SCHMIDT, CEO, GOOGLE



**¿Qué sucede cuando desaparecen los cuellos de botella que se interponen entre la oferta y la demanda en nuestra cultura y todo está al alcance de todos?**

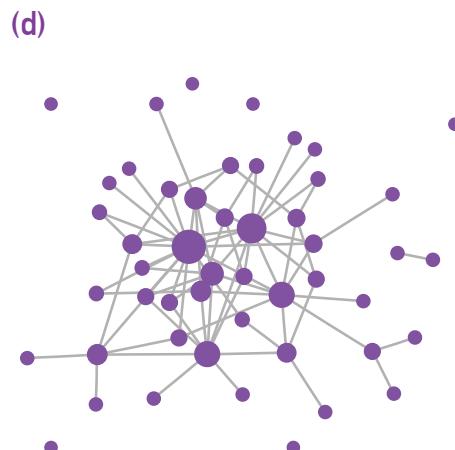
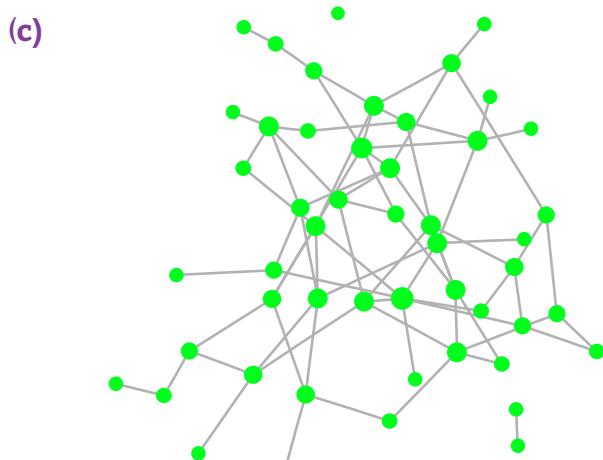
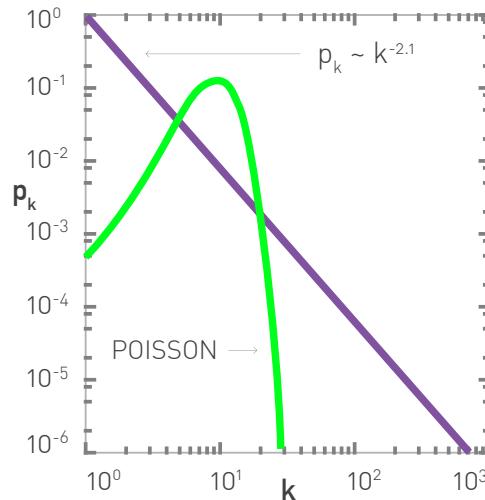
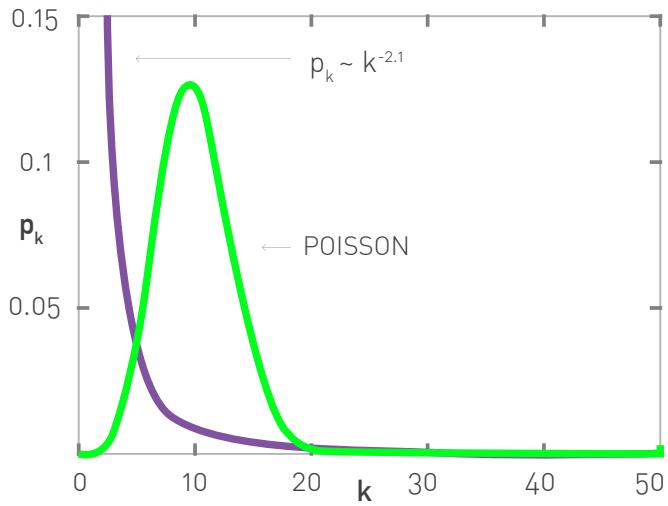
"The Long Tail" es una nueva y poderosa fuerza en nuestra economía: el surgimiento del nicho. A medida que el costo de llegar a los consumidores se reduce drásticamente, nuestros mercados están pasando de un modelo único de atractivo masivo a uno de variedad ilimitada para gustos únicos. Desde los estantes de los supermercados hasta las agencias de publicidad, la capacidad de ofrecer una gran variedad de opciones lo está cambiando todo y nos obliga a repensar dónde se encuentran nuestros mercados y cómo llegar a ellos. La selección ilimitada está revelando verdades sobre lo que quieren los consumidores y cómo quieren obtenerlo, desde DVD en Netflix hasta canciones en iTunes y publicidad en Google.

Sin embargo, esto no es solo una virtud de los mercados en línea; es un ejemplo de un modelo económico completamente nuevo para los negocios, uno que recién comienza a mostrar su poder. **Después de un siglo de obsesionarnos con los pocos productos que se encuentran a la cabeza de la curva de demanda, la nueva economía de la distribución nos permite centrarnos en los muchos más productos que se encuentran en la cola, que colectivamente pueden crear un nuevo mercado tan grande como el que tenemos. ya saben.**

The Long Tail se trata realmente de la economía de la abundancia. Las nuevas eficiencias en la distribución, la fabricación y el marketing esencialmente están restableciendo la definición de lo que es comercialmente viable en todos los ámbitos. Si el siglo XX fue de hits, el XXI será igualmente de nichos.

# Hubs

# La diferencia entre una ley de potencia y una distribución exponencial.

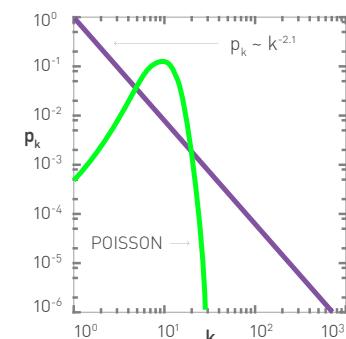


# La diferencia entre una ley de potencia y una distribución exponencial.

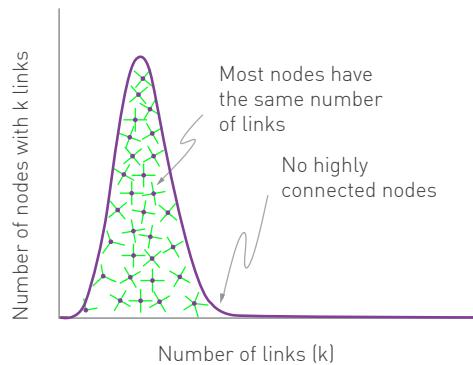
**Usemos la WWW para ilustrar las propiedades del régimen de alto-k.**

La probabilidad de tener un nodo con  $k \sim 100$  es

- Cercana a  $p_{100} \simeq 10^{-30}$  en una distribución de Poisson
- Cercana a  $p_{100} \simeq 10^{-4}$  si  $p_k$  sigue una ley de potencia
- En consecuencia, si la WWW fuera una red aleatoria, de acuerdo con la predicción de Poisson, esperaríamos  $10^{-18}$  nodos de grado  $k > 100$ , o ninguno.
- Para una distribución de grado de ley de potencia, esperamos alrededor de  $N_{k>100} = 10^9$  nodos de grado  $k > 100$



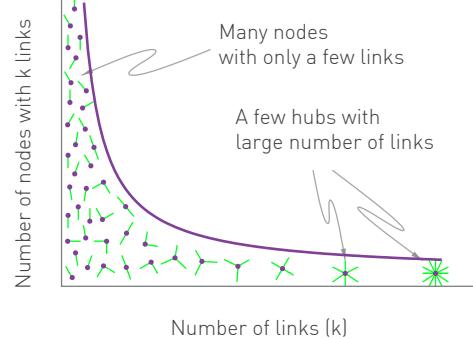
(a) POISSON



(b)



(c) POWER LAW



(d)



# El tamaño del hub más grande.

Todas las redes reales son finitas → exploremos sus consecuencias.

→ Tenemos un grado máximo esperado,  $k_{\max}$

## Estimando $k_{\max}$

$$\int_{k_{\max}}^{\infty} P(k) dk \approx \frac{1}{N}$$

Por qué: la probabilidad de tener un nodo mayor que  $k_{\max}$  no debe exceder la probabilidad de tener un nodo, es decir, una fracción 1 / N de todos los nodos

$$\int_{k_{\max}}^{\infty} P(k) dk = (\gamma - 1) k_{\min}^{\gamma-1} \int_{k_{\max}}^{\infty} k^{-\gamma} dk = \frac{(\gamma - 1)}{(-\gamma + 1)} k_{\min}^{\gamma-1} \left[ k^{-\gamma+1} \right]_{k_{\max}}^{\infty} = \frac{k_{\min}^{\gamma-1}}{k_{\max}^{\gamma-1}} \approx \frac{1}{N}$$

$$k_{\max} = k_{\min} N^{\frac{1}{\gamma-1}}$$

## El tamaño del hub más grande.

$$k_{\max} = k_{\min} N^{\frac{1}{\gamma-1}}$$

Red libre de escala:

$$\gamma = 2.1 \quad k_{\max} \simeq 85,000$$

Red libre aleatoria (hubs prohibidos):

$$k_{\max} \simeq 13$$

Red real WWW:  $k_{\max}=10,721$

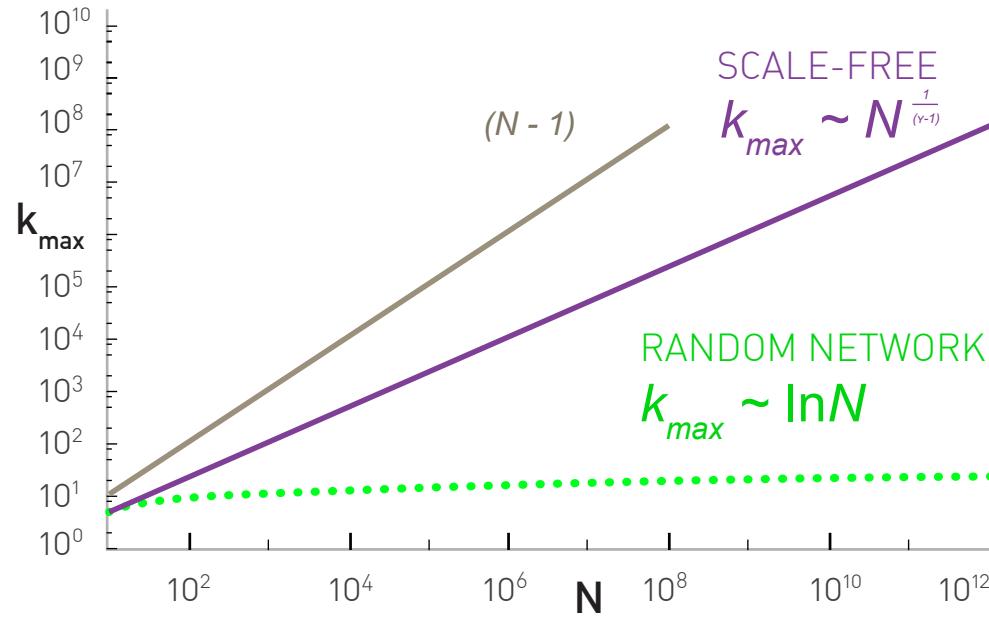
# Redes libre de escala finitas

Grado máximo esperado,  $k_{\max}$

$$k_{\max} = k_{\min} N^{\frac{1}{\gamma-1}}$$

- $k_{\max}$ , Aumenta con el tamaño de la red.  
→ Cuanto más grande es un sistema, más grande es su **hub** más grande.
- Para  $\gamma > 2$   $k_{\max}$  aumenta mas lento que  $N$   
→ el **hub** más grande contendrá una fracción decreciente de enlaces a medida que  $N$  aumente.
- Para  $\gamma = 2$   $k_{\max} \sim N$ .  
→ El tamaño del centro más grande es  $O(N)$  (lineal con  $N$ )
- Para  $\gamma < 2$   $k_{\max}$  aumenta más rápido que  $N$ : fenómenos de condensación  
→ el **hub** más grande tomará una fracción creciente de enlaces. ¡Anomalía!

# El tamaño del hub más grande.



$$k_{\max} = k_{\min} N^{\frac{1}{\gamma-1}}$$

# El significado de la propiedad libre de escala

# Redes libre de escala: Definición

**Definición :**

**Las redes con una ley de potencia en su distribución de grado se llaman  
‘Redes libre de escala’**

¿De dónde viene el nombre?

**Fenómenos críticos e invariancia de escala.**  
**(un desvío)**

# Comportamiento libre de escalas en el espacio.

$$\xi \sim |T - T_c|^{-\nu}$$

a  $T = T_c$ :

la longitud de  
correlación diverge

Las fluctuaciones emergen  
en todas las escalas:

*comportamiento libre de  
escala*



# Scale invariance at the critical point

by Douglas Ashton

[www.kineticallyconstrained.com](http://www.kineticallyconstrained.com)

# FENOMENOS CRITICOS

- La longitud de correlación diverge en el punto crítico:  
¡todo el sistema está correlacionado!
- Invariancia de escala: no hay una escala  
característica para la fluctuación (**comportamiento libre de escala**).
- Universalidad: los exponentes son independientes de  
los detalles del sistema.

## Divergencias en distribuciones sin escala.

$$P(k) = Ck^{-\gamma} \quad k = [k_{\min}, \infty)$$

$$\int_{k_{\min}}^{\infty} P(k) dk = 1$$

$$C = \frac{1}{\int_{k_{\min}}^{\infty} k^{-\gamma} dk} = (\gamma - 1)k_{\min}^{\gamma-1}$$

$$P(k) = (\gamma - 1)k_{\min}^{\gamma-1}k^{-\gamma}$$

$$\langle k^m \rangle = \int_{k_{\min}}^{\infty} k^m P(k) dk \quad \langle k^m \rangle = (\gamma - 1)k_{\min}^{\gamma-1} \int_{k_{\min}}^{\infty} k^{m-\gamma} dk = \frac{(\gamma - 1)}{(m - \gamma + 1)} k_{\min}^{\gamma-1} \left[ k^{m-\gamma+1} \right]_{k_{\min}}^{\infty}$$

Exponente debe ser negativo

If  $m - \gamma + 1 < 0$ :

$$\langle k^m \rangle = -\frac{(\gamma - 1)}{(m - \gamma + 1)} k_{\min}^m$$

If  $m - \gamma + 1 > 0$ ,

La integral diverge.

Para un  $\gamma$  fijo, esto significa que todos los momentos con  $m > \gamma - 1$  divergen.

# DIVERGENCIA DE LOS MOMENTOS SUPERIORES

$$\langle k^m \rangle = (\gamma - 1) k_{\min}^{\gamma-1} \int_{k_{\min}}^{\infty} k^{m-\lambda} dk = \frac{(\gamma - 1)}{(m - \gamma + 1)} k_{\min}^{\gamma-1} \left[ k^{m-\gamma+1} \right]_{k_{\min}}^{\infty}$$

Para un  $\gamma$  fijo, esto significa que todos los momentos con  $m > \gamma - 1$  divergen.

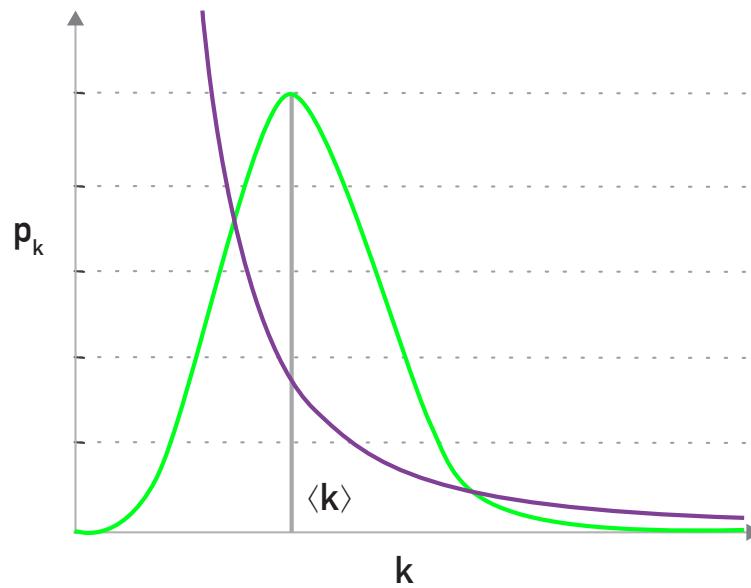
Network	Size	$\langle k \rangle$	$\kappa$	$\gamma_{out}$	$\gamma_{in}$
WWW	325 729	4.51	900	2.45	2.1
WWW	$4 \times 10^7$	7		2.38	2.1
WWW	$2 \times 10^8$	7.5	4000	2.72	2.1
WWW, site	260 000				1.94
Internet, domain*	3015–4389	3.42–3.76	30–40	2.1–2.2	2.1–2.2
Internet, router*	3888	2.57	30	2.48	2.48
Internet, router*	150 000	2.66	60	2.4	2.4
Movie actors*	212 250	28.78	900	2.3	2.3
Co-authors, SPIRES*	56 627	173	1100	1.2	1.2
Co-authors, neuro.*	209 293	11.54	400	2.1	2.1
Co-authors, math.*	70 975	3.9	120	2.5	2.5
Sexual contacts*	2810			3.4	3.4
Metabolic, <i>E. coli</i>	778	7.4	110	2.2	2.2
Protein, <i>S. cerev.</i> *	1870	2.39		2.4	2.4
Ythan estuary*	134	8.7	35	1.05	1.05
Silwood Park*	154	4.75	27	1.13	1.13
Citation	783 339	8.57			3
Phone call	$53 \times 10^6$	3.16		2.1	2.1
Words, co-occurrence*	460 902	70.13		2.7	2.7
Words, synonyms*	22 311	13.48		2.8	2.8

Network	$N$	$L$	$\langle k \rangle$	$\langle k_{in}^2 \rangle$	$\langle k_{out}^2 \rangle$	$\langle k^2 \rangle$	$\gamma_{in}$	$\gamma_{out}$	$\gamma$
Internet	192,244	609,066	6.34	-	-	240.1	-	-	3.42*
WWW	325,729	1,497,134	4.60	1546.0	482.4	-	2.00	2.31	-
Power Grid	4,941	6,594	2.67	-	-	10.3	-	-	Exp.
Mobile-Phone Calls	36,595	91,826	2.51	12.0	11.7	-	4.69*	5.01*	-
Email	57,194	103,731	1.81	94.7	1163.9	-	3.43*	2.03*	-
Science Collaboration	23,133	93,437	8.08	-	-	178.2	-	-	3.35*
Actor Network	702,388	29,397,908	83.71	-	-	47,353.7	-	-	2.12*
Citation Network	449,673	4,689,479	10.43	971.5	198.8	-	3.03*	4.00*	-
E. Coli Metabolism	1,039	5,802	5.58	535.7	396.7	-	2.43*	2.90*	-
Protein Interactions	2,018	2,930	2.90	-	-	32.3	-	-	2.89*-

