

Redes y Sistemas Complejos



Cristian Candia, Ph.D.

Universidad del Desarrollo (UDD), Chile

Director at the Computational Research in Social Sciences Lab

Associate Professor, Data Science Institute, School of Engineering

Northwestern University, United States.

External Faculty, Northwestern Institute on Complex Systems (NICO)

Kellogg School of Management.

Capybara Spa (AI & Network Science for Preventive, Traceable School Coexistence Compliance)

Founder & Chief Scientific and Technological Officer (CSTO)

¿Por qué necesitamos Baselines?

⚠ Evitar Interpretaciones Espurias

Sin un modelo nulo, es fácil atribuir significado a patrones que surgen puramente por azar o densidad de la red.

El **Model Checking** permite validar si una propiedad observada es realmente una característica del sistema o un subproducto estadístico.

🔍 Identificar la "Señal"

La comparación **Real vs. Baseline** ayuda a aislar la señal no explicable simplemente por el grado o la densidad.

Criterios de Comparación:

- Coeficiente de Clustering
- Distancias Promedio
- Colas de Distribución (Tails)

Sin baseline → no hay inferencia.

Modelos generativos y nulos en redes

Medir una red real no basta. Para interpretar una métrica necesitamos una pregunta contrafactual:

¿qué valor esperaríamos ver si la red fuese aleatoria, o si solo preservara ciertas restricciones básicas?

Definiciones

Modelo generativo

Especifica un **mecanismo de formación** de la red. No solo reproduce una estructura, sino que propone una historia/mecanismo plausible de cómo se creó.

Modelo nulo

Es una **hipótesis de referencia**. Su función principal no es explicar la red, sino permitir comparar lo observado contra una expectativa controlada.

Baseline

Es cualquier modelo de comparación que usaremos como punto de referencia. Puede ser:

- un **nulo estricto**;
- un **modelo generativo estilizado**;
- o un modelo que controla una propiedad puntual, como la secuencia de grados.

Distinción clave

No todo baseline es un modelo nulo.

En esta clase:

- **ER** será el nulo aleatorio básico;
- **WS** y **BA** serán modelos generativos estilizados usados como baselines comparativos, no como nulos estrictos.

¿Por qué no basta con mirar la red real?

Muchas propiedades que parecen “interesantes” pueden emerger por razones triviales:

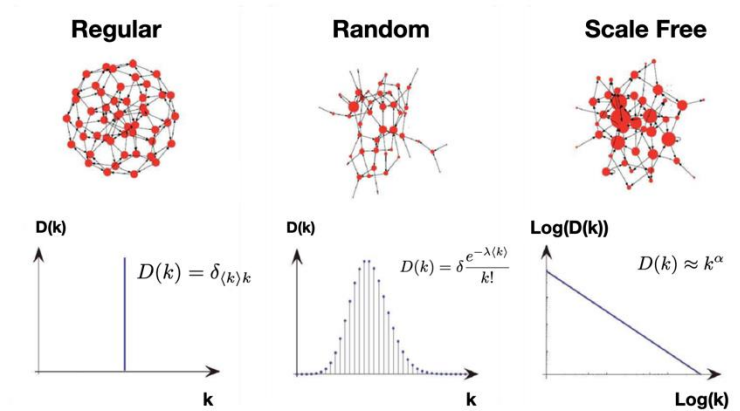
- tamaño de la red
- densidad
- distribución de grados
- proyección de una red bipartita
- simple aleatoriedad

Por eso, una métrica aislada no es evidencia suficiente.

El procedimiento correcto es:

1. medir la red real
2. generar una referencia adecuada
3. comparar
4. concluir qué parte del patrón observado **excede** lo esperable bajo ese baseline

Barabási lo formula de manera muy clara para el caso aleatorio: el random network model funciona como una **referencia fundamental**. Si una propiedad aparece también allí, puede emerger por azar, si no aparece, requiere una explicación más profunda.



Qué significa “hacer model checking” en redes

Model checking en esta unidad significa preguntar si un modelo de referencia reproduce las propiedades relevantes de la red real.

No basta con decir “mi red tiene clustering alto” o “mi red tiene hubs”.

La pregunta correcta es:

¿alto respecto de qué? ¿hubs respecto de qué control?

Disciplina práctica

- fijar explícitamente qué preserva el baseline
- simular varias realizaciones
- comparar métricas agregadas y distribuciones
- identificar la señal residual no explicada

Comparaciones típicas

- clustering real vs clustering esperado
- distancia media real vs distancia esperada
- cola de grados real vs cola simulada
- motifs/triángulos reales vs esperados
- en bipartitas: redundancia y overlap reales vs aleatorios

Tipología mínima

A. Nulo aleatorio básico

- Erdős-Rényi $G(n, p)$

B. Modelos generativos estilizados usados como baseline mecánico

- Watts-Strogatz (small-world)
- Barabási-Albert (growth + preferential attachment)

C. Nulos más útiles para análisis

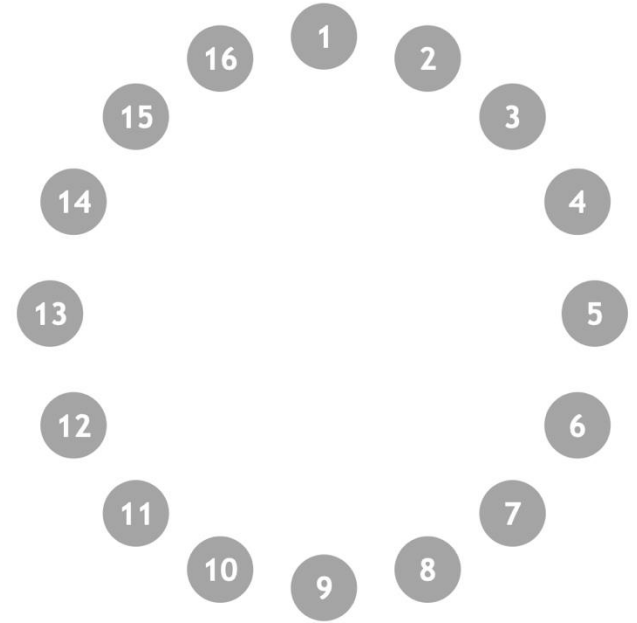
- configuration model
- degree-preserving rewiring
- stochastic block model como control mesoestructural
- nulos bipartitos que preservan tamaño y grados de ambos modos