Muchos exponentes de grado son más pequeños que 3

→  $\langle k^2 \rangle$  diverge en el límite  $N \rightarrow \infty$

# El significado de sin escala



## Random Network

Nodo elegido al azar:  
Scale:  $\langle k \rangle$

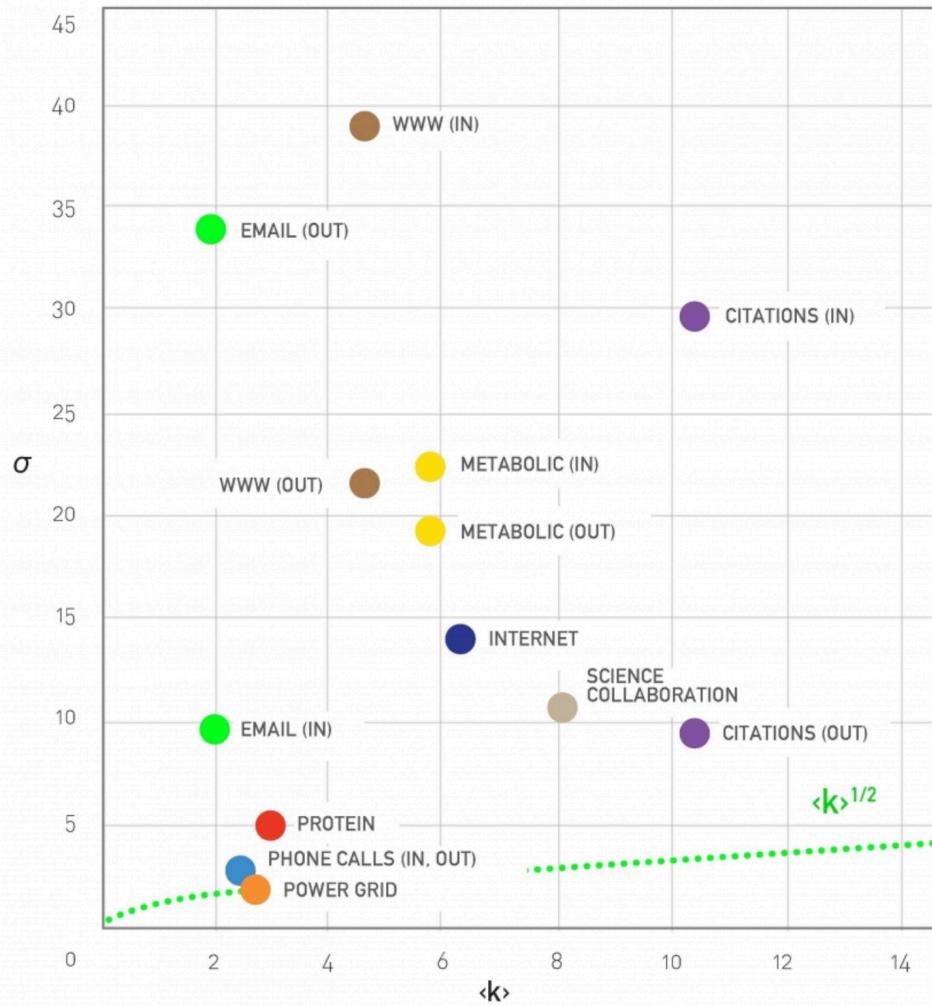
$$k = \langle k \rangle \pm \langle k \rangle^{1/2}$$

## Scale-Free Network

Nodo elegido al azar:  
Scale: none

$$k = \langle k \rangle \pm \infty$$

# El significado de sin escala



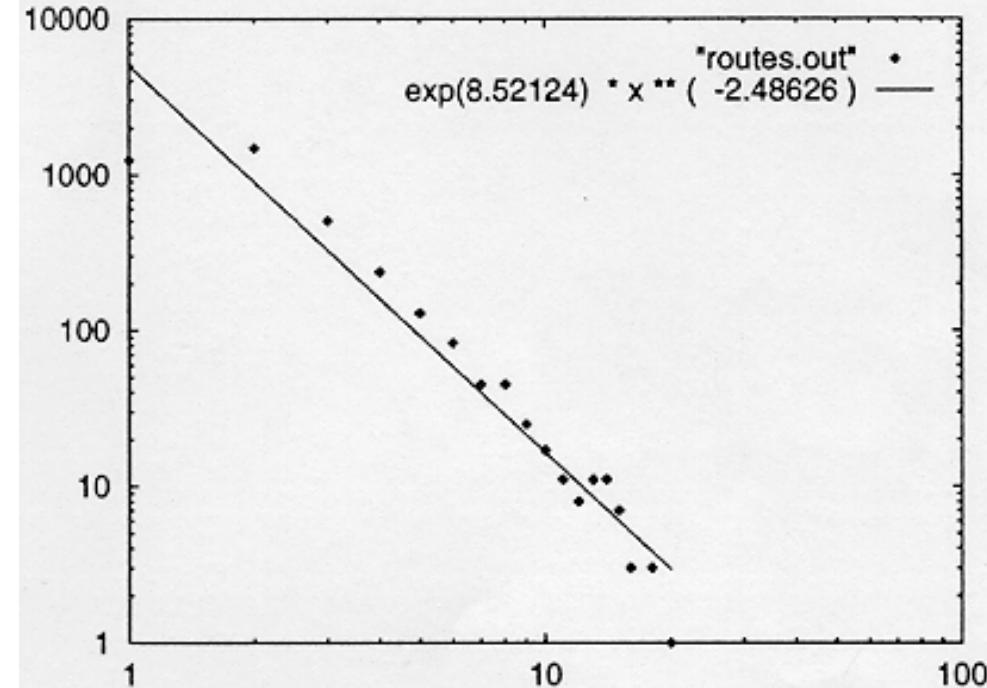
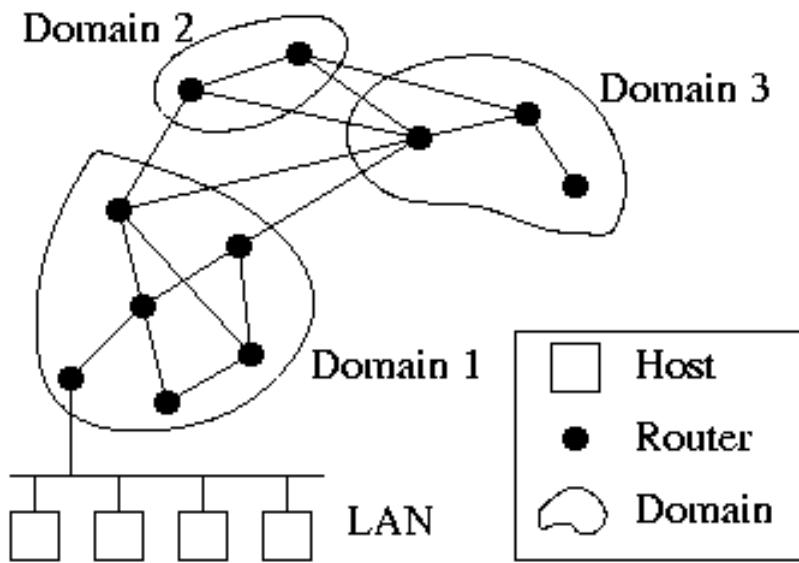
$$k = \langle k \rangle \pm \sigma_k$$

# Universalidad

# Backbone de internet

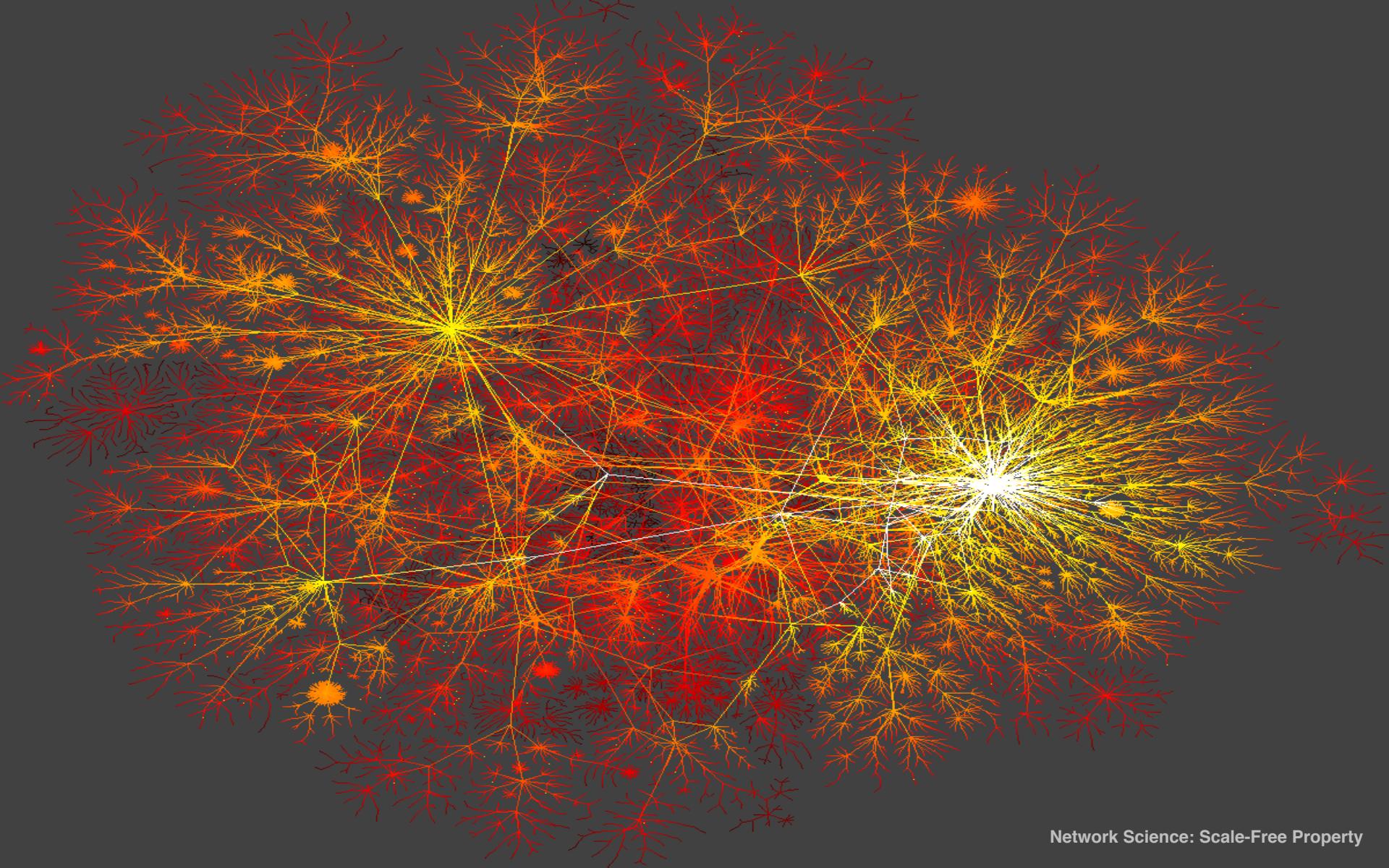
**Nodos:** computadores, routers

**Enlaces:** líneas físicas



(Faloutsos, Faloutsos and Faloutsos, 1999)

Network Science: Scale-Free Property



Network Science: Scale-Free Property

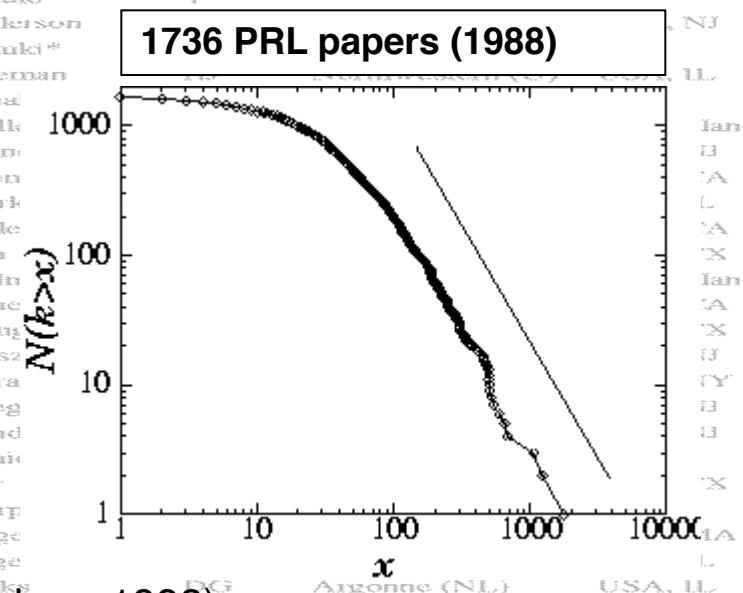
# INDICE DE CITACIONES EN CIENCIA

Out of over 500,000 Examined  
 (see <http://www.sst.nrel.gov>)

Nodos: papers

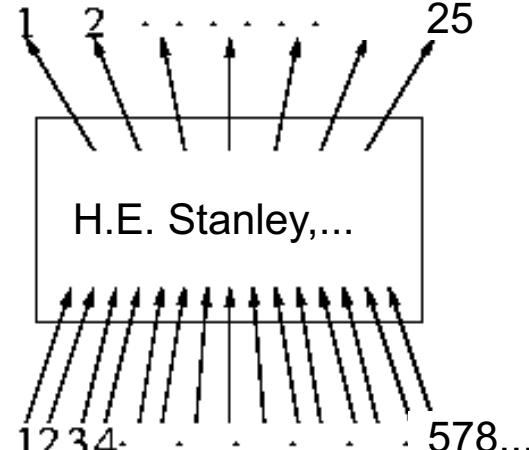
Enlaces: citaciones

		Institute	Country	Field	avg. cites	total art.	total cites	rank by total cit.
Witten	E	Princeton (U)	USA, NJ	High-energy (D)	168	138	23235	1
Cava	RJ	UCSB (U)	USA, CA	Semic				2
Batlogg	B	Bell Labs (D)	USA, NJ	Superx				3
Ploog	K	Max-Planck (NL)	Germany	Semic				4
Ellis	J	Euro Nuclear Cent.	Switzerland	Astrop				5
Fisk	Z	Florida State (U)	USA, FL	Solid t				6
Cardona	M	Max Planck (NL)	Germany	Semic				7
Nanopoulos	DV	Texas A&M (U)	USA, TX	High-e				8
Heeger	AJ	UCSB (U)	USA, CA	Polym				9
Lee*	PA							10
Suzuki*	T							11
Anderson								12
Suzuki*								13
Freeman								14
Tanaka								15
Muller								16
Schmidt								17
Chen								18
Mork								19
Mille								20
Chu								21
Bednorz								22
Cohen								23
Meissner								24
Waszczak								25
Shiraishi								26
Wiegert								27
Vandamme								28
Uchida								29
Horn								30
Maple								31
Birge								32
Jorge								33
links								34
								35



(S. Redner, 1998)

\* citation total may be skewed because of multiple authors with the same name

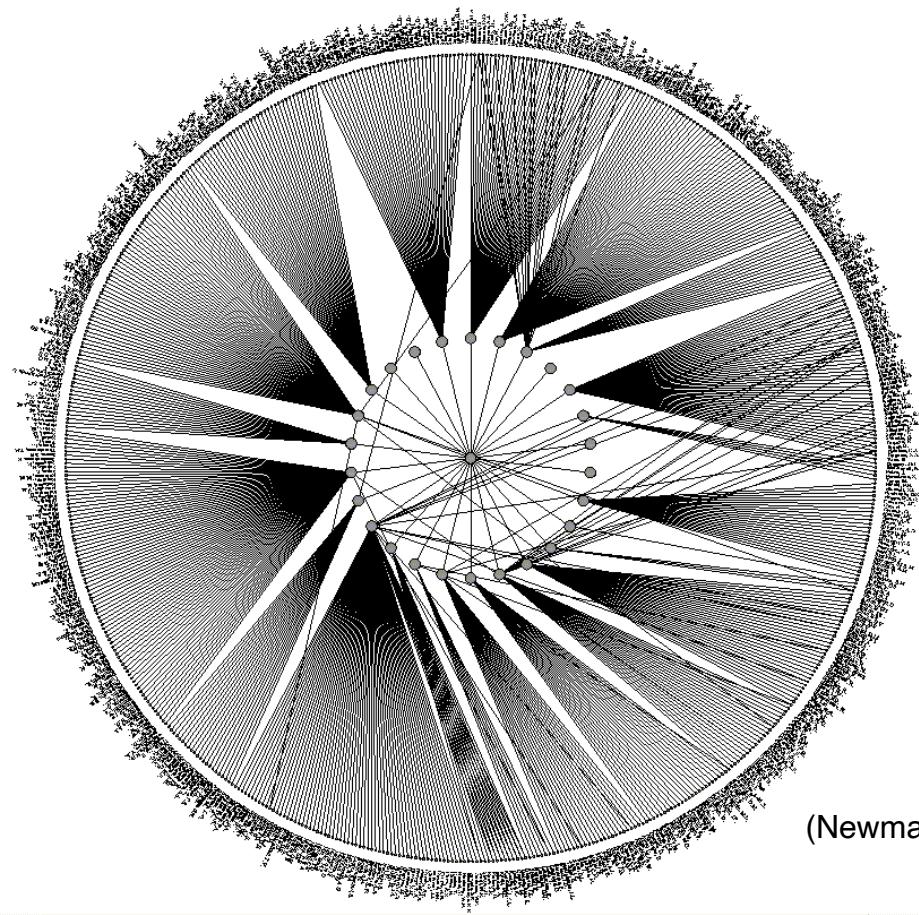


$$P(k) \sim k^{-\gamma} \quad (\gamma = 3)$$

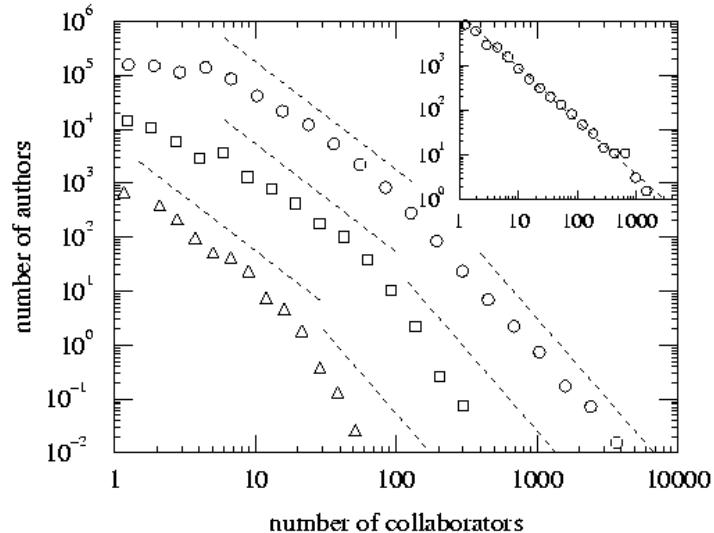
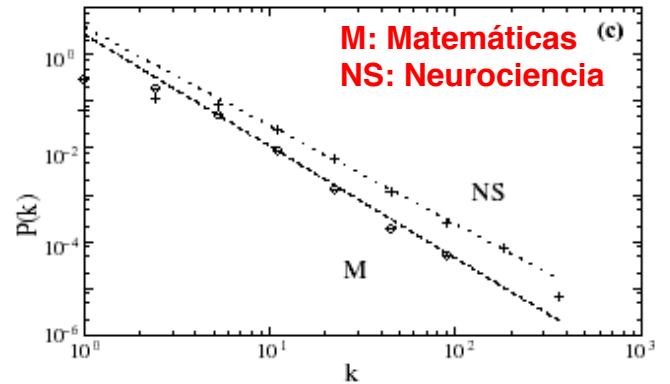
# CO-AUTORIA EN CIENCIA