Idea clave

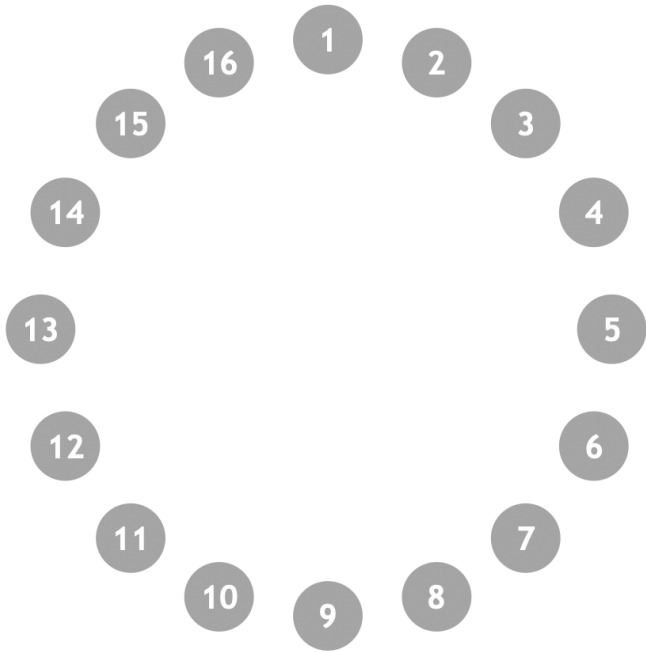
ER, WS y BA no compiten por explicar lo mismo:

- ER controla aleatoriedad homogénea
- WS controla regularidad local + shortcuts
- BA controla crecimiento y acumulación preferencial

the party algorithm

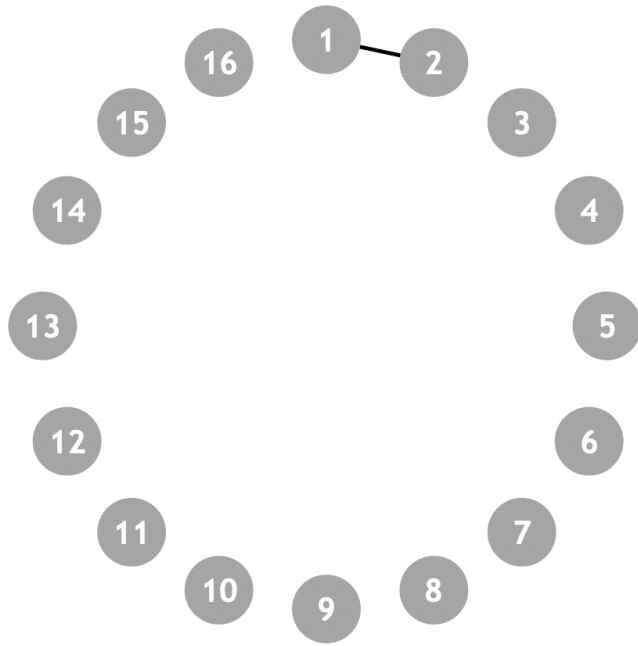


the party algorithm



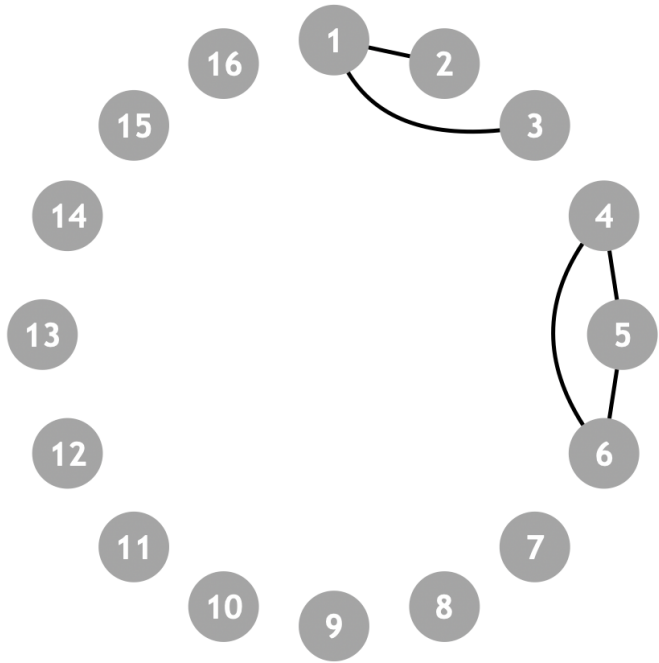
Hagamos un seguimiento de quién habla con quién

the party algorithm



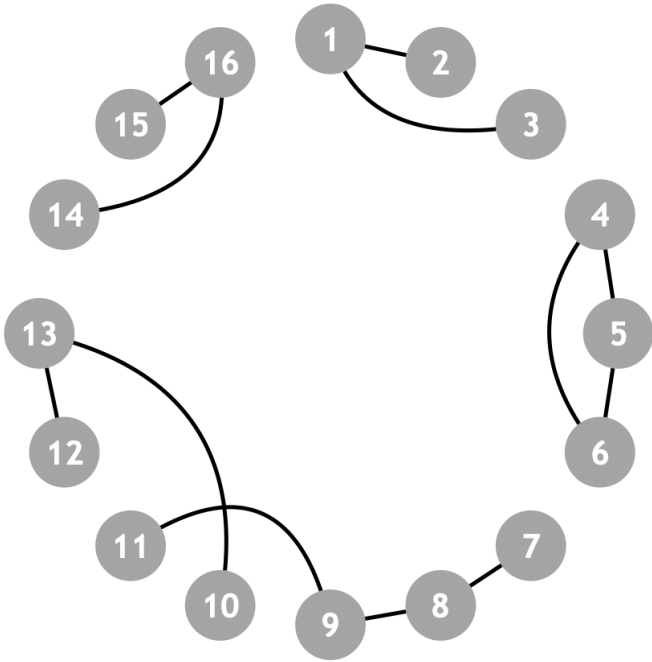
Hagamos un seguimiento de quién habla con quién

the party algorithm



Hagamos un seguimiento de quién habla con quién

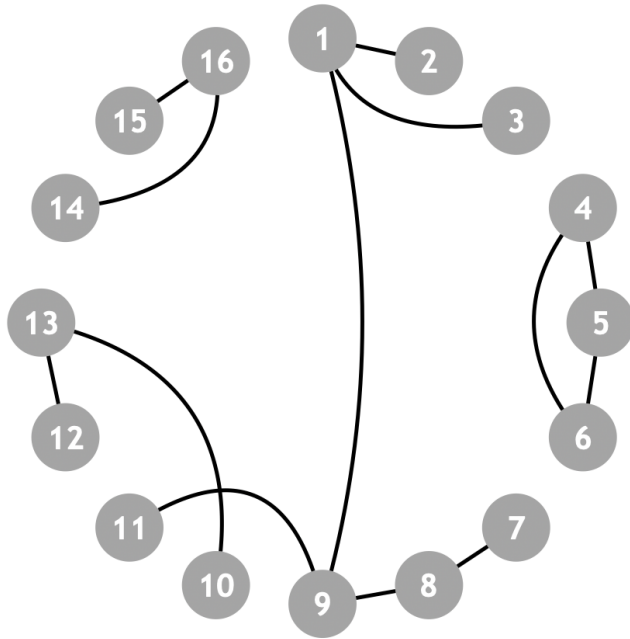
the party algorithm



Hagamos un seguimiento de quién habla con quién

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

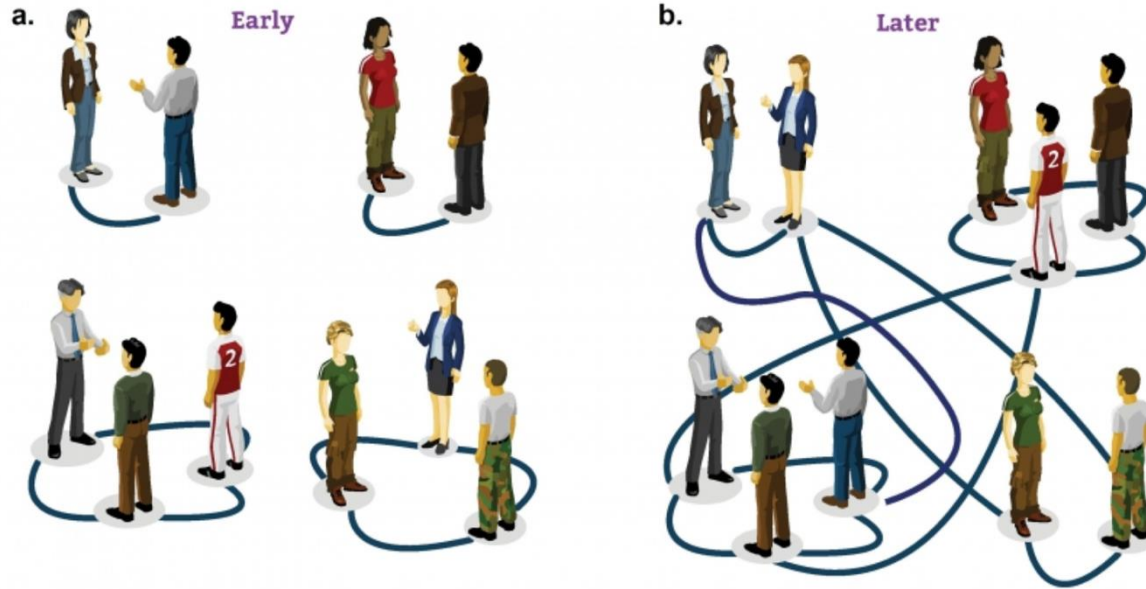
the party algorithm



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

Un algoritmo simple para redes sociales

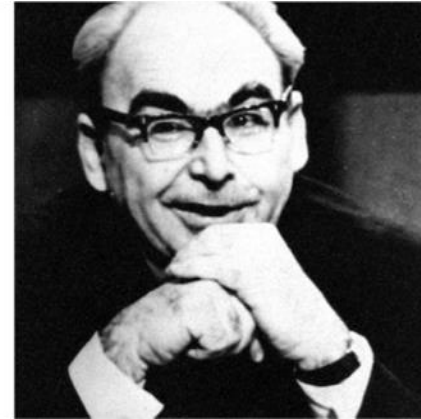
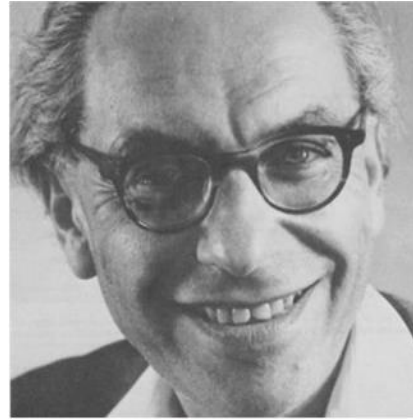


Invitados a una cena se reunen al azar y establecen relaciones sociales.

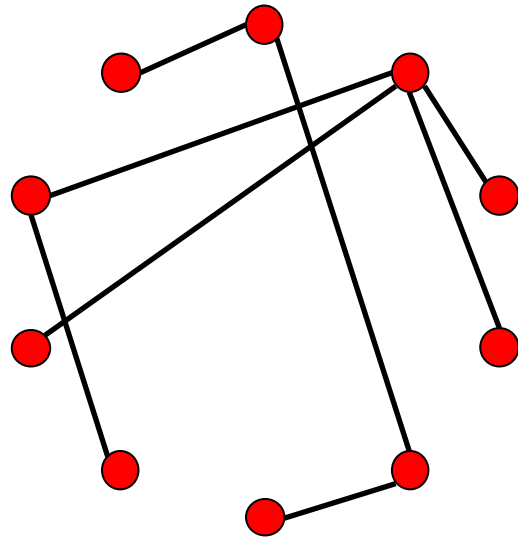
El modelo de redes aleatorias

MODELO DE REDES ALEATORIAS

Pául Erdős
(1913-1996)



Alfréd Rényi
(1921-1970)



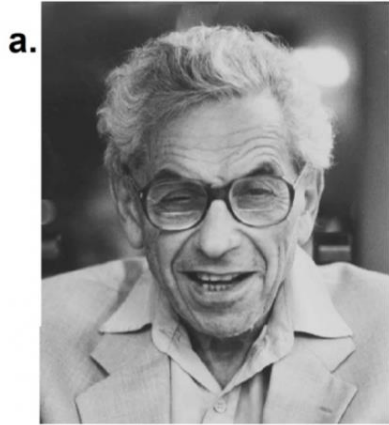
Erdős-Rényi model (1960)

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

$$p = 1/6 \quad N = 10$$

$$\langle k \rangle \sim 1.5$$

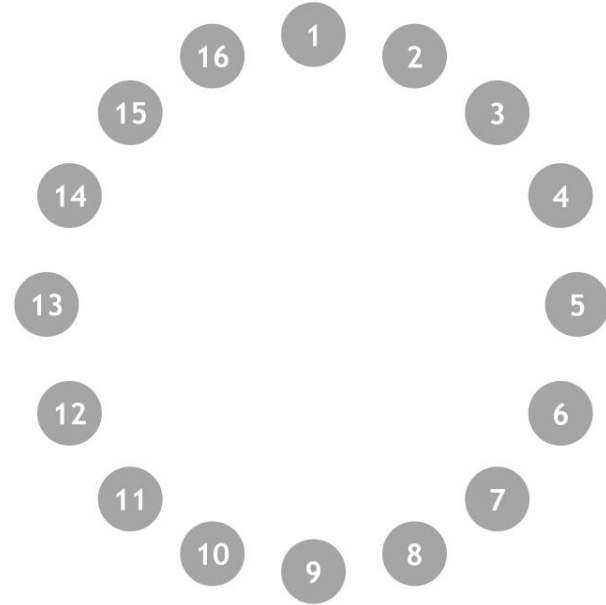
Random Networks



Pál Erdős



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

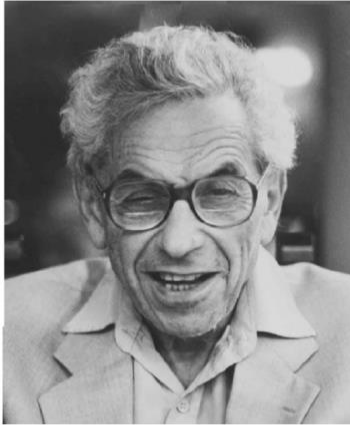
Conectar 1 con 2?

is Random() < p

Si

No

a.