**Nodos:** Científicos (autores)

**Links:** Publicaciones conjuntas



(Newman, 2000, Barabasi et al 2001)

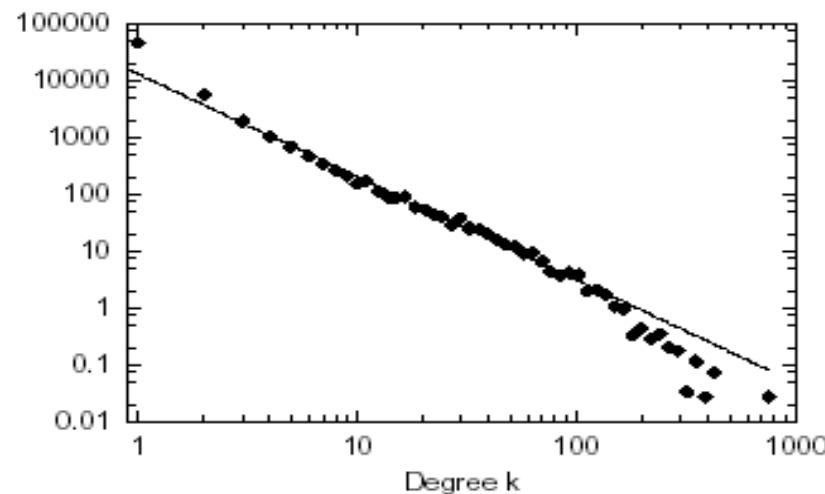


# COMUNIDADES ONLINE

**Nodos:** Usuarios online

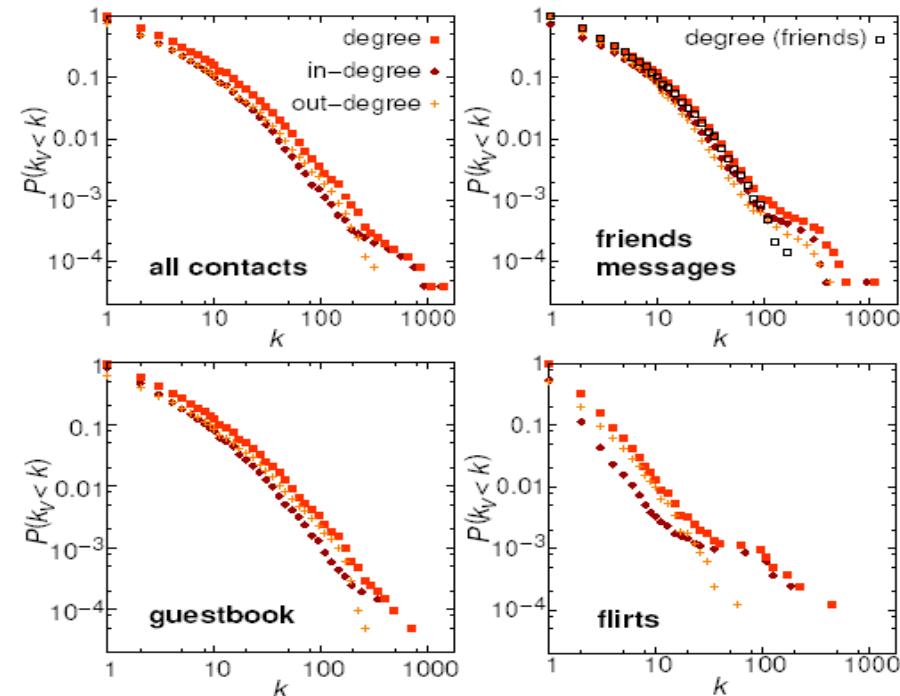
**Enlaces:** contactos email

Archivos log de Kiel University  
112 días, N=59,912 nodos



Ebel, Mielsch, Bornholdtz, PRE 2002.

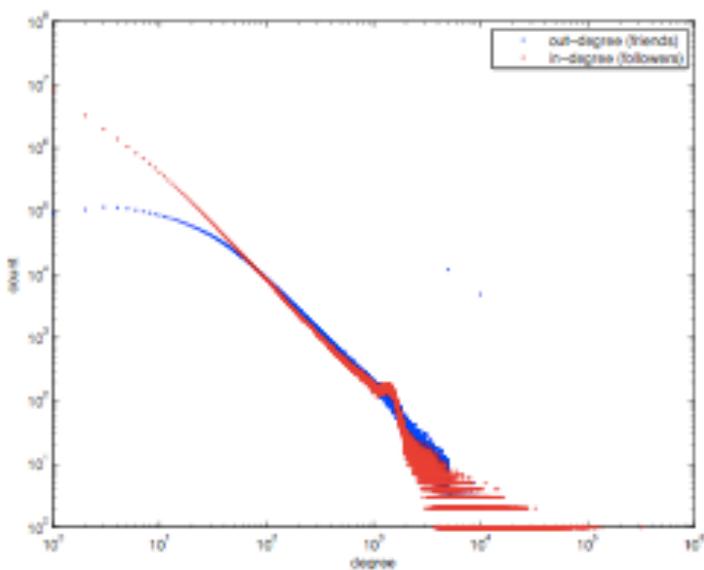
Pussokram.com comunidad online;  
512 days, 25,000 users.



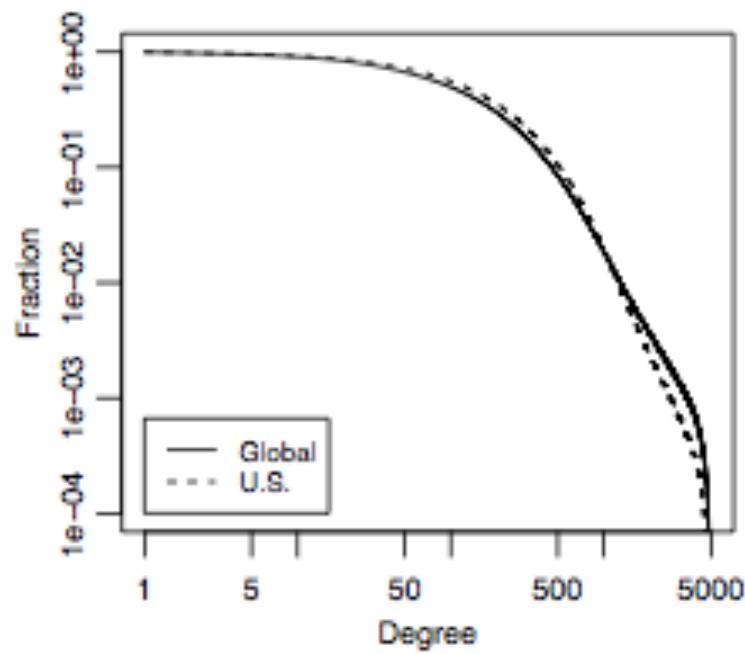
Holme, Edling, Liljeros, 2002.

# COMUNIDADES ONLINE

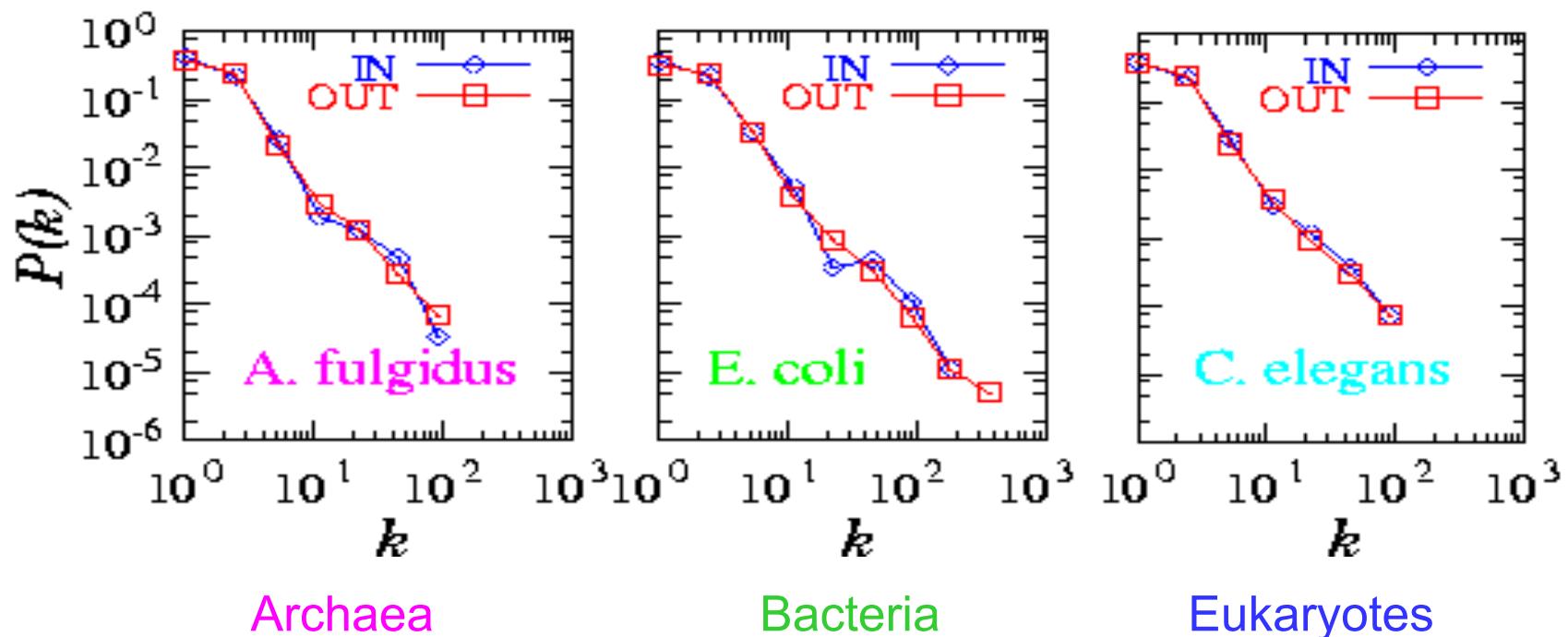
Twitter:



Facebook



# RED METABOLICA



¡Los organismos de los tres dominios de la vida son **libres de escala!**

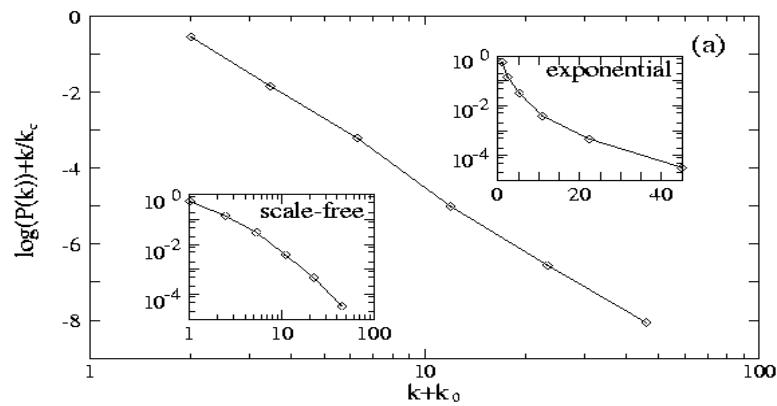
$$P_{in}(k) \approx k^{-2.2}$$

$$P_{out}(k) \approx k^{-2.2}$$

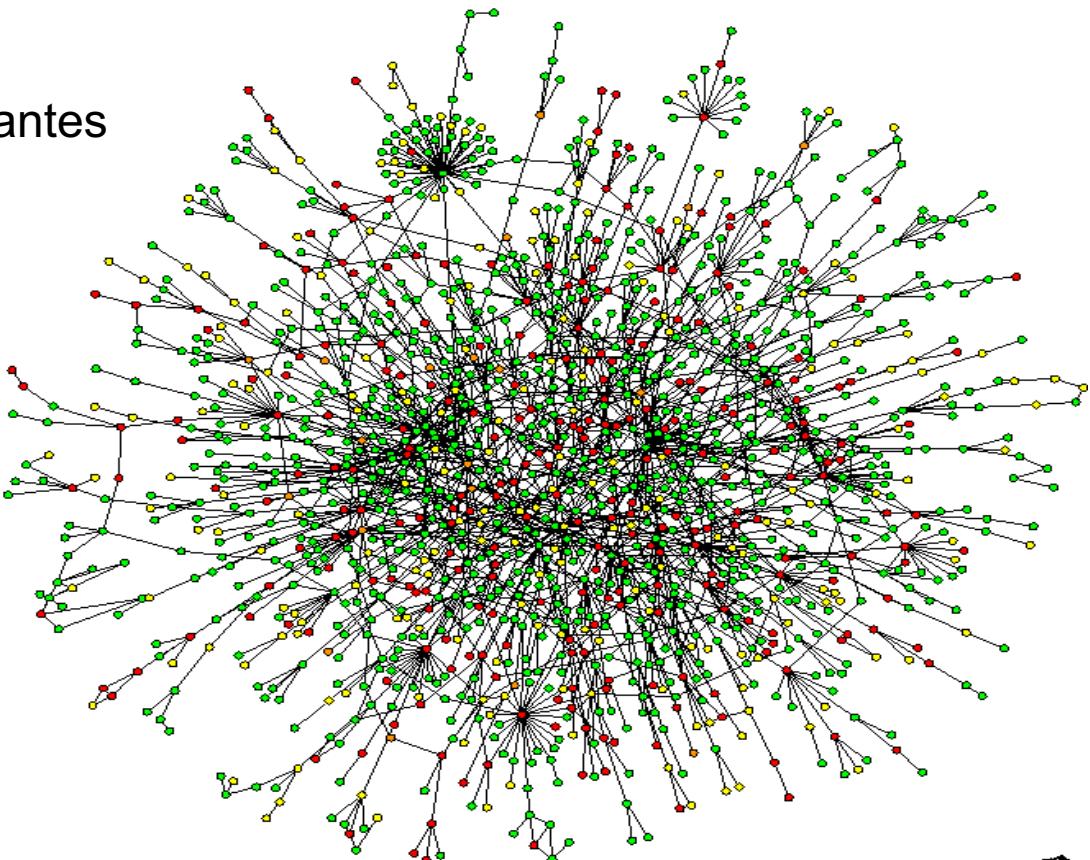
# TOPOLOGÍA DE LA RED DE PROTEÍNAS

Nodos: proteínas

Enlaces: interacciones físicas vinculantes



$$P(k) \sim (k + k_0)^{-\gamma} \exp\left(-\frac{k + k_0}{k_\tau}\right)$$

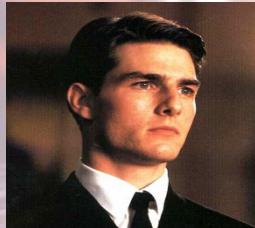


# RED DE ACTORES

Nodos: actores

Enlaces: película conjunta

IMDb Internet Movie Database



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

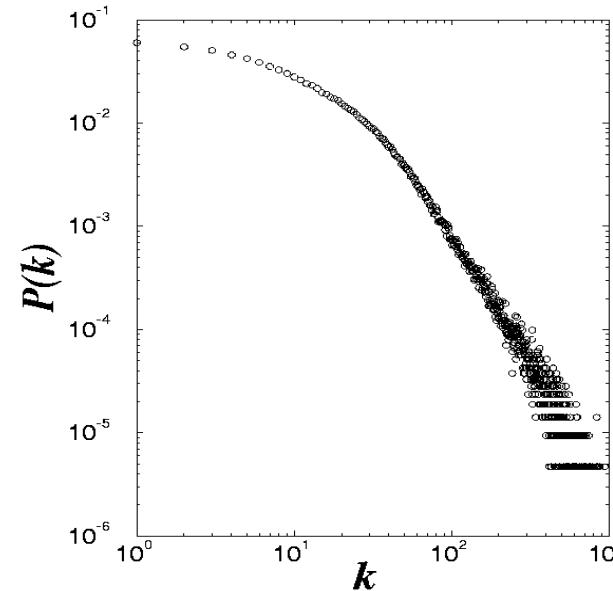
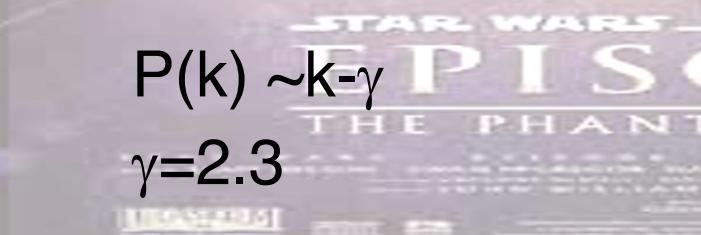


$N = 212,250$  actores

$\langle k \rangle = 28.78$

$P(k) \sim k^{-\gamma}$

$\gamma=2.3$

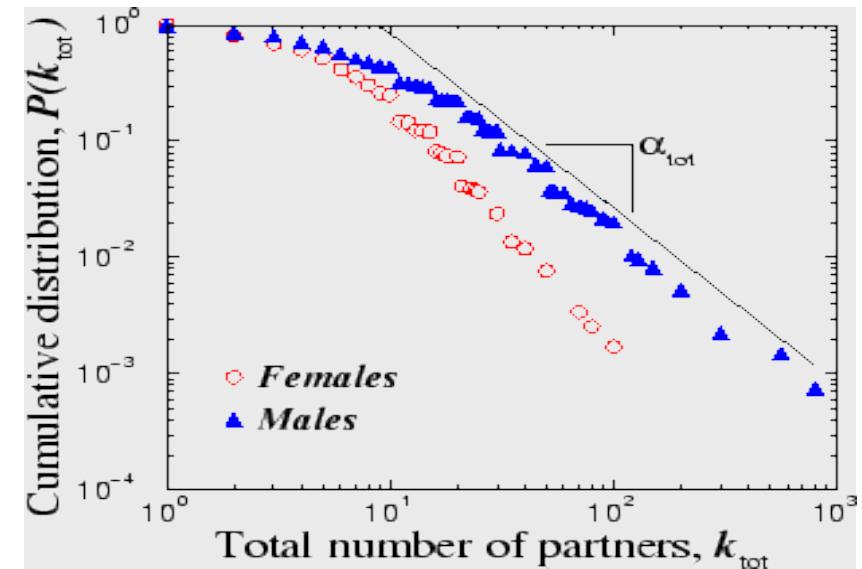


# SWEDISH SE-WEB



**Nodos:** Personas (Mujeres; Hombres)

**Enlaces:** Relaciones sexuales

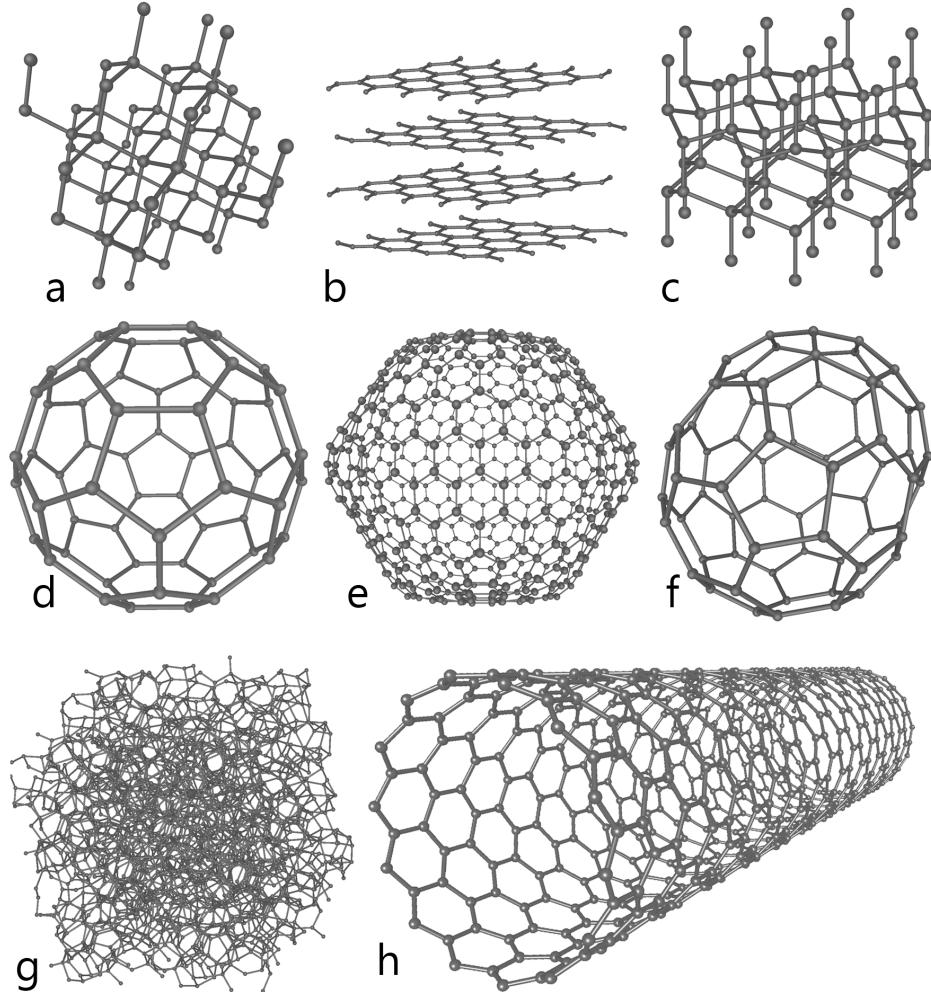


4781 Suizos; 18-74;  
59% tasa de respuesta.

Liljeros et al. Nature 2001

# No todas las redes son libres de escala

- *Redes que aparecen en la ciencia de los materiales, como la red que describe los enlaces entre los átomos en materiales cristalinos o amorfos, donde cada nodo tiene exactamente el mismo grado.*
- *La red neuronal del gusano C.elegans.*
- *La red eléctrica, compuesta por generadores e interruptores conectados por líneas de transmisión.*