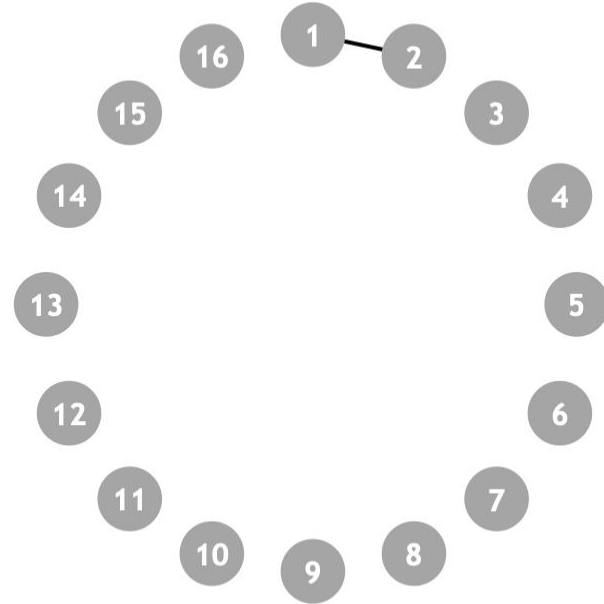


Pál Erdős

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

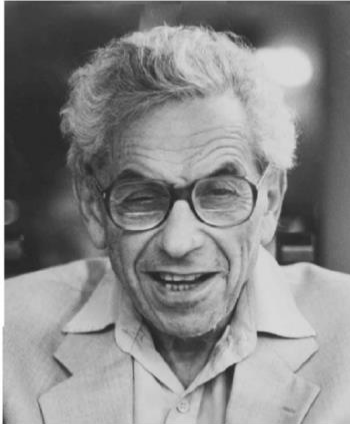
Random Networks

Conectar 1 con 3?

is Random() < p

Si
No

a.

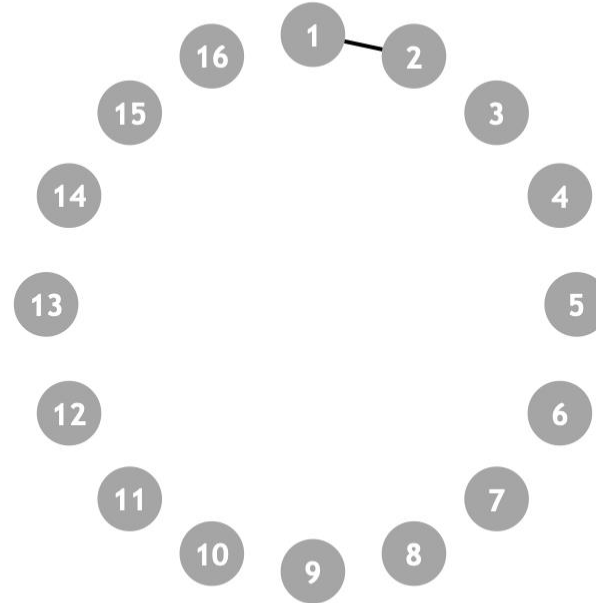


Pál Erdős

b.



Alfréd Rényi

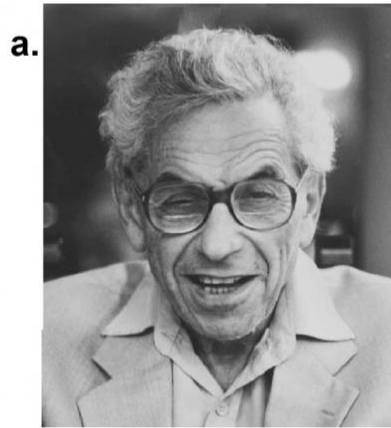


Erdős-Rényi algorithm

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

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

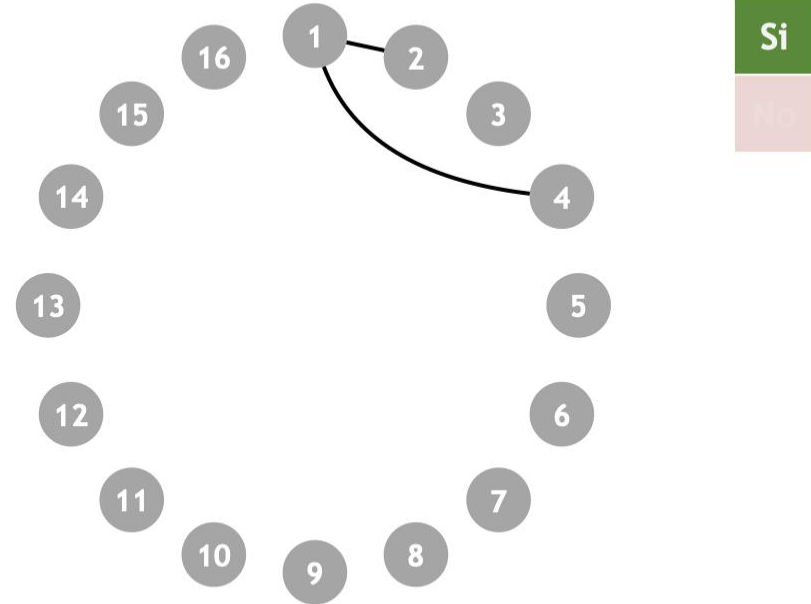
Random Networks



Pál Erdős



Alfréd Rényi



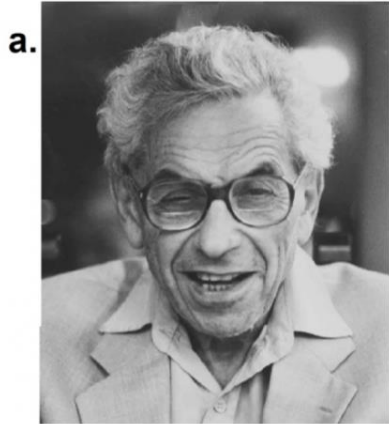
Erdős-Rényi algorithm

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

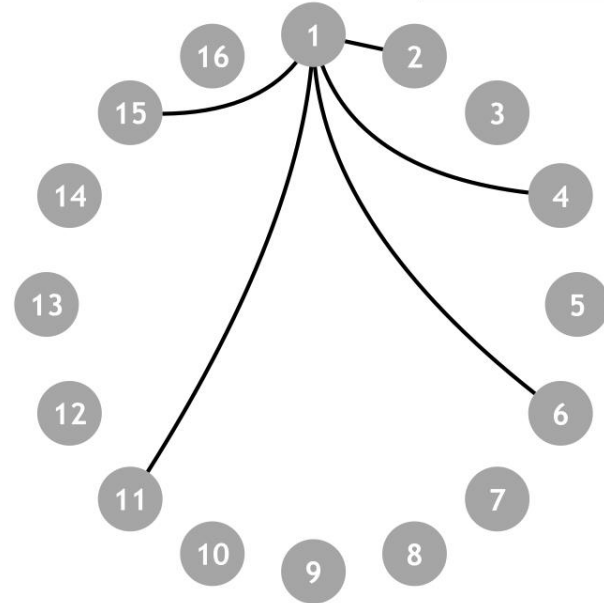
Repita para todas las
Posibles conexiones
de 1