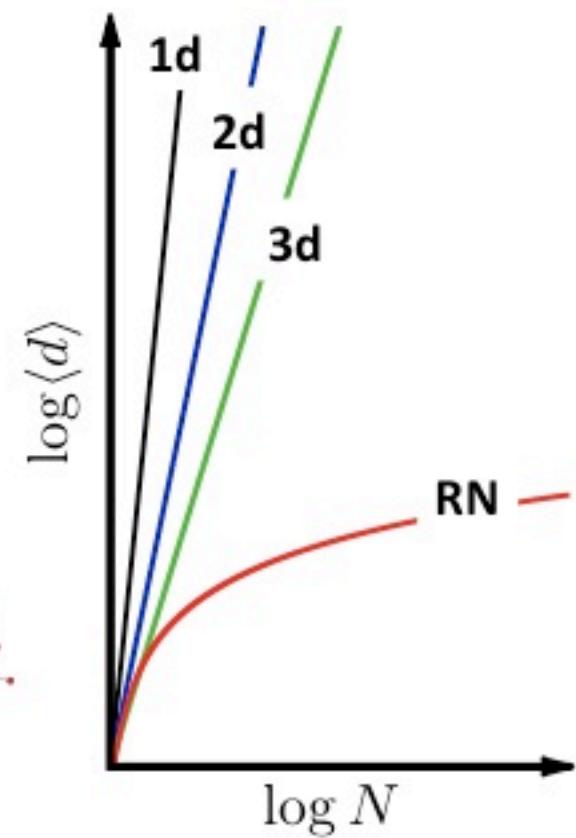
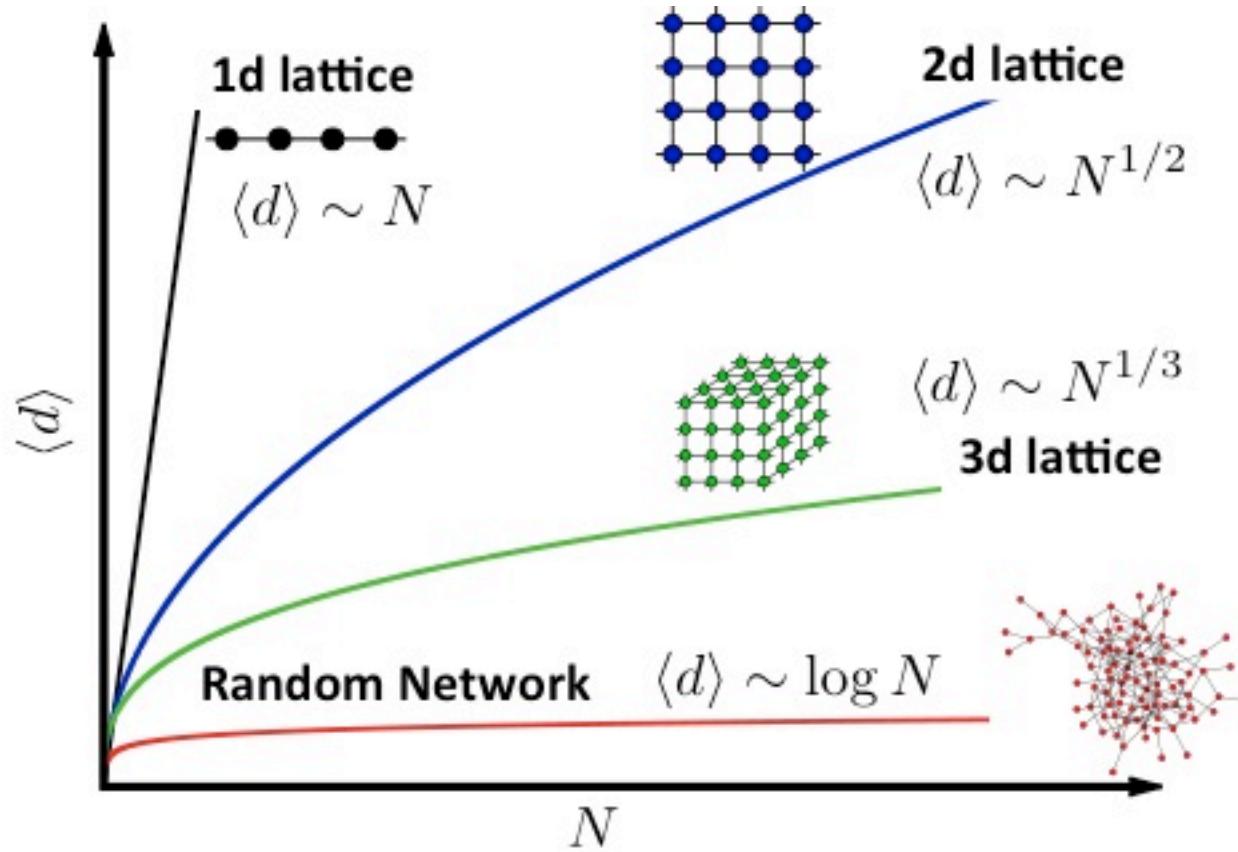
# Propiedad ultra-pequeña

# COMPORTAMIENTO DE MUNDO PEQUEÑO EN REDES SIN ESCALA

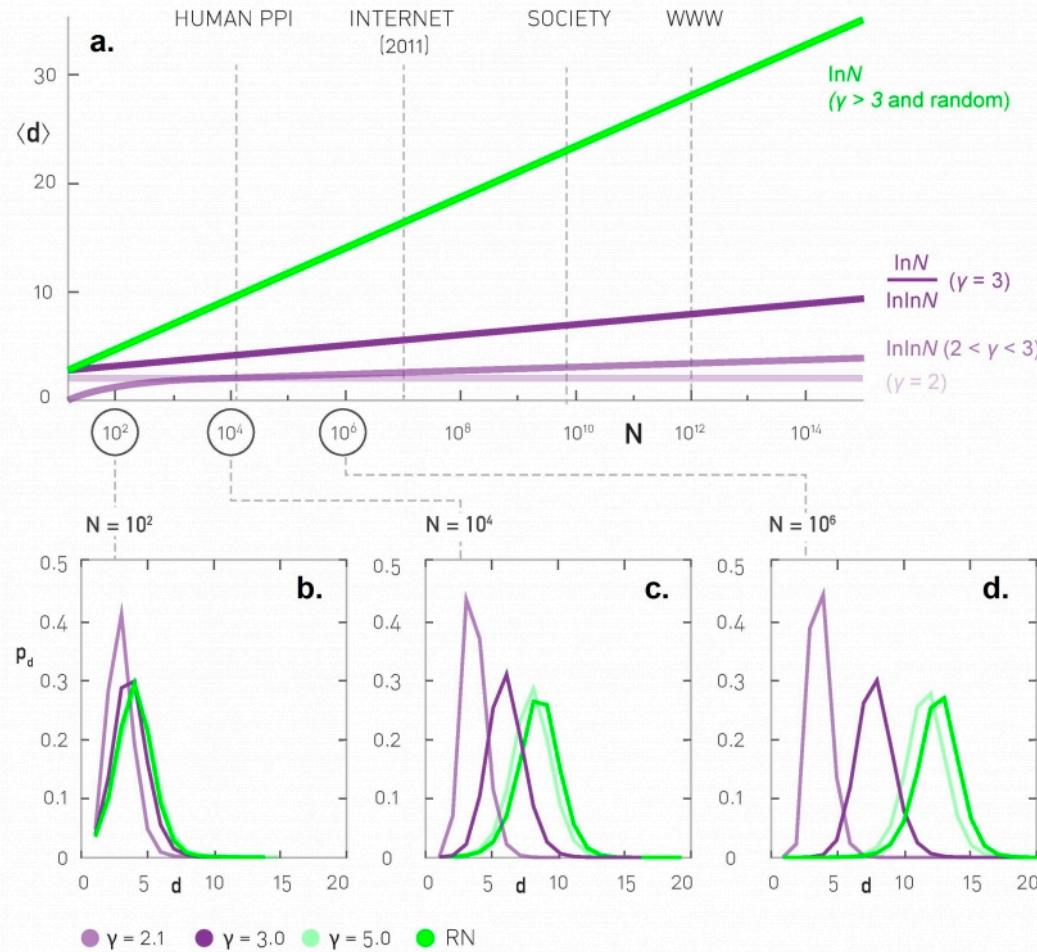
**Ultra  
Small  
World**

$\langle l \rangle \sim$	$const.$ $\gamma = 2$	<p>El tamaño del hub más grande es de orden O (N). La mayoría de los nodos se pueden conectar dentro de dos enlaces (la gran mayoría están conectados al hub), por lo que la longitud promedio de la ruta será independiente del tamaño del sistema.</p>
	$\frac{\ln \ln N}{\ln(\gamma - 1)}$ $2 < \gamma < 3$	<p>La longitud media de la ruta aumenta más lentamente que logarítmicamente. En una red aleatoria, todos los nodos tienen un grado comparable, por lo que la mayoría de las rutas tendrán una longitud comparable. En una red sin escala, la gran mayoría de la ruta pasa por unos pocos hubs de alto grado, lo que reduce las distancias entre los nodos.</p>
	$\frac{\ln N}{\ln \ln N}$ $\gamma = 3$	<p>Algunos modelos clave producen <math>\gamma = 3</math>, por lo que el resultado es de particular importancia para ellos. Esto fue derivado por primera vez por Bollobas et al. para el diámetro de la red en el contexto de un modelo dinámico, pero también se aplica a la longitud de ruta promedio.</p>
<b>Small World</b>	$\ln N$ $\gamma > 3$	<p>El segundo momento de la distribución es finito, por lo que de muchas maneras la red se comporta como una red aleatoria. Por lo tanto, la longitud promedio de la ruta sigue el resultado que obtuvimos anteriormente para el modelo de red aleatoria.</p>

# ¿Por qué sorprenden los pequeños mundos? Sorprendente en comparación con qué?



# COMPORTAMIENTO DE MUNDO PEQUEÑO EN REDES SIN ESCALA

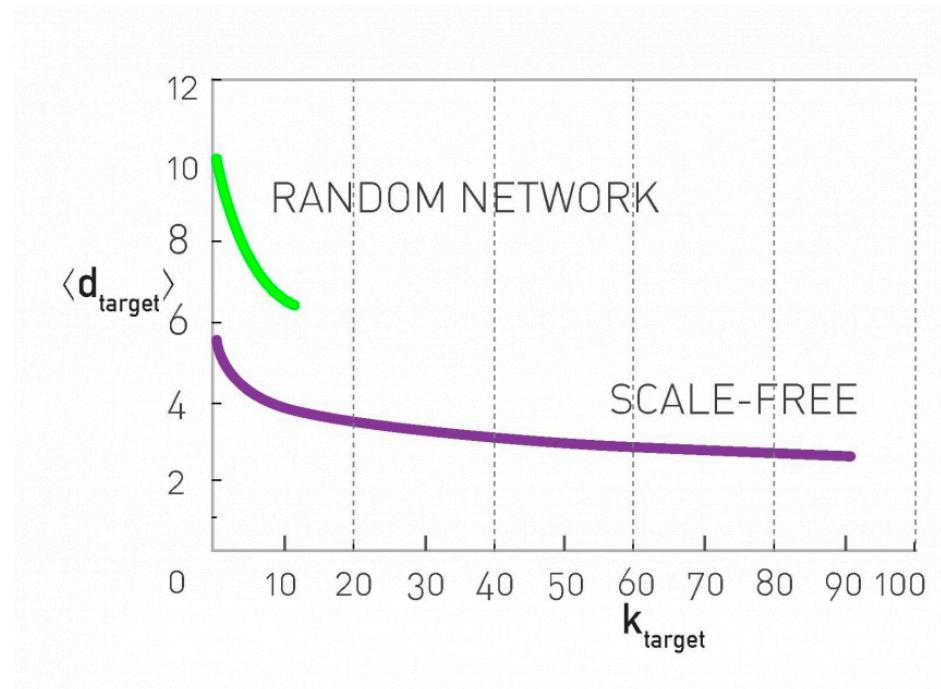


$$\langle d \rangle \sim \begin{cases} \text{const.} & \gamma = 2, \\ \frac{\ln \ln N}{\ln(\gamma - 1)} & 2 < \gamma < 3, \\ \frac{\ln N}{\ln \ln N} & \gamma = 3, \\ \ln N & \gamma > 3. \end{cases}$$

# Siempre estamos cerca de los hubs.

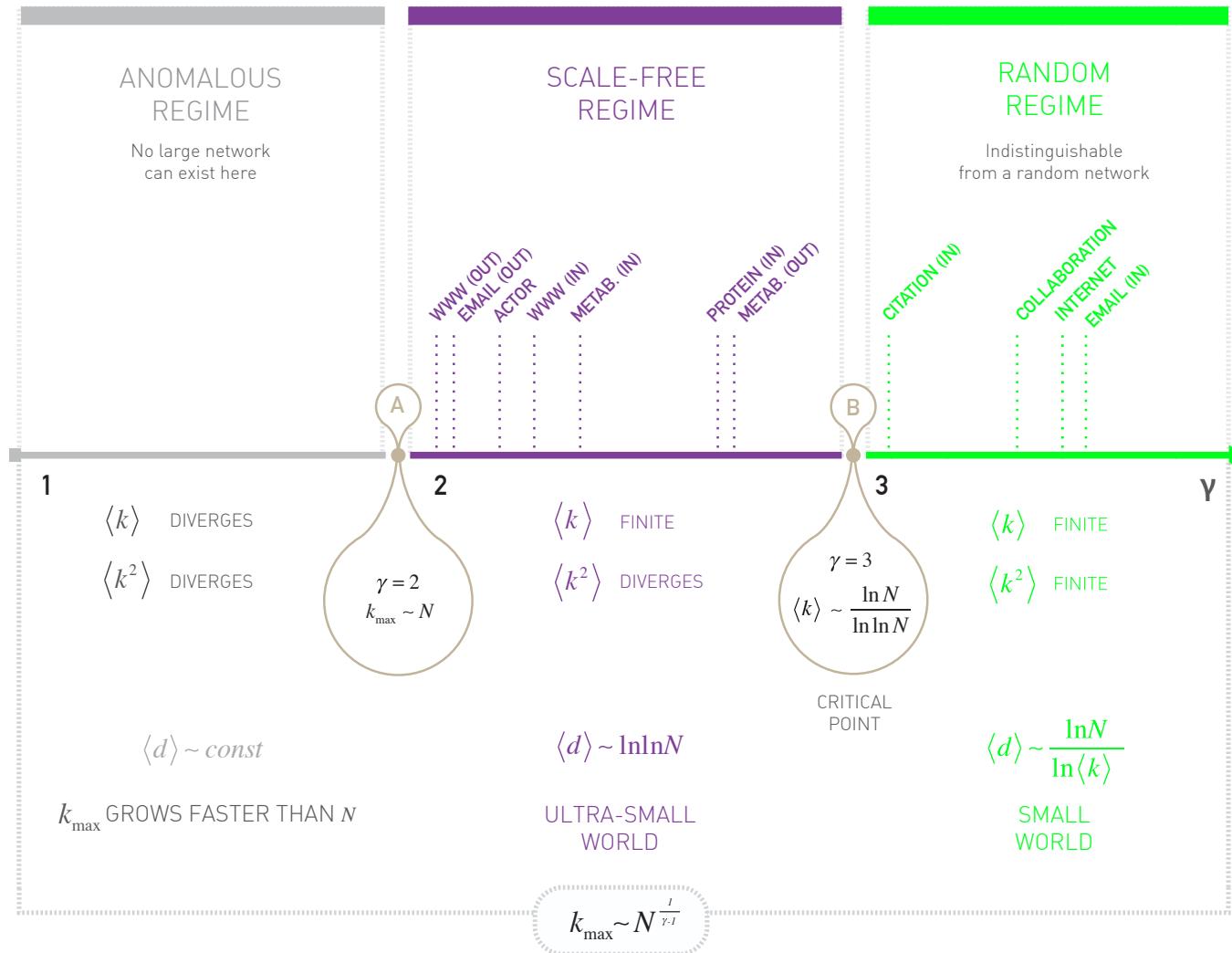
"Siempre es más fácil encontrar a alguien que conozca a una figura famosa o popular que a una persona insignificante."