Pál Erdős



Alfréd Rényi



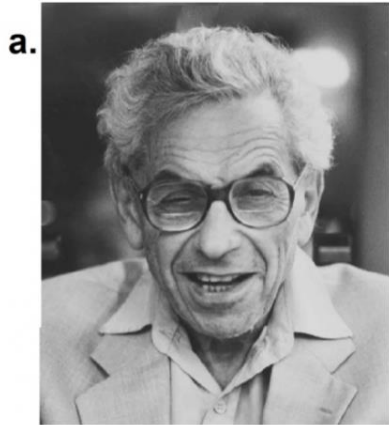
Erdős-Rényi algorithm

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

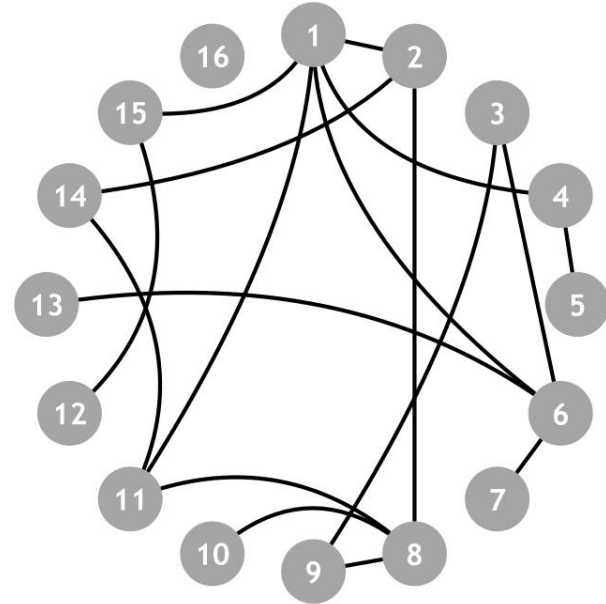
Repite para todos
Los posibles enlaces



Pál Erdős



Alfréd Rényi



Erdős-Rényi algorithm

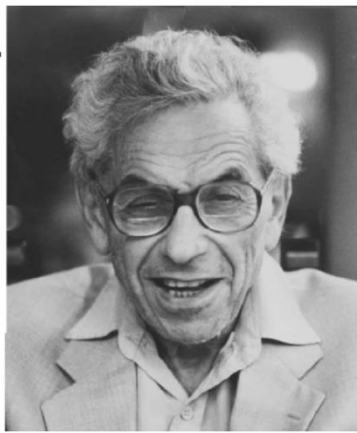
condición inicial Comenzando con N nodos desconectados ($L = 0$ enlaces)

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

El número de enlaces es variable.

MODELO DE REDES ALEATORIAS

Erdős-Rényi algorithm



Pál Erdős



Alfréd Rényi

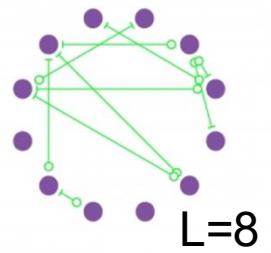
condición inicial

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

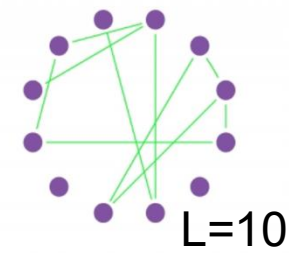
paso iterativo

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

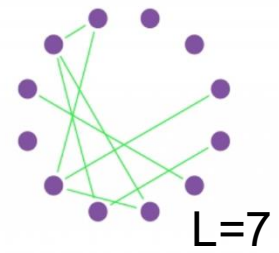
$N=12$
 $p = 1/6$



$L=8$

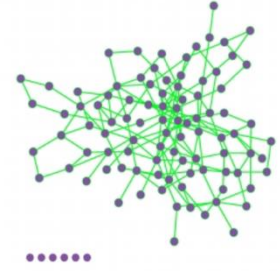
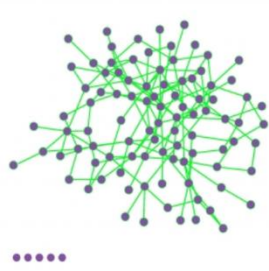
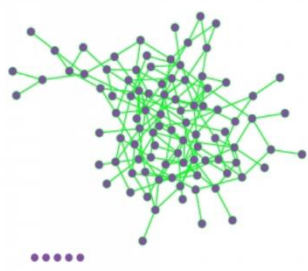