(Frigyes Karinthy, 1929)



# El rol del exponente del grado.

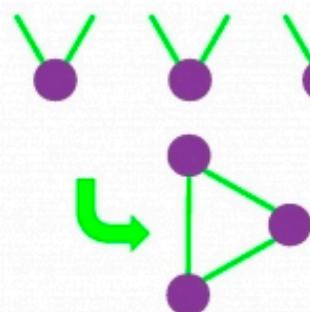
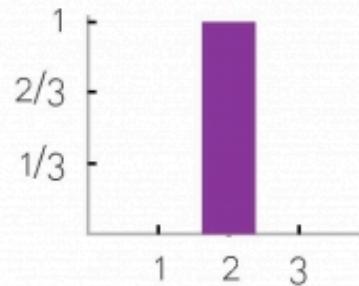
# RESUMEN DEL COMPORTAMIENTO DE REDES SIN ESCALA



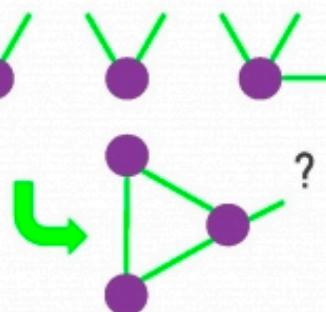
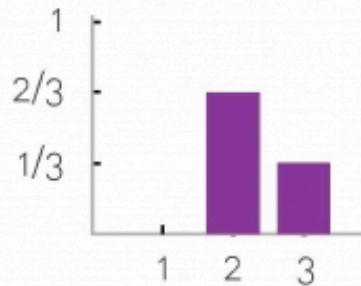
# Gáficamente: No hay redes grandes para $\gamma < 2$

El HUB más grande crece más rápido que N. Si no permitimos auto-buclees y enlaces múltiples, entonces el hub más grande se quedará sin nodos para conectarse una vez que su grado exceda de N-1.

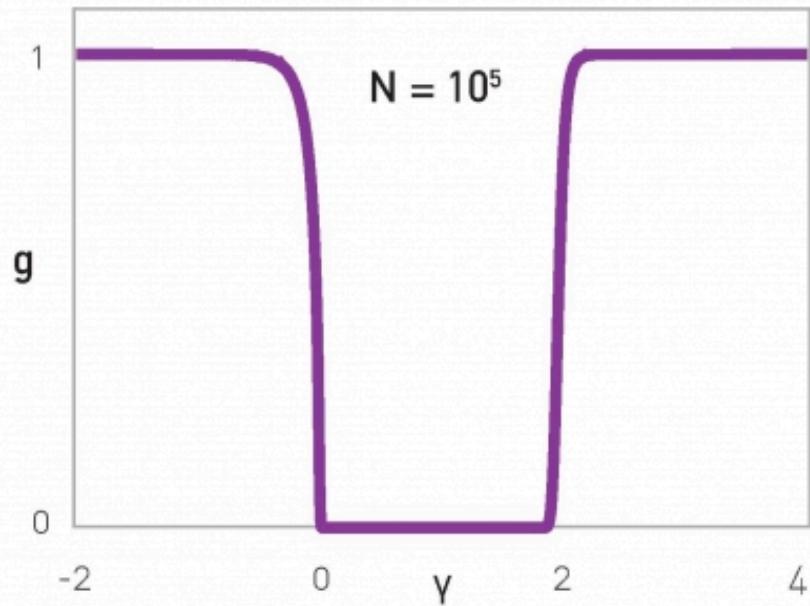
a. Graphical



b. Not Graphical



c.



En redes libre de escala:  $k_{\max} = k_{\min} N^{\frac{1}{\gamma-1}}$

Para  $\gamma < 2$ :  $1/(\gamma-2) > 1$

# ¿Por qué no vemos redes con exponentes en el rango de $\gamma = 4,5,6$ , etc.?

Para documentar una red sin escala, necesitamos un escalamiento de 2-3 órdenes de magnitud. Es decir,  $K_{\max} \sim 10^3$

Sin embargo, eso limita el tamaño del sistema que necesitamos para documentarlo. Por ejemplo, para medir un exponente  $\gamma = 5$ , necesitamos en grado máximo un tamaño de sistema del orden de

$$K_{\max} = K_{\min} N^{\frac{1}{\gamma-1}}$$

$$N = \left( \frac{K_{\max}}{K_{\min}} \right)^{\gamma-1} \approx 10^8$$

Mobile Call Network

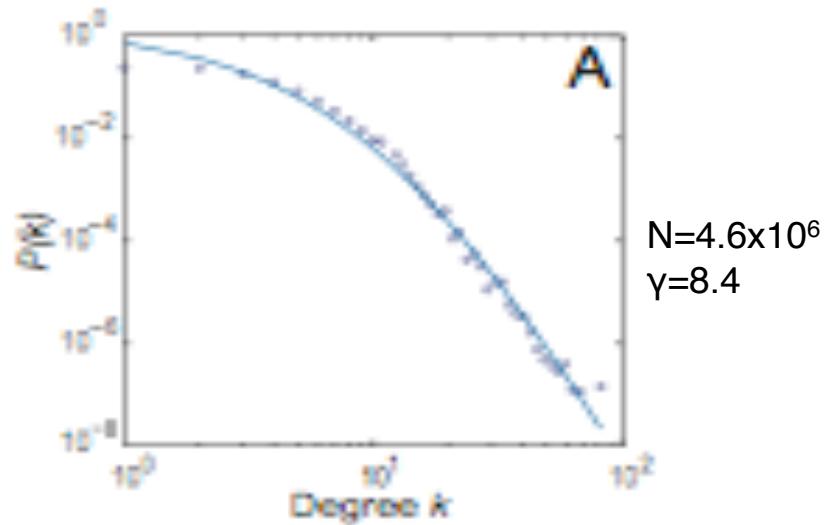
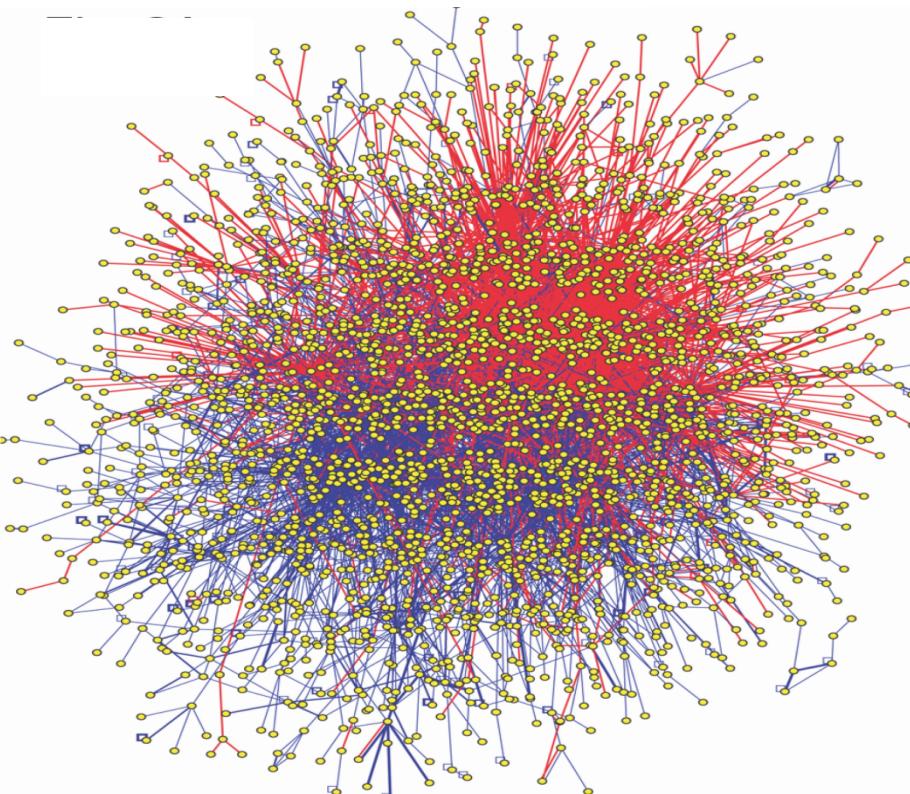


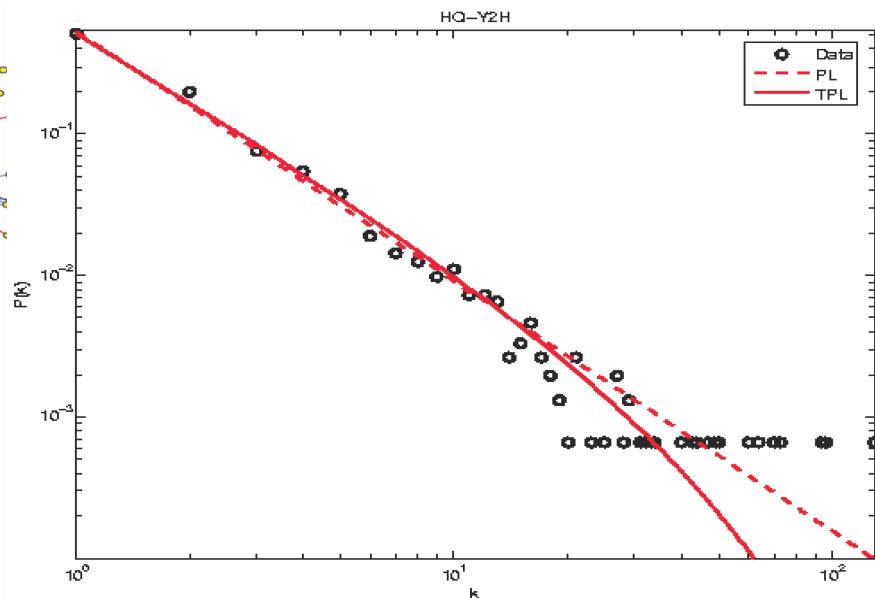
Fig. 1. Characterizing the large-scale structure and the tie strengths of the mobile call graph. (A) and (B) Vertex degree (A) and tie strength distribution (B). Each distribution was fitted with  $P(x) = a(x + x_0)^{-\gamma} \exp(-x/x_0)$ , shown as a blue curve, where  $x$  corresponds to either  $k$  or  $w$ . The parameter values for the fits are  $k_0 = 10.9$ ,  $\gamma_k = 8.4$ ,  $k_c = \infty$  (A, degree), and  $w_0 = 280$ ,  $\gamma_w = 1.9$ ,  $w_c = 3.45 \times$

# **GRAFICANDO LEYES DE POTENCIA**

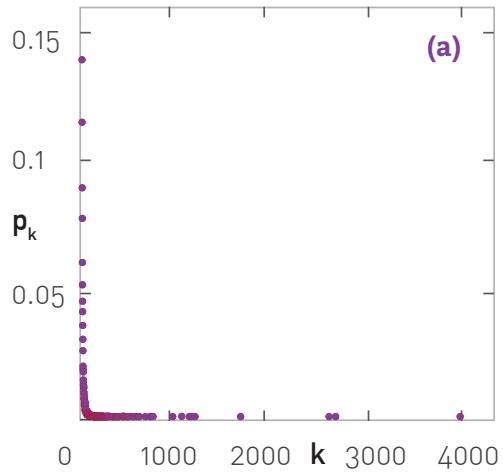
# RED DE INTERACCION HUMANA



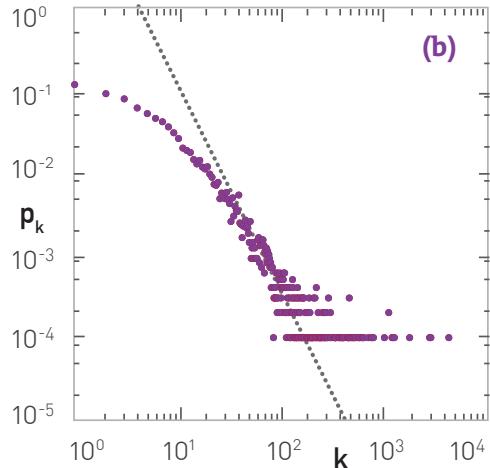
2,800 Y2H interactions  
4,100 binary LC interactions  
(HPRD, MINT, BIND, DIP, MIPS)



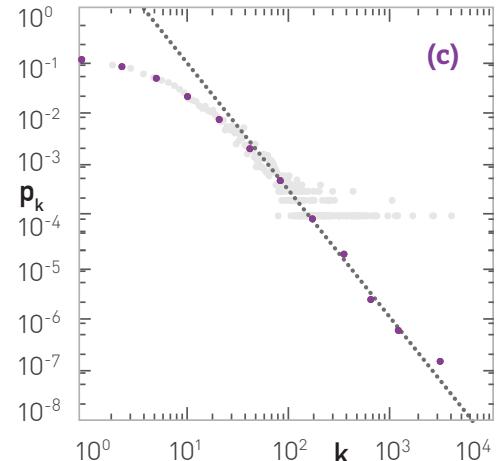
LINEAR SCALE



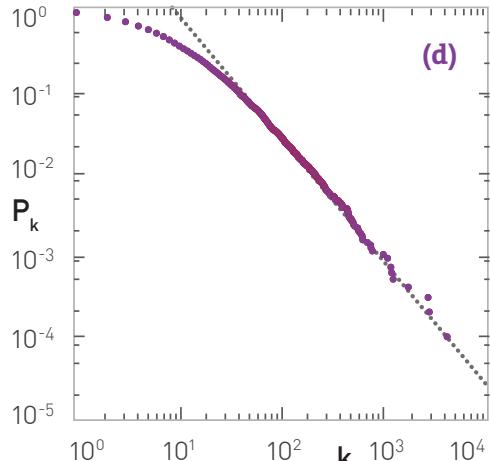
LINEAR BINNING



LOG-BINNING



CUMULATIVE



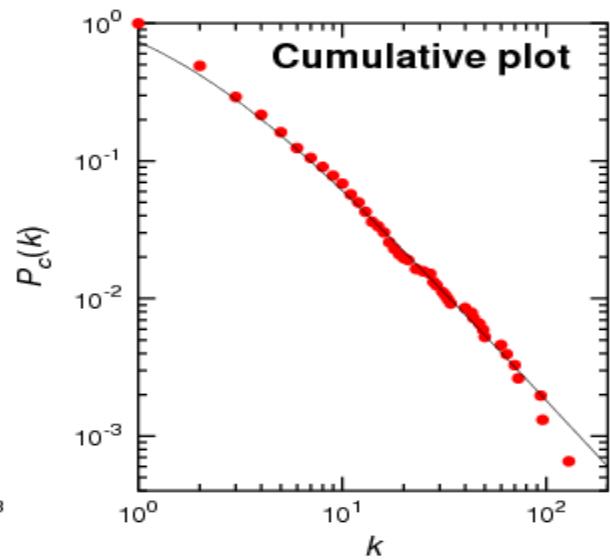
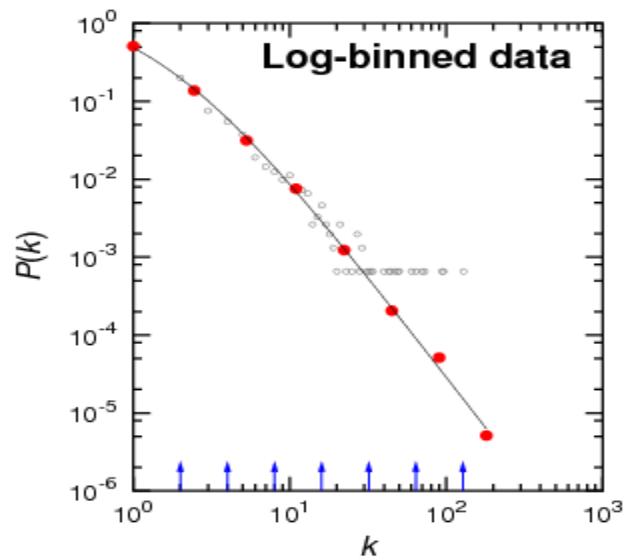
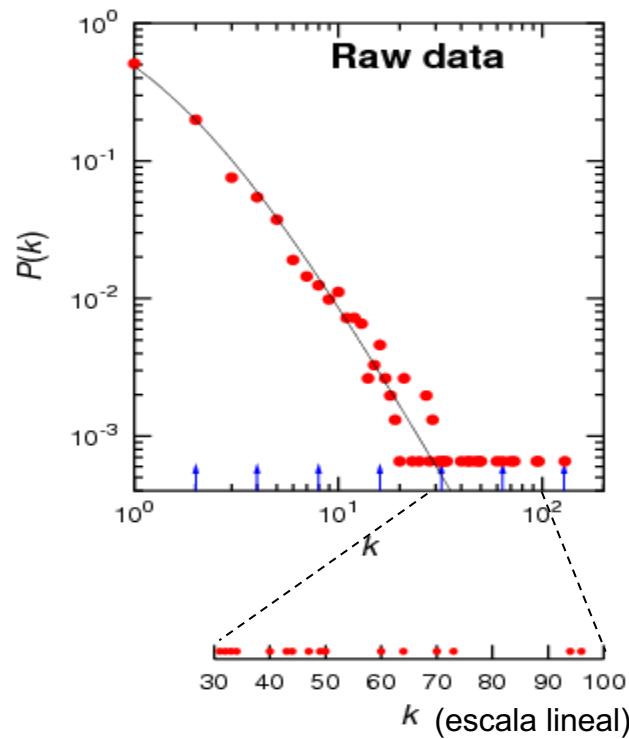
Usa gráficos log-log

Evita el bining lineal

Usa bining logarítmico

Usa la distribución acumulada

# DATA DE INTERACCION HUMANA POR RUAL ET AL.



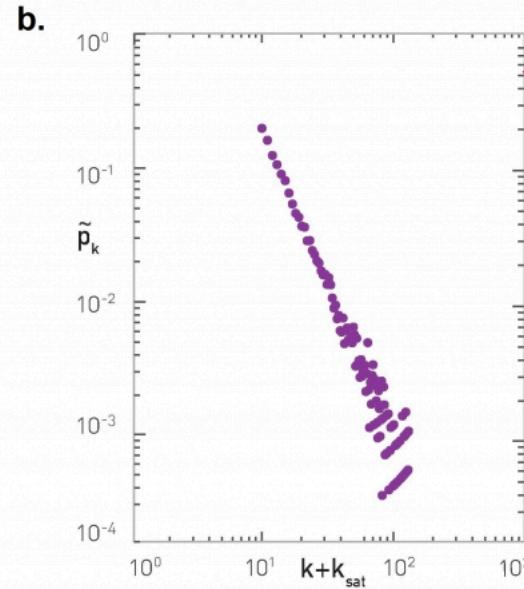
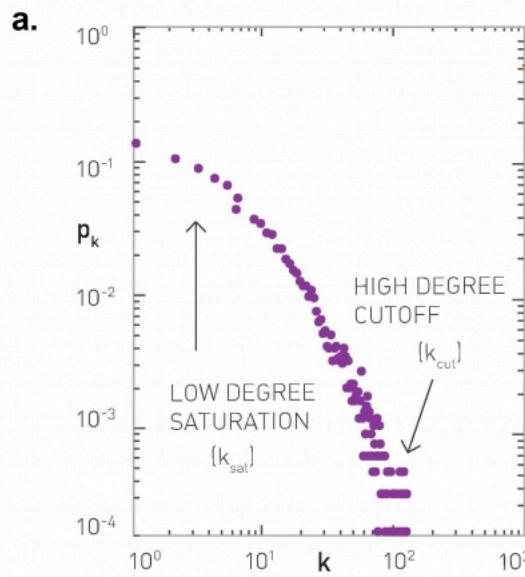
$$P(k) \sim (k+k_0)^{-\gamma}$$
$$k_0 = 1.4, \gamma = 2.6.$$

# Fallas comunes

## Reescalando la distribución del grado

En redes reales, la distribución de grados con frecuencia se desvía de una ley de potencia pura al mostrar una saturación para grados bajos y un corte en grados altos.

Al trazar la función p reescalada en  $(k + k_{sat})$ , la distribución de grados sigue una ley de potencia para todos los grados.