$L=10$



$L=7$

$N=100$
 $p = 0.03$



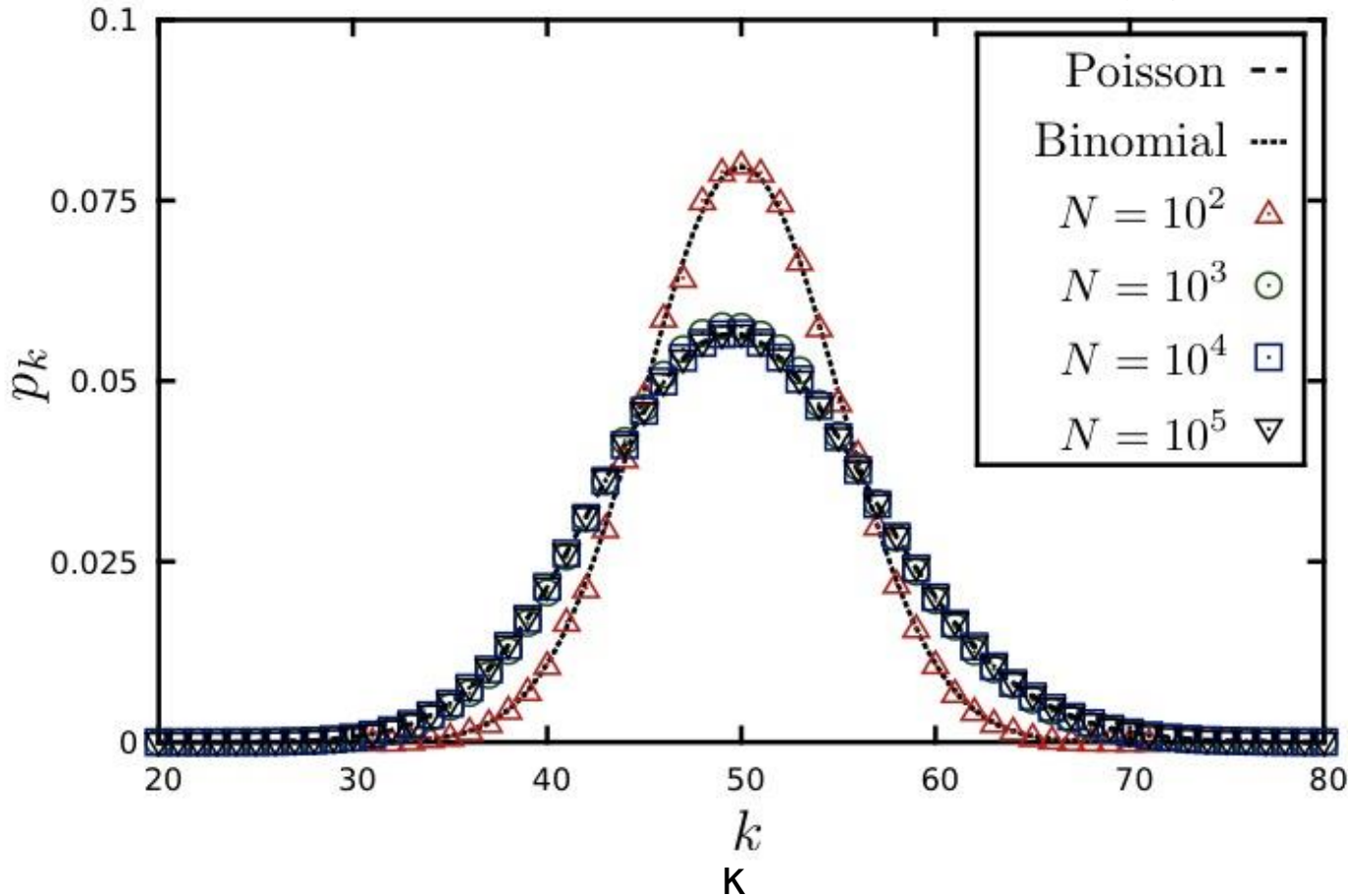
Parece una red real, ¿no es así?

Distribución de grado de una red aleatoria

DISTRIBUCION DE GRADO DE UNA RED ALEATORIA

$\langle k \rangle = 50$

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



$P(k): k \ll N$
Indistinguibles

DISTRIBUCION DE GRADO DE UNA RED ALEATORIA

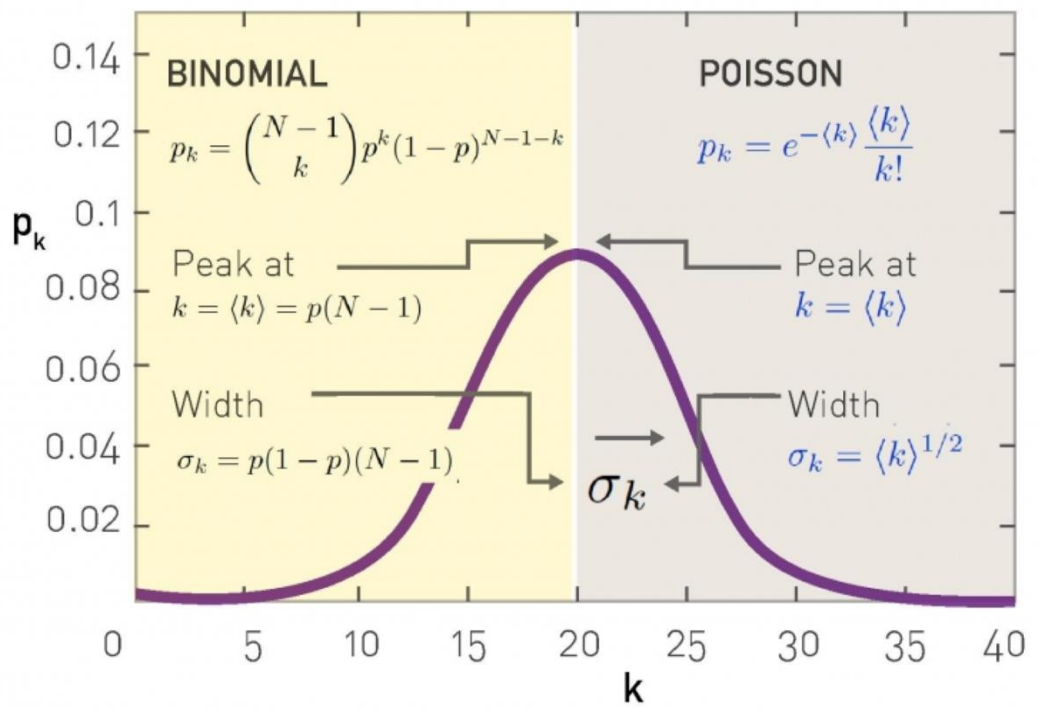
Resultado exacto

- Distribución Binomial -

Límite para N grande

- Distribución de Poisson -

Función de distribución de probabilidad (PDF)



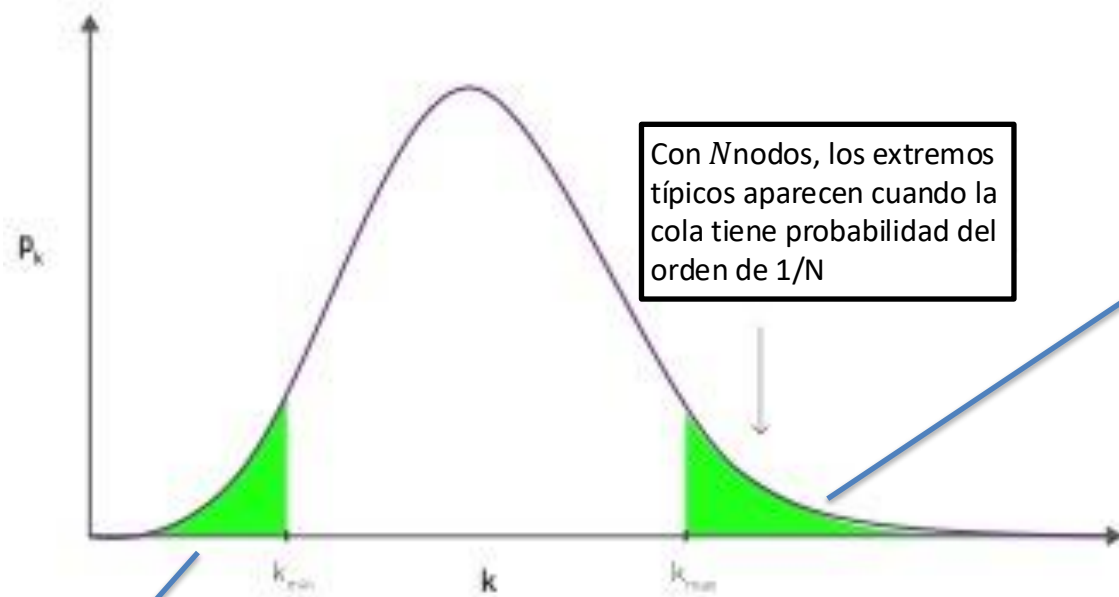
Las redes reales **no** son Poisson

Grado máximo y mínimo

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

$$N = 8 \times 10^9 \quad \frac{1}{N} = \frac{1}{8 \times 10^9} = 1.25 \times 10^{-10}$$

$$\sigma_k = \sqrt{\langle k \rangle} = \sqrt{1000} \approx 31.62$$



Con N nodos, los extremos típicos aparecen cuando la cola tiene probabilidad del orden de $1/N$

$$\Pr(K > k_{\max}) \approx \frac{1}{N}$$

$$k_{\max} \approx 1000 + 6.3(31.6) \approx 1000 + 199 \approx 1199$$

$$\Pr(K \leq k_{\min}) \approx \frac{1}{N}$$

$$k_{\min} \approx 1000 - 6.3(31.6) \approx 1000 - 199 \approx 801$$

Si $K \sim \text{Poisson}(\langle k \rangle)$ y N es grande:

$$\Pr(Z \leq z) = 1 - \frac{1}{N}$$

$$z = \Phi^{-1}\left(1 - \frac{1}{N}\right)$$

$$z \approx 6.3$$

$$k_{\max} \approx \langle k \rangle + z_N \sqrt{\langle k \rangle}$$

$$k_{\min} \approx \langle k \rangle - z_N \sqrt{\langle k \rangle}$$

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 $k_{\max} \sim 1,199$ $1 - P(k_{\max}) = 1 - e^{-\langle k \rangle} \sum_{k=0}^{k_{\max}} \frac{\langle k \rangle^k}{k!} = e^{-\langle k \rangle} \sum_{k=k_{\max}+1}^{\infty} \frac{\langle k \rangle^k}{k!} \approx e^{-\langle k \rangle} \frac{\langle k \rangle^{k_{\max}+1}}{(k_{\max}+1)!}$

→ El individuo menos conectado tiene grado $k_{\min} \sim 801$ $P(k_{\min}) = e^{-\langle k \rangle} \frac{\langle k \rangle^{k_{\min}}}{k_{\min}!}$

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.

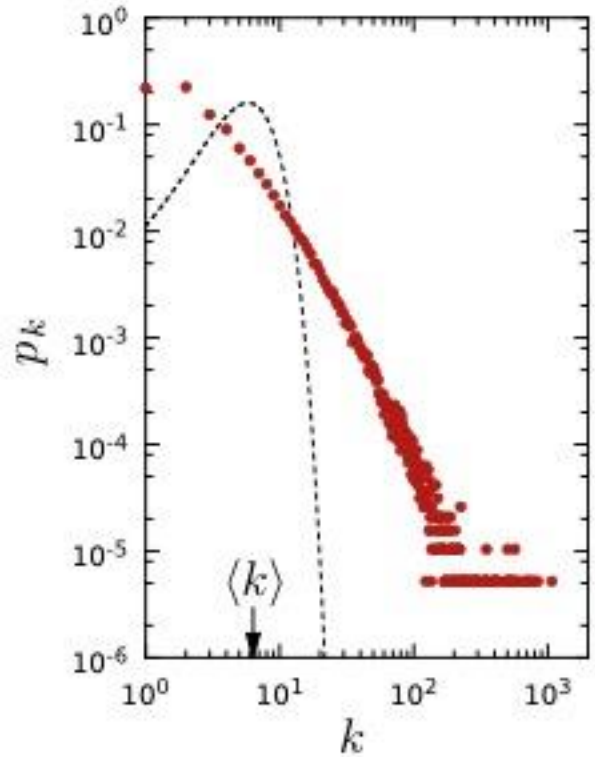
→ una sociedad aleatoria consistiría principalmente en individuos promedio, con todos con aproximadamente el mismo número de amigos.

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

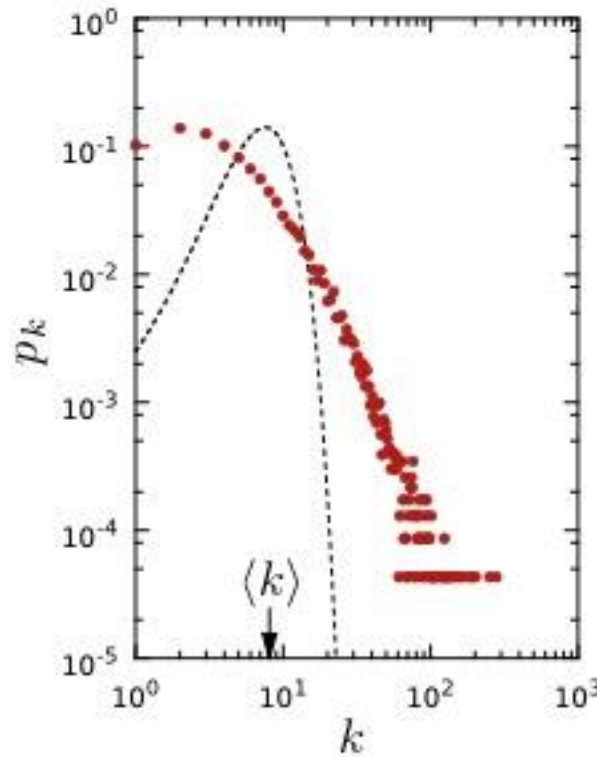
ENFRENTANDO LA REALIDAD: Distribución en grado de redes reales.