Dada la presencia generalizada de tales cortes, la distribución de grados se ajusta ocasionalmente a:

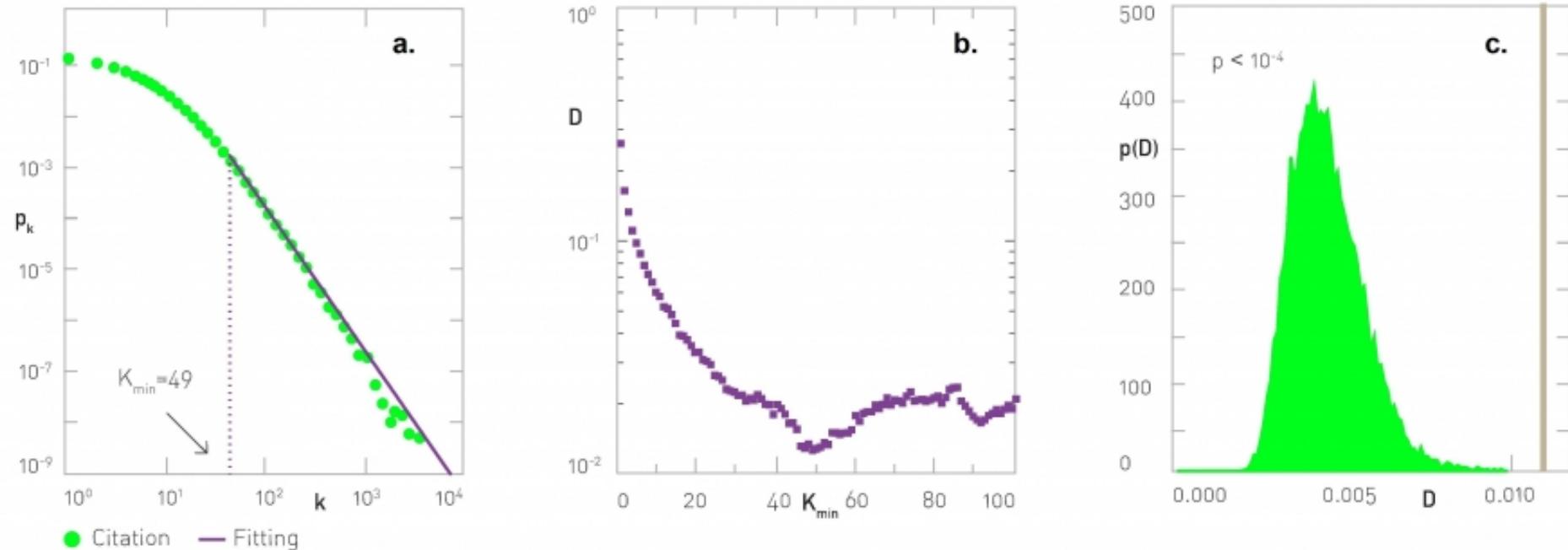
$$p_x = a(k + k_{sat})^{-\gamma} \exp\left(-\frac{k}{k_{cut}}\right) \quad (4.39)$$

Para extraer la extensión completa de la escala se grafica:

$$\tilde{p}_x = p_x \exp\left(\frac{k}{k_{cut}}\right) \quad (4.40)$$

$$\tilde{k} = k + k_{sat} \quad \tilde{p} \sim \tilde{k}^{-\gamma}$$

# Procedimiento de ajuste



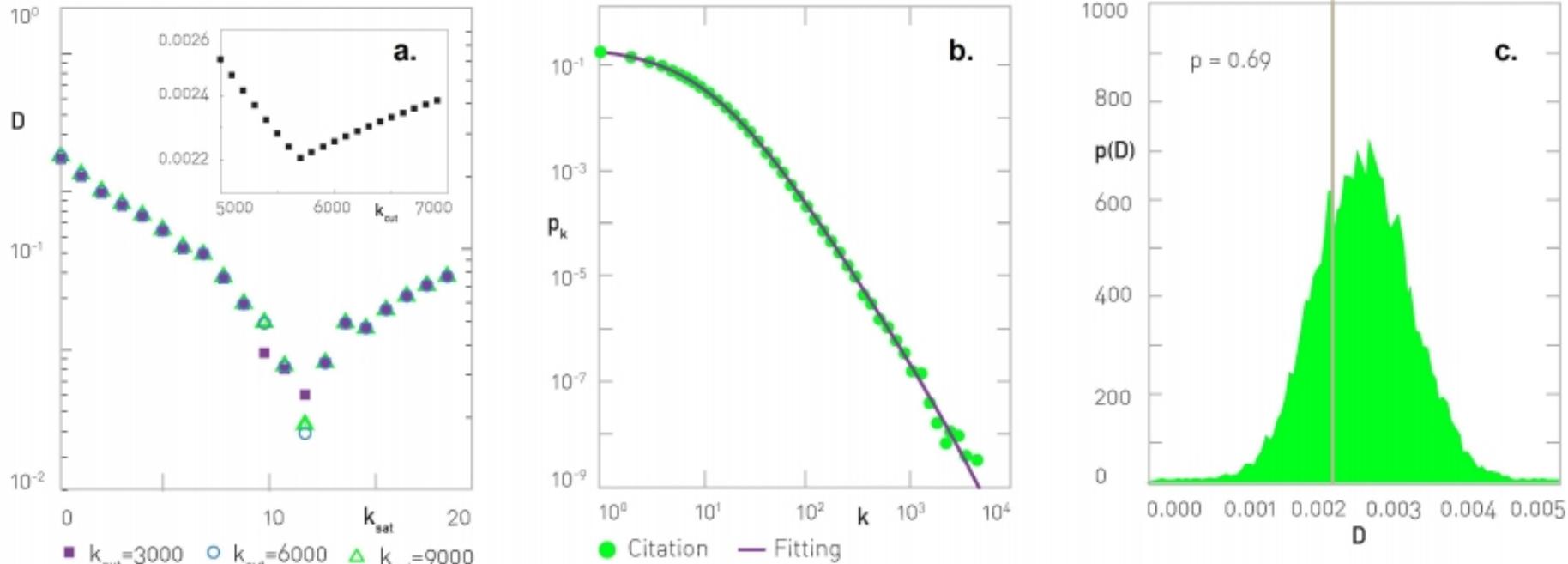
$$p_k = \frac{1}{\zeta(\gamma, K_{\min})} k^{-\gamma}$$

$$P_k = 1 - \frac{\zeta(\gamma, k)}{\zeta(\gamma, K_{\min})}$$

$$D = \max_{k \geq K_{\min}} |S(k) - P_k|$$

Estudiar el paquete powerlaws:  
<https://aaronclauset.github.io/powerlaws/>

# Procedimiento de ajuste



$$p_k = \frac{1}{\sum_{k'=1}^N (k'+k_{sat})^{-\gamma} e^{-k'/k_{cut}}} (k + k_{sat})^{-\gamma} e^{-k/k_{cut}} \quad (4.47)$$

$$\log L(\gamma | k_{sat}, k_{cut}) = \sum_{i=1}^N \log p(k_i | \gamma, k_{sat}, k_{cut})$$

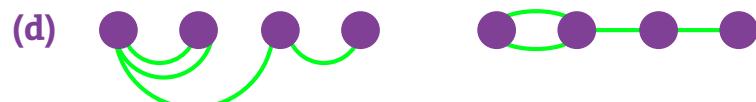
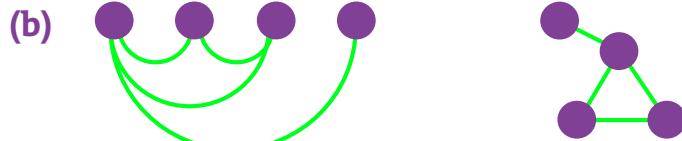
# Generando redes con un $p_k$ predefinido

# Modelo de configuración

- (1) **Secuencia de grados:** Asigne un grado a cada nodo, representado como cabos o medios enlaces. La secuencia de grados se genera analíticamente a partir de una distribución preseleccionada, o se extrae de la matriz de adyacencia de una red real. Debemos comenzar desde un número par de cabos, de lo contrario nos quedarán con cabos sin par.

- (2) **Ensamblado de la red:** seleccione aleatoriamente un par de conexiones y conectelas. Luego, elija aleatoriamente otro par de los cabos restantes y conéctelos. Este procedimiento se repite hasta que todos los cabos se emparejan. Dependiendo del orden en que se eligieron los cabos, obtenemos diferentes redes. Algunas redes incluyen ciclos (b), otras con bucles (c) o con enlaces múltiples (d). Sin embargo, el número esperado de bucles y múltiples tiende a cero en el límite.

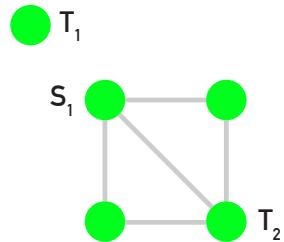
$$k_1=3 \quad k_2=2 \quad k_3=2 \quad k_4=1$$



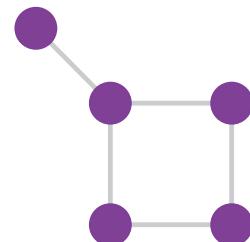
$$P_{ij} = \frac{k_i k_j}{2L - 1}$$

# Aleatorización preservando la distribución de grado

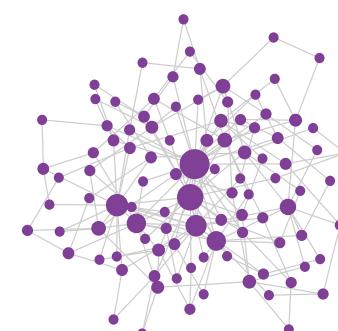
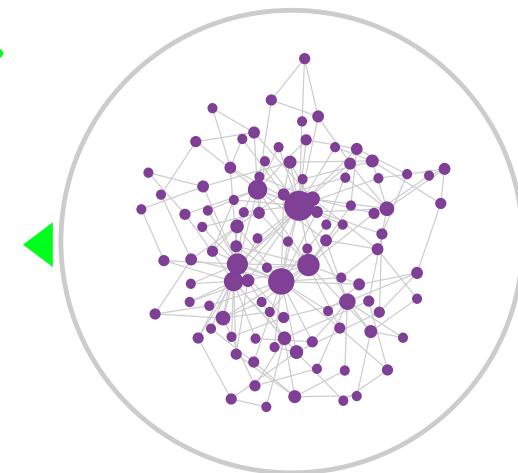
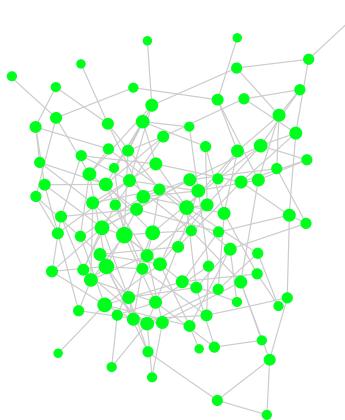
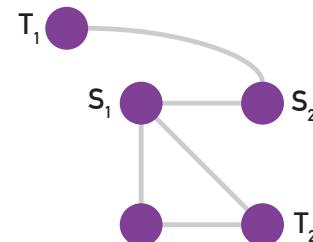
FULL  
RANDOMIZATION



ORIGINAL NETWORK



DEGREE-PRESERVING  
RANDOMIZATION

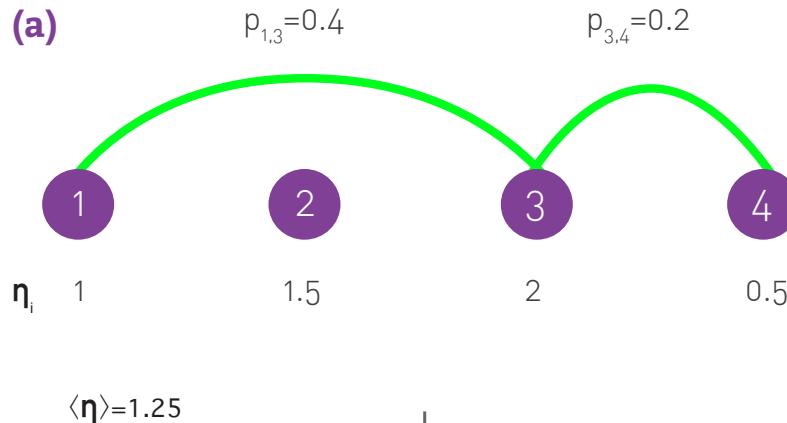


## Ejercicio Fuertemente Recomendado

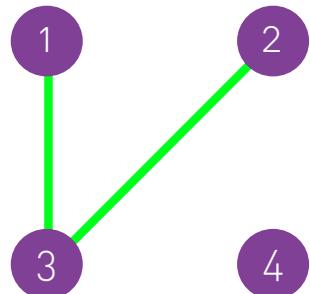
- Estudiar del paquete networkx los modelos de randomización
- `generators.degree\_seq.configuration\_model`
- [https://networkx.org/documentation/latest/reference/generated/networkx.generators.degree\\_seq.configuration\\_model.html](https://networkx.org/documentation/latest/reference/generated/networkx.generators.degree_seq.configuration_model.html)

# Modelo de parámetro oculto (revisar de manera autónoma)

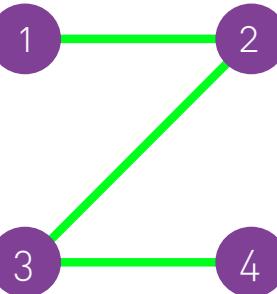
(a)



(b)



(c)



$$p(\eta_i, \eta_j) = \frac{\eta_i \eta_j}{\langle \eta \rangle N}$$

$$p_k = \int \frac{e^{-\eta} \eta^k}{k!} p(\eta) d\eta.$$

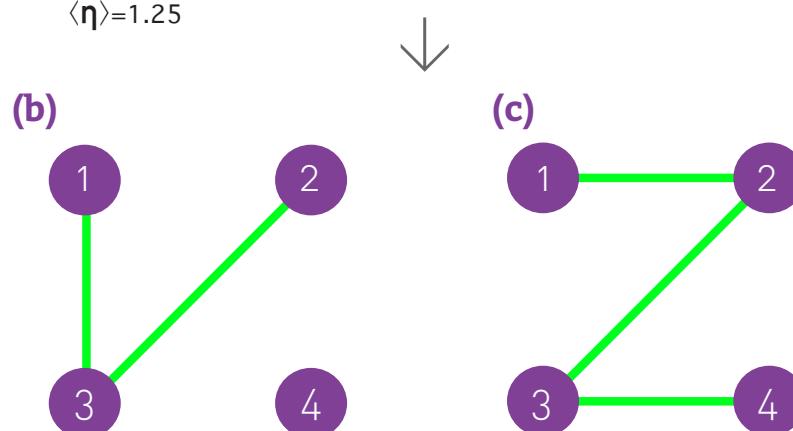
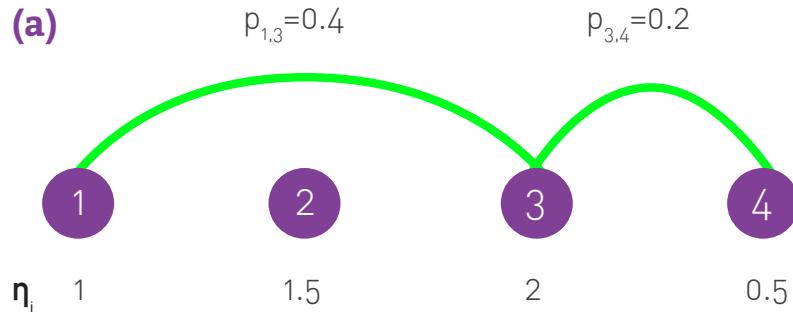
$$\{\eta_1, \eta_2, \dots, \eta_N\}$$

$$p_k = \frac{1}{N} \sum_j \frac{e^{-\eta_j} \eta_j^k}{k!}.$$

$$\eta_j = \frac{c}{i^\alpha}, i = 1, \dots, N$$

$$p_k \sim k^{-(1 + \frac{1}{\alpha})}$$

# Modelo de parámetro oculto



$$p_k = \int \frac{e^{-\eta} \eta^k}{k!} p(\eta) d\eta.$$

$$\{\eta_1, \eta_2, \dots, \eta_N\}, \quad p_k = \frac{1}{N} \sum_j \frac{e^{-\eta_j} \eta_j^k}{k!}. \quad p_k \sim k^{-(1+\frac{1}{\alpha})}$$

Comience con  $N$  nodos aislados y asigne a cada nodo un "parámetro oculto"  $\eta$ , que puede seleccionarse al azar de una distribución  $p(\eta)$ . A continuación conectamos cada par de nodos con probabilidad.

$$p(\eta_i, \eta_j) = \frac{\eta_i \eta_j}{\langle \eta \rangle N}$$

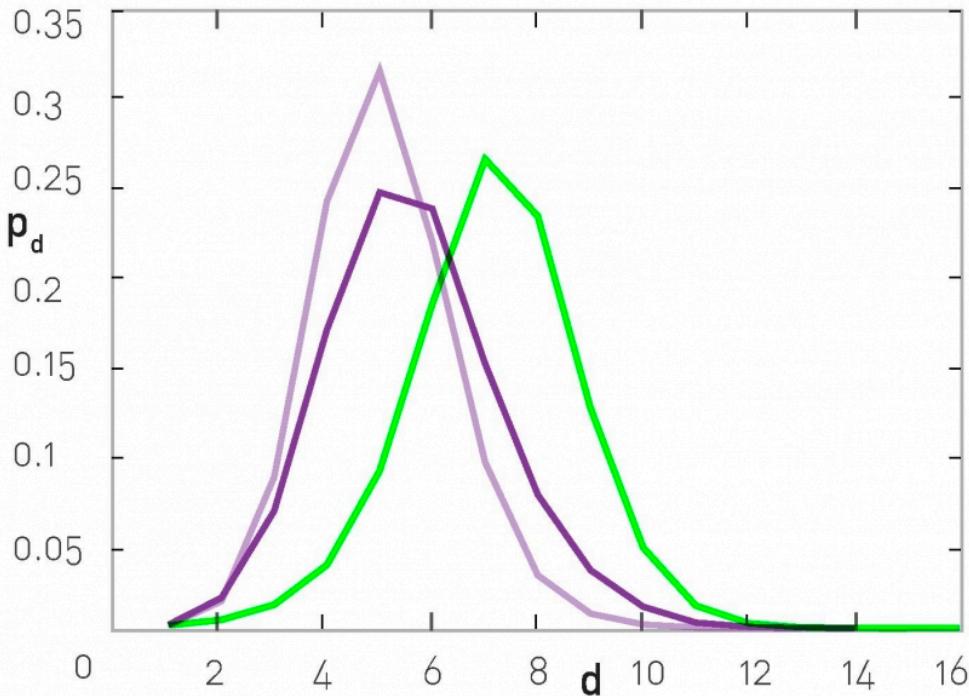
Por ejemplo, la figura muestra la probabilidad de conectar los nodos (1,3) y (3,4).

Después de conectar los nodos, terminamos con las redes mostradas en (b) o (c), que representan dos realizaciones independientes generadas por la misma secuencia de parámetros ocultos (a). El número esperado de enlaces en la red obtenida es

$$L = \frac{1}{2} \sum_N^{i,j} \frac{\eta_i \eta_j}{\langle \eta \rangle N} = \frac{1}{2} \langle \eta \rangle N$$

$$\eta_j = \frac{c}{i^\alpha}, i = 1, \dots, N$$

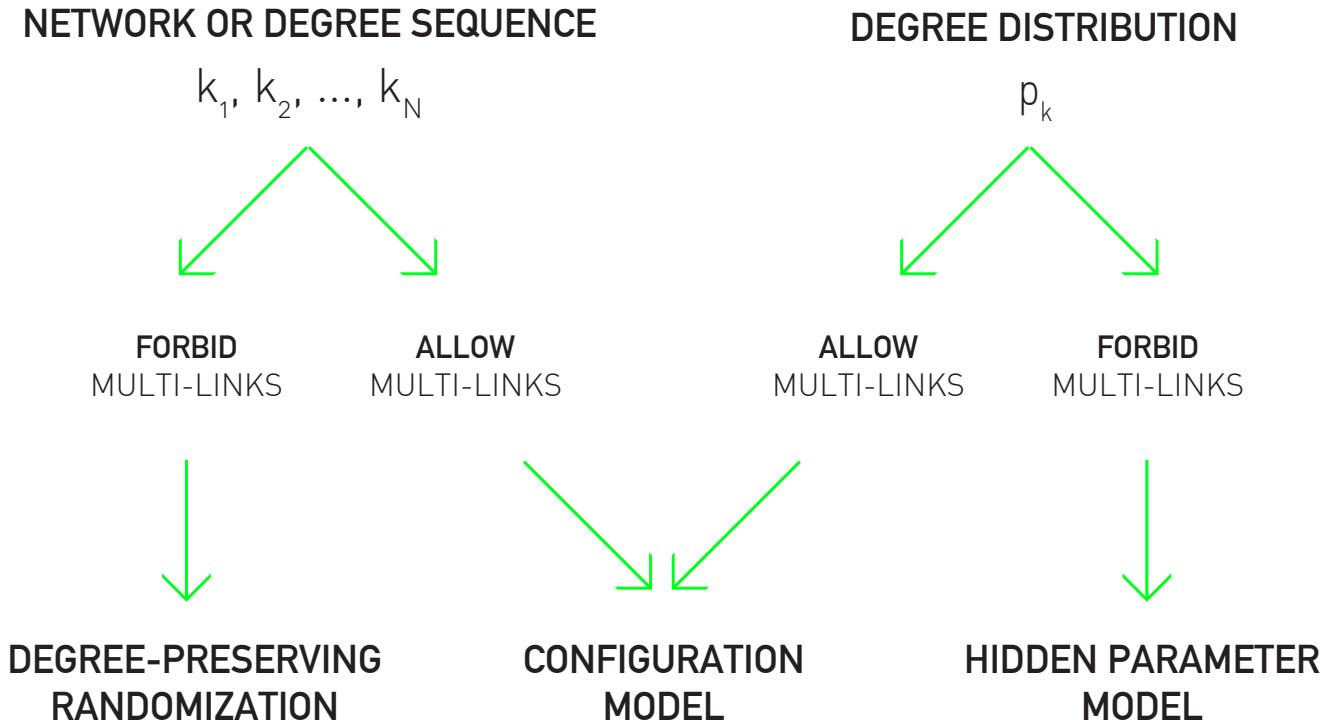
# Caso de estudio: Red PPI



We have:  $\langle d \rangle = 5.61 \pm 1.64$  (original),  $\langle d \rangle = 7.13 \pm 1.62$  (full randomization),  $\langle d \rangle = 5.08 \pm 1.34$  (degree-preserving randomization).

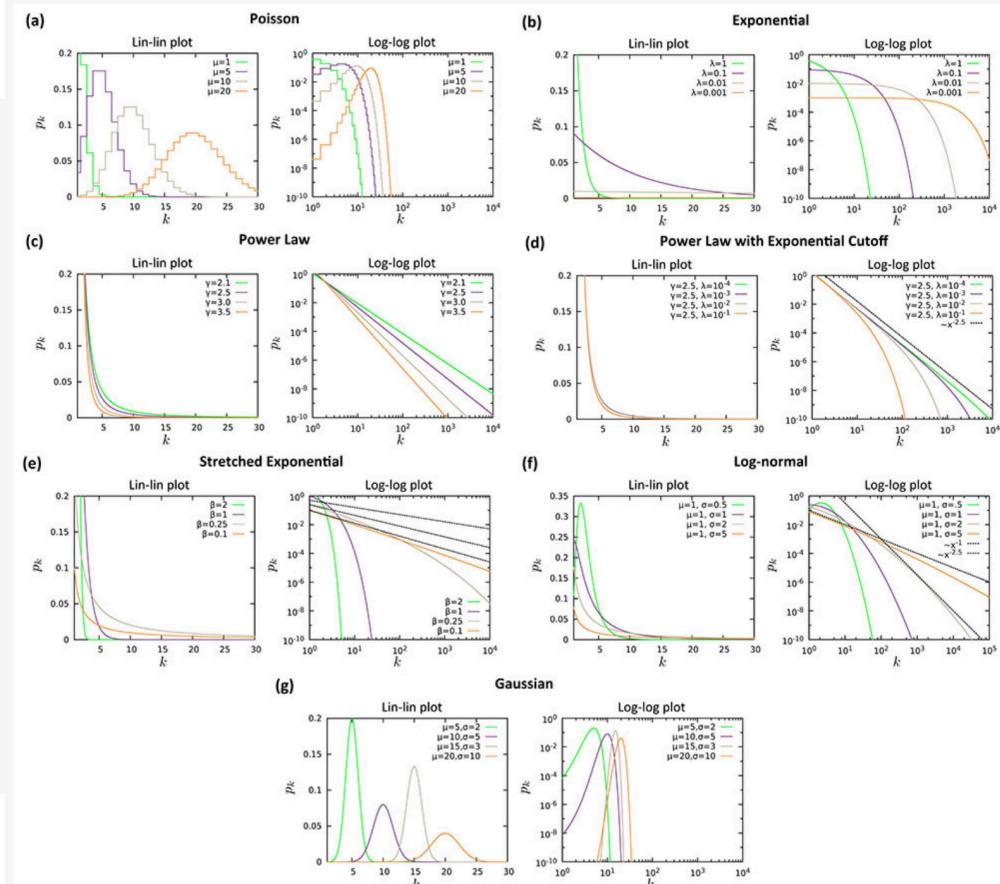
- Original network
- Degree preserving randomization
- Full randomization

# Arbol de decisión



# Tipos de distribuciones en redes

NAME	$p_k/p(x)$	$\langle x \rangle$	$\langle x^2 \rangle$
Poisson (discrete)	$e^{-\mu}\mu^x/x!$	$\mu$	$\mu(1+\mu)$
Exponential (discrete)	$(1-e^{-\lambda})e^{-\lambda x}$	$1/(e^\lambda - 1)$	$(e^\lambda + 1)/(e^\lambda - 1)^2$
Exponential (continuous)	$\lambda e^{-\lambda x}$	$1/\lambda$	$2/\lambda^2$
Power law (discrete)	$x^{-\alpha}/\zeta(\alpha)$	$\begin{cases} \zeta(\alpha-2)/\zeta(\alpha), & \text{if } \alpha > 2 \\ \infty, & \text{if } \alpha \leq 1 \end{cases}$	$\begin{cases} \zeta(\alpha-1)/\zeta(\alpha), & \text{if } \alpha > 1 \\ \infty, & \text{if } \alpha \leq 2 \end{cases}$
Power law (continuous)	$\alpha x^{-\alpha}$	$\begin{cases} \alpha/(\alpha-1), & \text{if } \alpha > 2 \\ \infty, & \text{if } \alpha \leq 1 \end{cases}$	$\begin{cases} \alpha/(\alpha-2), & \text{if } \alpha > 1 \\ \infty, & \text{if } \alpha \leq 2 \end{cases}$
Power law with cutoff (continuous)	$\frac{\lambda^{1-\alpha}}{\Gamma(1-\alpha)} x^{-\alpha} e^{-\lambda x}$	$\lambda^{-1} \frac{\Gamma(2-\alpha)}{\Gamma(1-\alpha)}$	$\lambda^{-2} \frac{\Gamma(3-\alpha)}{\Gamma(1-\alpha)}$
Stretched exponential	$\beta \lambda^\beta x^{\beta-1} e^{-(\lambda x)^\beta}$	$\lambda^{-1} \Gamma(1+\beta^{-1})$	$\lambda^{-2} \Gamma(1+2\beta^{-1})$
Log-normal (continuous)	$\frac{1}{x\sqrt{2\pi\sigma^2}} e^{-(\ln x - \mu)^2/(2\sigma^2)}$	$e^{\mu + \sigma^2/2}$	$e^{2(\mu + \sigma^2)}$
Normal (continuous)	$\frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x - \mu)^2/(2\sigma^2)}$	$\mu$	$\mu^2 + \sigma^2$





Para generar redes libres de escala, Laszlo Barabási y Réka Albert sugirieron el algoritmo

# **Barabási-Albert model of Growth and Preferential Attachment**

# **Growth and Preferential Attachment**

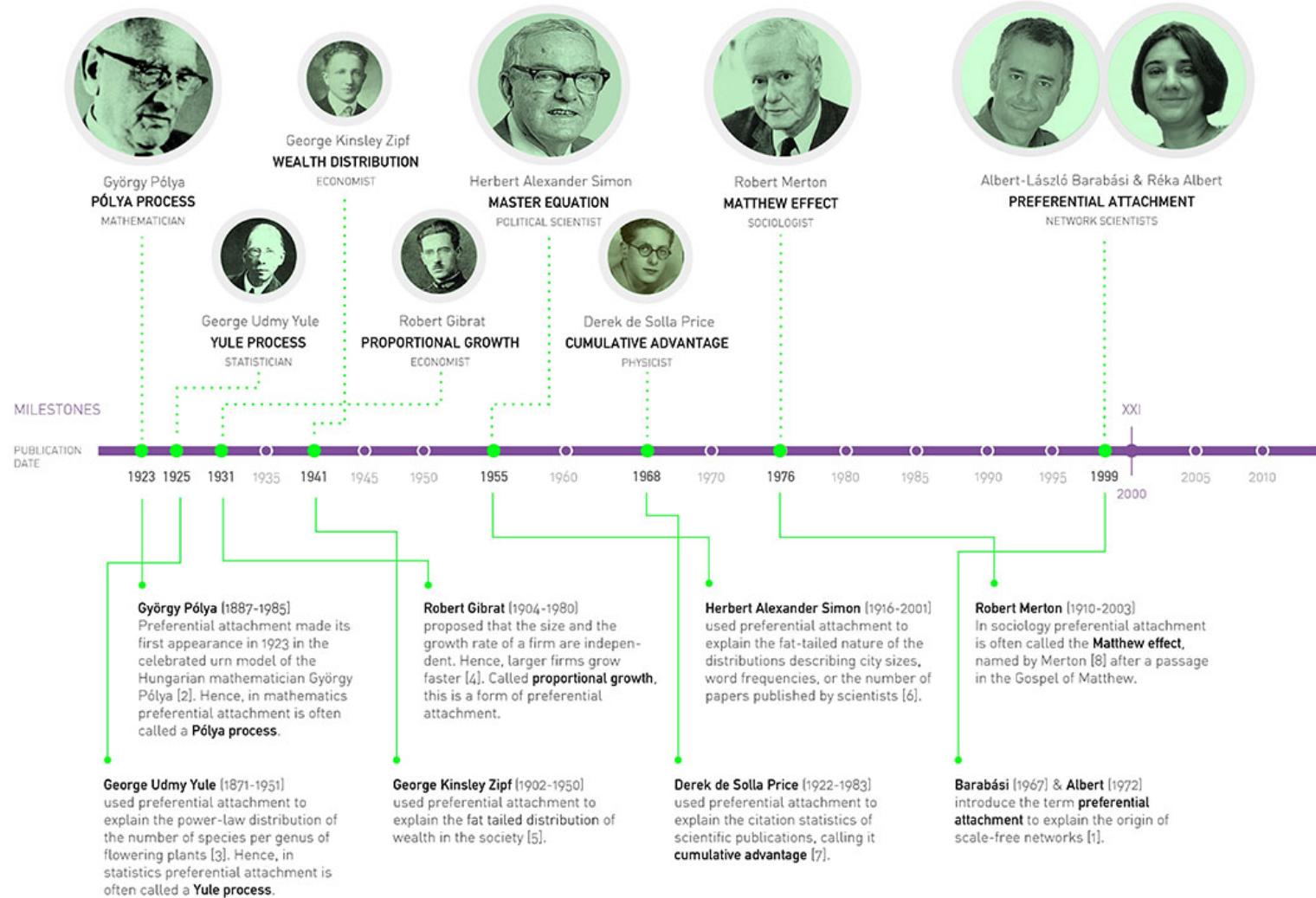


Crecimiento significa que tendremos una red en la que el número de nodos crecerá iterativamente en el tiempo hasta que alcancemos el tamaño de población deseado.



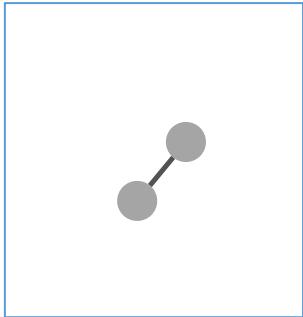
Los nuevos nodos se unirán a los nodos existentes, pero lo harán preferentemente a los más conectados.

# Prefential Attachment



# Growth and Preferential Attachment

## Step 0



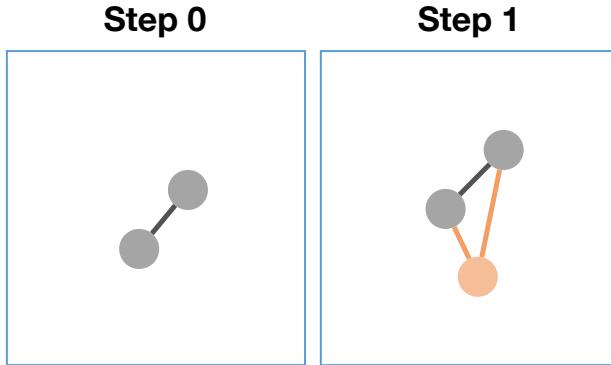
El algoritmo consta de los siguientes pasos:

- Comenzando con  $m_0$  nodos, los enlaces entre los cuales se eligen arbitrariamente, siempre que cada nodo tenga al menos un enlace.
- Para  $N-m_0$  iteraciones, agregue una nueva que se adjunte con  $m (< m_0)$  enlaces a nodos preexistentes con probabilidad:

$$\Pi(k_i) = \frac{k_i}{\sum_j k_j}$$

la probabilidad de que un enlace del nuevo nodo se conecte al nodo  $i$  depende del grado  $k_i$  como

# Growth and Preferential Attachment



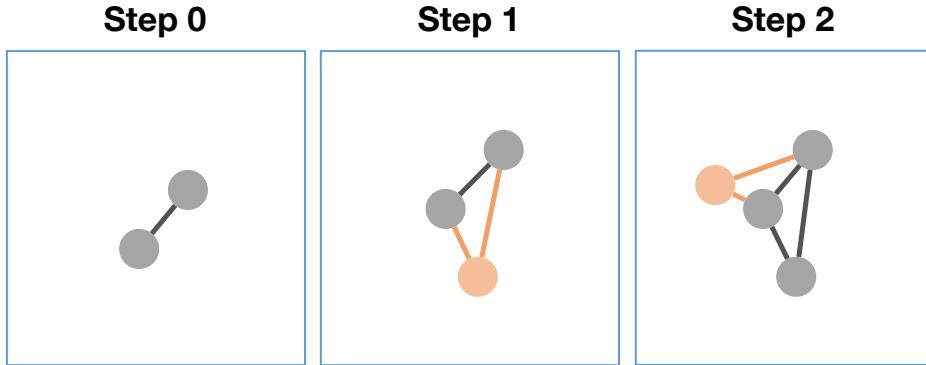
El algoritmo consta de los siguientes pasos:

- Comenzando con  $m_0$  nodos, los enlaces entre los cuales se eligen arbitrariamente, siempre que cada nodo tenga al menos un enlace.
- Para  $N-m_0$  iteraciones, agregue una nueva que se adjunte con  $m$  ( $< m_0$ ) enlaces a nodos preexistentes con probabilidad:

$$\Pi(k_i) = \frac{k_i}{\sum_j k_j}$$

la probabilidad de que un enlace del nuevo nodo se conecte al nodo  $i$  depende del grado  $k_i$  como

# Growth and Preferential Attachment



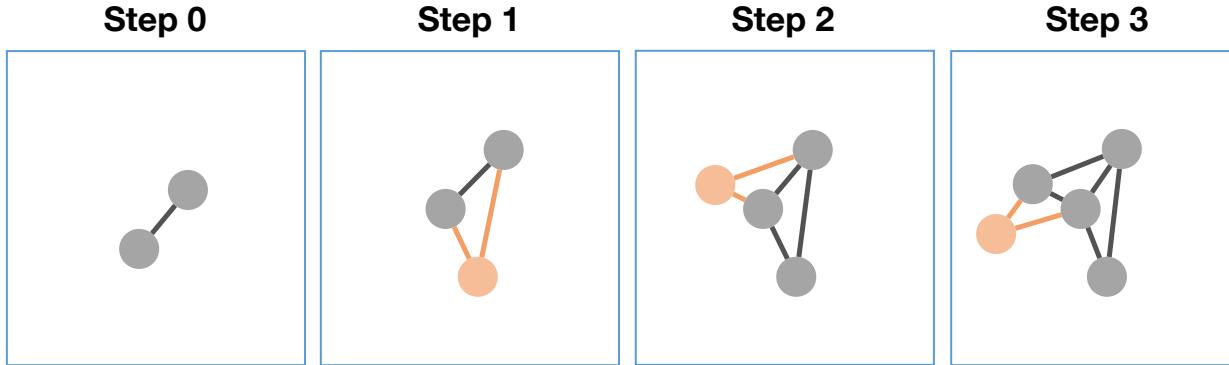
El algoritmo consta de los siguientes pasos:

- Comenzando con  $m_0$  nodos, los enlaces entre los cuales se eligen arbitrariamente, siempre que cada nodo tenga al menos un enlace.
- Para  $N-m_0$  iteraciones, agregue una nueva que se adjunte con  $m$  ( $< m_0$ ) enlaces a nodos preexistentes con probabilidad:

$$\Pi(k_i) = \frac{k_i}{\sum_j k_j}$$

la probabilidad de que un enlace del nuevo nodo se conecte al nodo  $i$  depende del grado  $k_i$  como

# Growth and Preferential Attachment



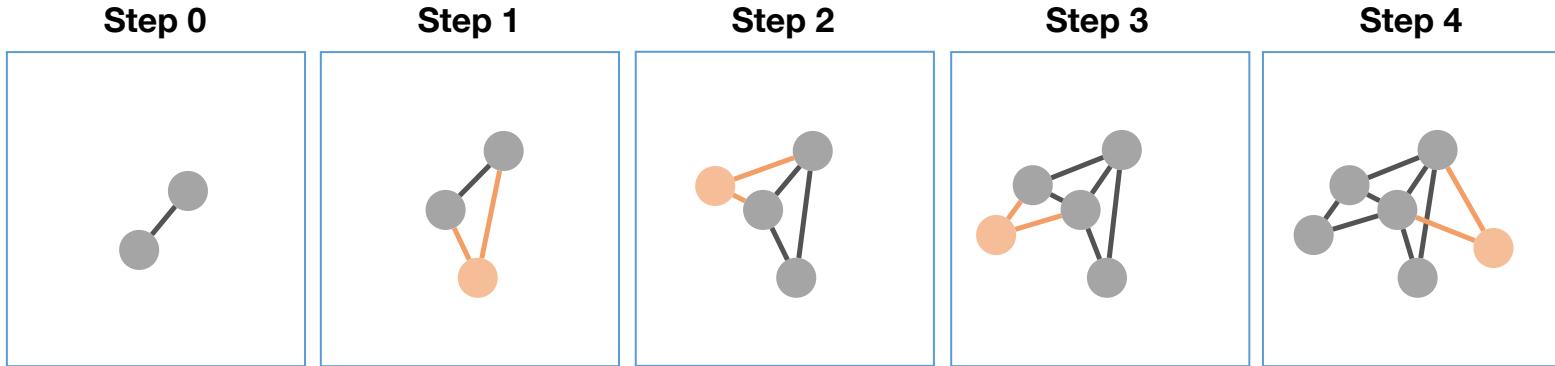
El algoritmo consta de los siguientes pasos:

- Comenzando con  $m_0$  nodos, los enlaces entre los cuales se eligen arbitrariamente, siempre que cada nodo tenga al menos un enlace.
- Para  $N-m_0$  iteraciones, agregue una nueva que se adjunte con  $m$  ( $< m_0$ ) enlaces a nodos preexistentes con probabilidad:

$$\Pi(k_i) = \frac{k_i}{\sum_j k_j}$$

la probabilidad de que un enlace del nuevo nodo se conecte al nodo  $i$  depende del grado  $k_i$  como

# Growth and Preferential Attachment



El algoritmo consta de los siguientes pasos:

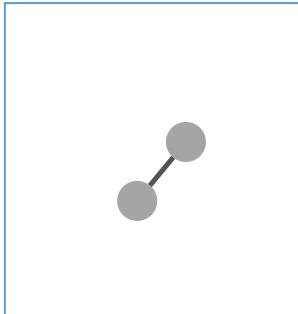
- Comenzando con  $m_0$  nodos, los enlaces entre los cuales se eligen arbitrariamente, siempre que cada nodo tenga al menos un enlace.
- Para  $N-m_0$  iteraciones, agregue una nueva que se adjunte con  $m$  ( $< m_0$ ) enlaces a nodos preexistentes con probabilidad:

$$\Pi(k_i) = \frac{k_i}{\sum_j k_j}$$

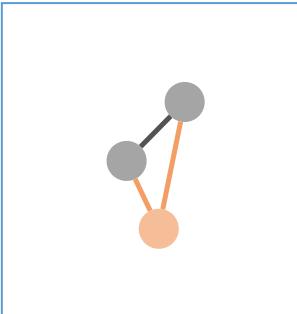
la probabilidad de que un enlace del nuevo nodo se conecte al nodo  $i$  depende del grado  $k_i$  como

# Growth and Preferential Attachment

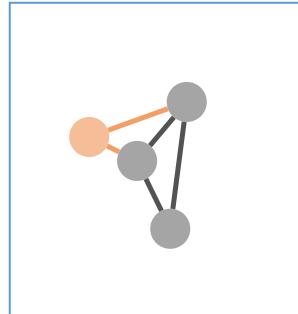
Step 0



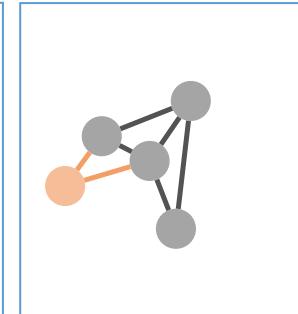
Step 1



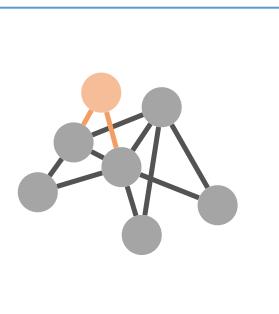
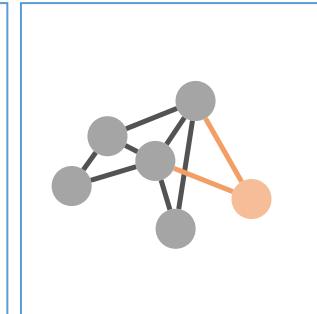
Step 2



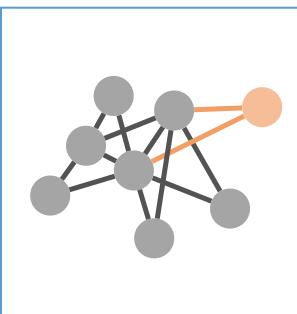
Step 3



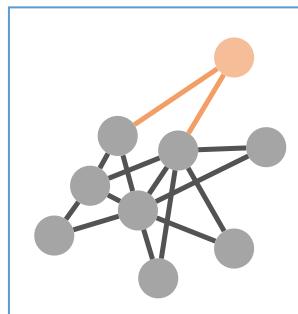
Step 4



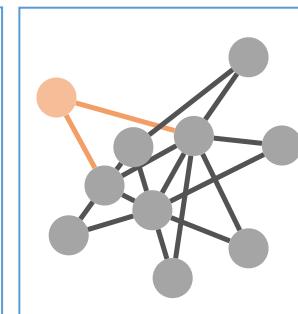
Step 5



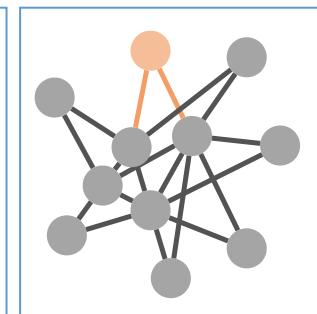
Step 6



Step 7

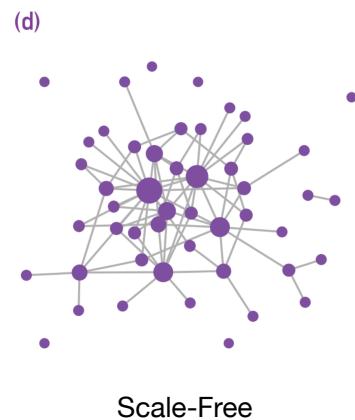
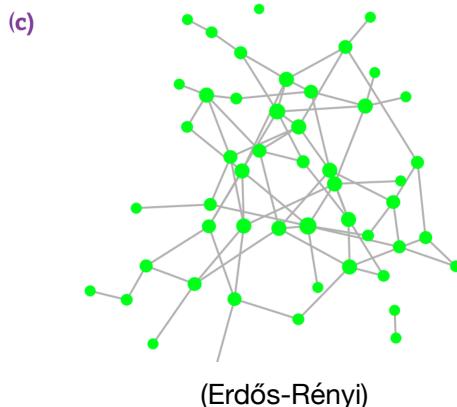
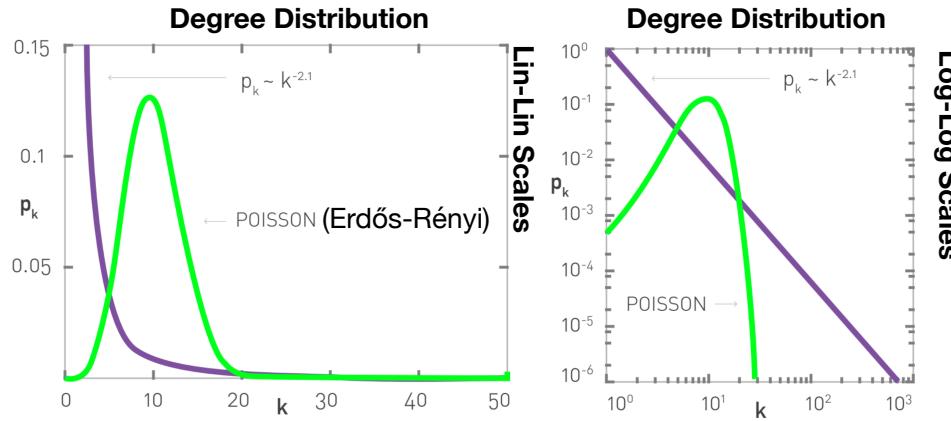


Step 8



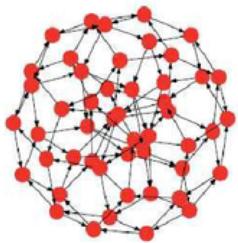
Step 9

# Scale-Free and Random Networks

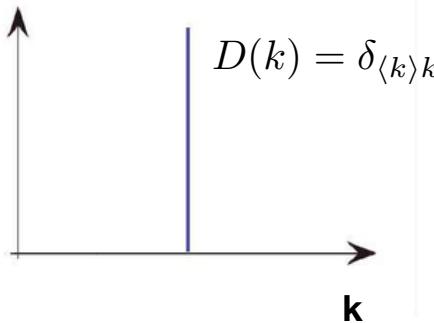


¡Las redes sin escala tienen una distribución de grados que sigue una ley de potencia!

## Lattices

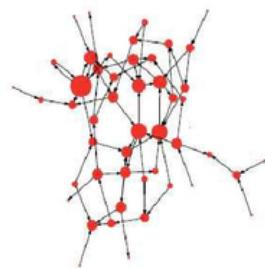


$D(k)$

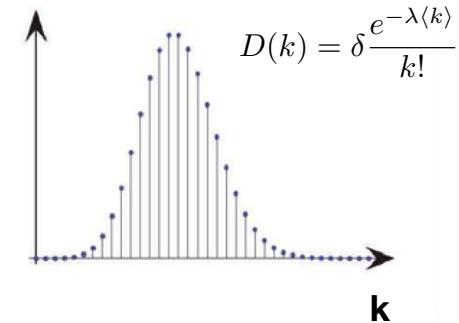


Regular

## Erdős-Rényi



$D(k)$

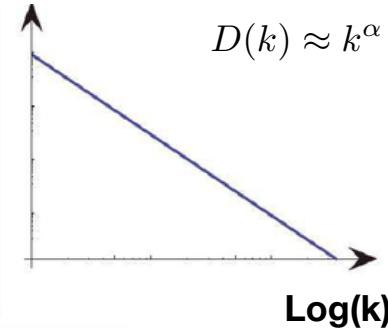


Random

## Barabási-Albert



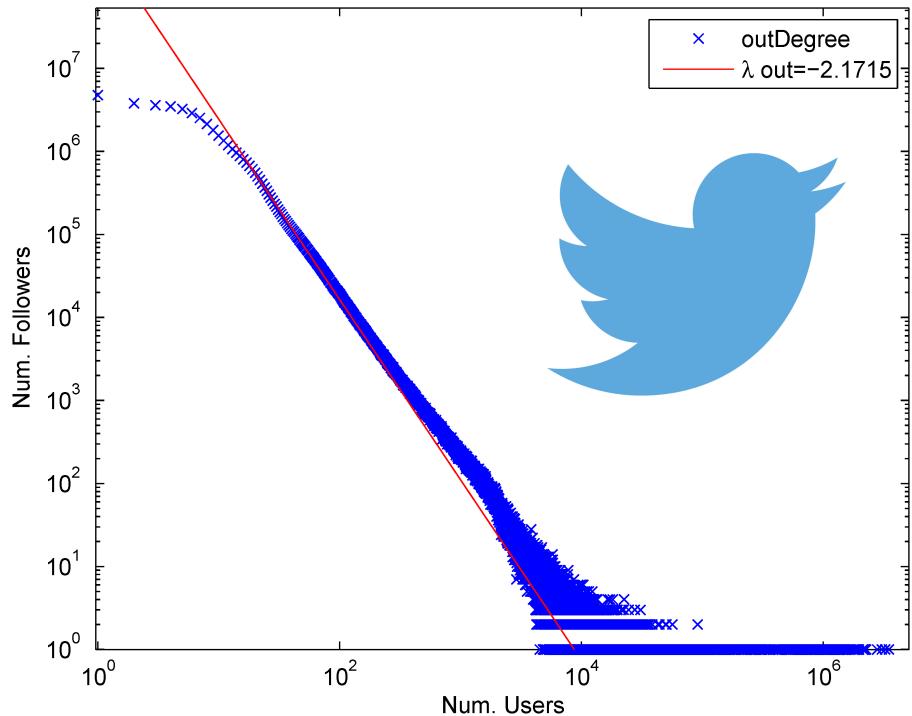
$\text{Log}(D(k))$



Scale Free

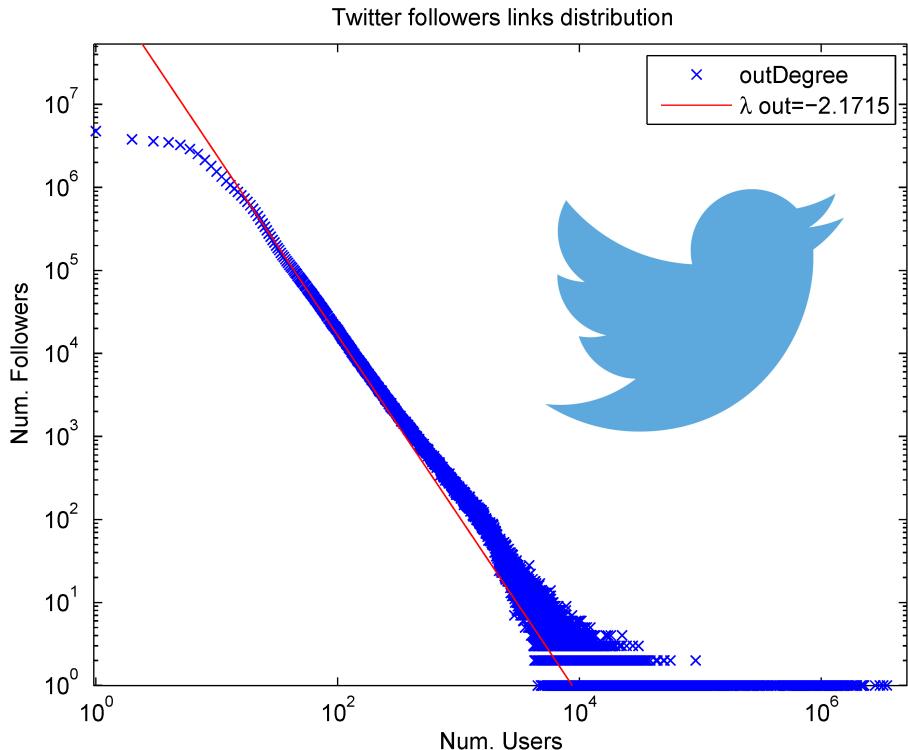
# Scale Free Networks

Twitter followers links distribution



- La ubicuidad de la propiedad sin escala sugiere algún mecanismo subyacente fundamental;
- Crecimiento y apego preferencial:
  - Crecimiento: se agregan nodos a la red con el tiempo;
  - Apego preferencial: estos nodos prefieren vincularse a nodos altamente conectados;
- El modelo simple de Barabási-Albert muestra que el crecimiento y el apego preferencial conducen a redes sin escala;
- Los ricos se hacen más ricos: los nodos altamente conectados tienden a seguir adquiriendo más conexiones
- Resultado claro del apego preferencial
  - En última instancia, conduce a una propiedad sin escala
  - Ventaja del primero: los nodos que se agregan antes tienen grados más altos

# ¿Es el modelo B.A. un buen modelo?



- Captura la intuición de que las redes reales crecen y que los nuevos nodos prefieren conectarse a nodos altamente conectados;
- Predice el comportamiento sin escala observado en redes reales;
- Que la preferencia sea directamente proporcional al número de enlaces parece simplista;
- La predicción de que los primeros nodos siempre están más conectados parece extraña;

## Posibles extensiones:

- apego preferencial sublineal/superlineal
- métricas de fitness para nodos
- eliminación de nodos (edad fuera de la red)
- modelo mejor vinculado con la dinámica específica en cuestión.

# “There is no general theory of networks”

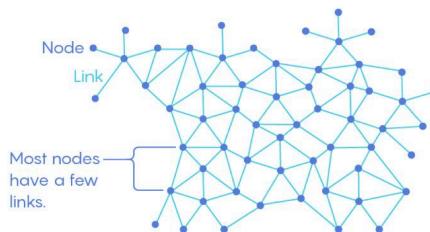
## To Be or Not to Be Scale-Free

Scientists study complex networks by looking at the distribution of the number of links (or “degree”) of each node.

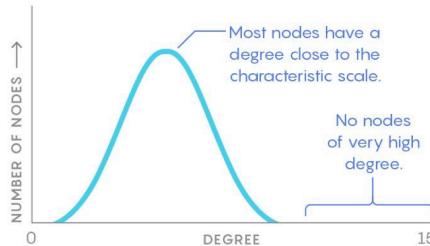
Some experts see so-called scale-free networks everywhere. But a new study suggests greater diversity in real-world networks.

### Random Network

Randomly connected networks have nodes with similar degrees. There are no (or virtually no) “hubs”—nodes with many times the average number of links.

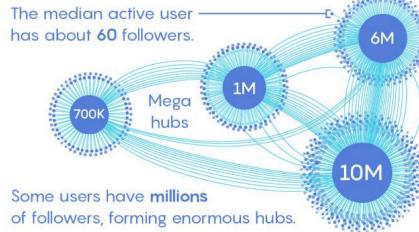


The distribution of degrees is shaped roughly like a bell curve that peaks at the network’s “characteristic scale.”

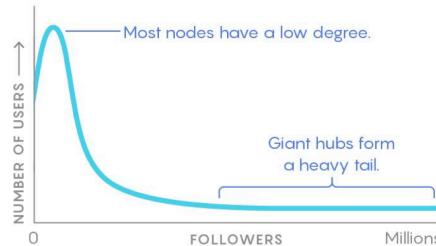


### Twitter's Scale-Free Network

Most real-world networks of interest are not random. Some nonrandom networks have massive hubs with vastly higher degrees than other nodes.

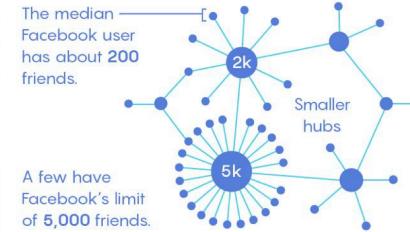


The degrees roughly follow a power law distribution that has a “heavy tail.” The distribution has no characteristic scale, making it scale-free.

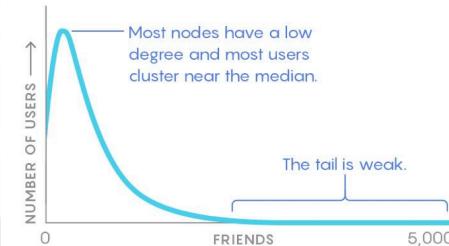


### Facebook's In-Between Network

Researchers have found that most nonrandom networks are not strictly scale-free. Many have a weak heavy tail and a rough characteristic scale.



This network has fewer and smaller hubs than in a scale-free network. The distribution of nodes has a scale and does not follow a pure power law.



[https://www.quantamagazine.org/scant-evidence-of-power-laws-found-in-real-world-networks-20180215/?utm\\_content=buffer34d54&utm\\_medium=social&utm\\_source=twitter.com&utm\\_campaign=buffer](https://www.quantamagazine.org/scant-evidence-of-power-laws-found-in-real-world-networks-20180215/?utm_content=buffer34d54&utm_medium=social&utm_source=twitter.com&utm_campaign=buffer)