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

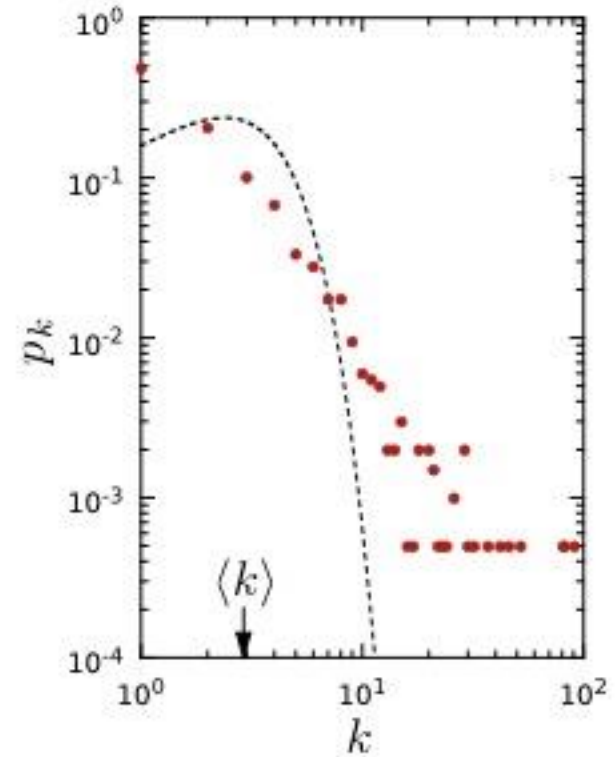
Internet



Science Collaboration

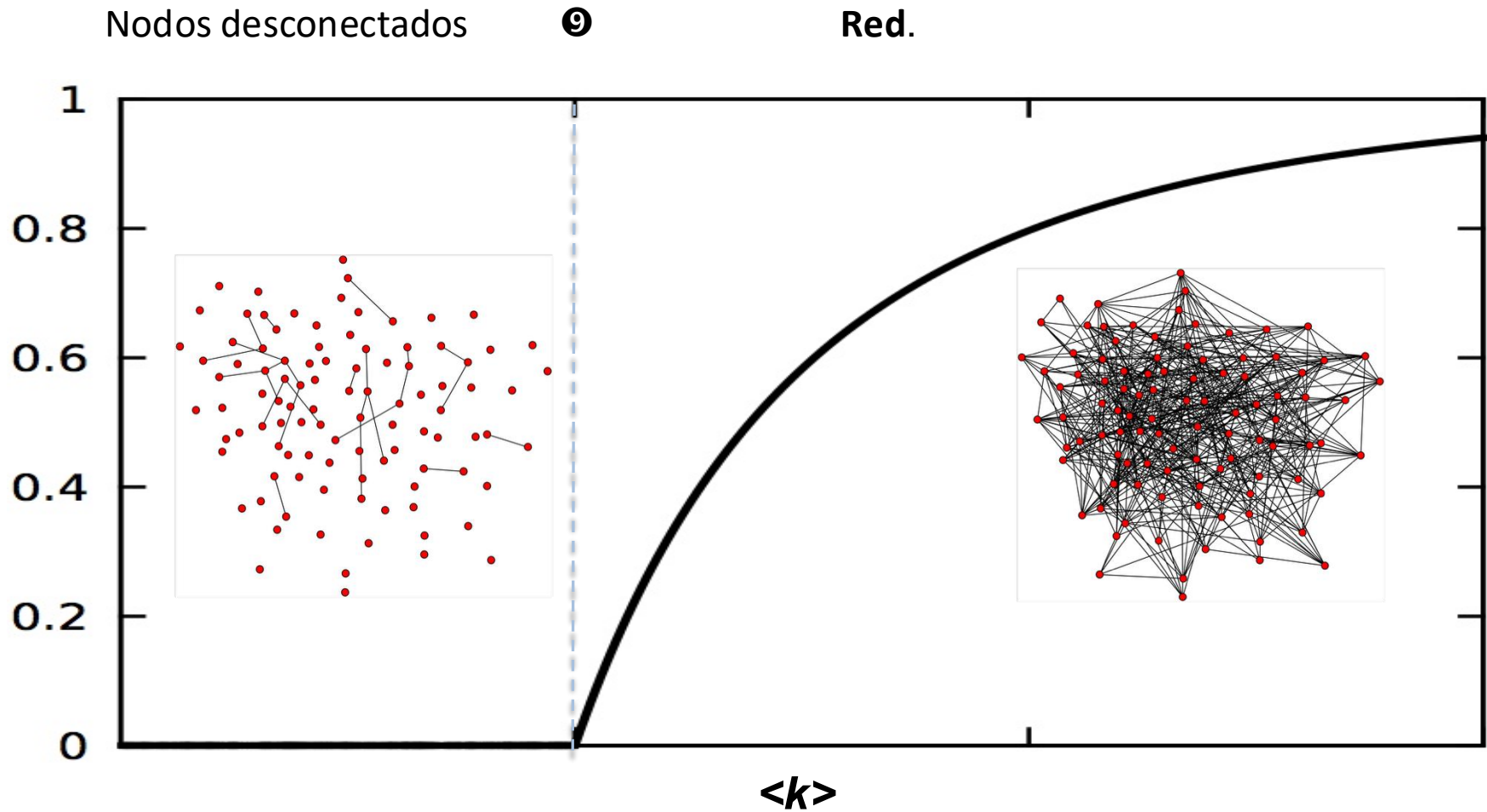


Protein Interactions



La evolución de una red aleatoria.

Evolución de una red aleatoria



Nodos desconectados

⑨

Red.

¿Cómo se produce esta transición?

DISTRIBUCION DEL TAMAÑO DEL GRUPO (CLUSTER)

Probabilidad de que un nodo seleccionado al azar pertenezca a un grupo de tamaño s :

$$p_s \sim s^{-3/2} \exp[-(\langle k \rangle - 1)s + (s - 1) \ln \langle k \rangle]$$

Esto implica:

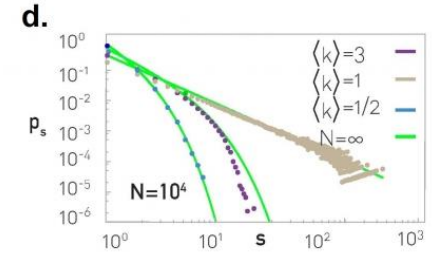
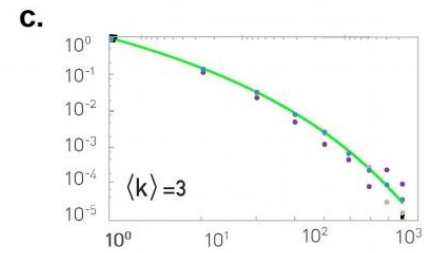
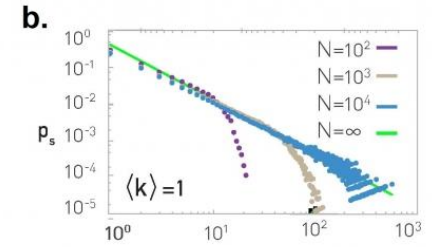
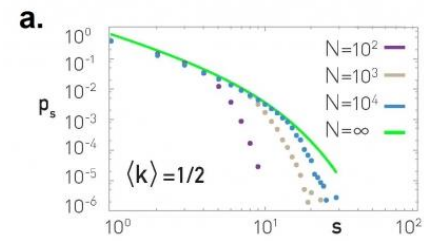
- Si $\langle k \rangle < 1$: p_s decae exponencialmente, por lo que predominan componentes pequeños.

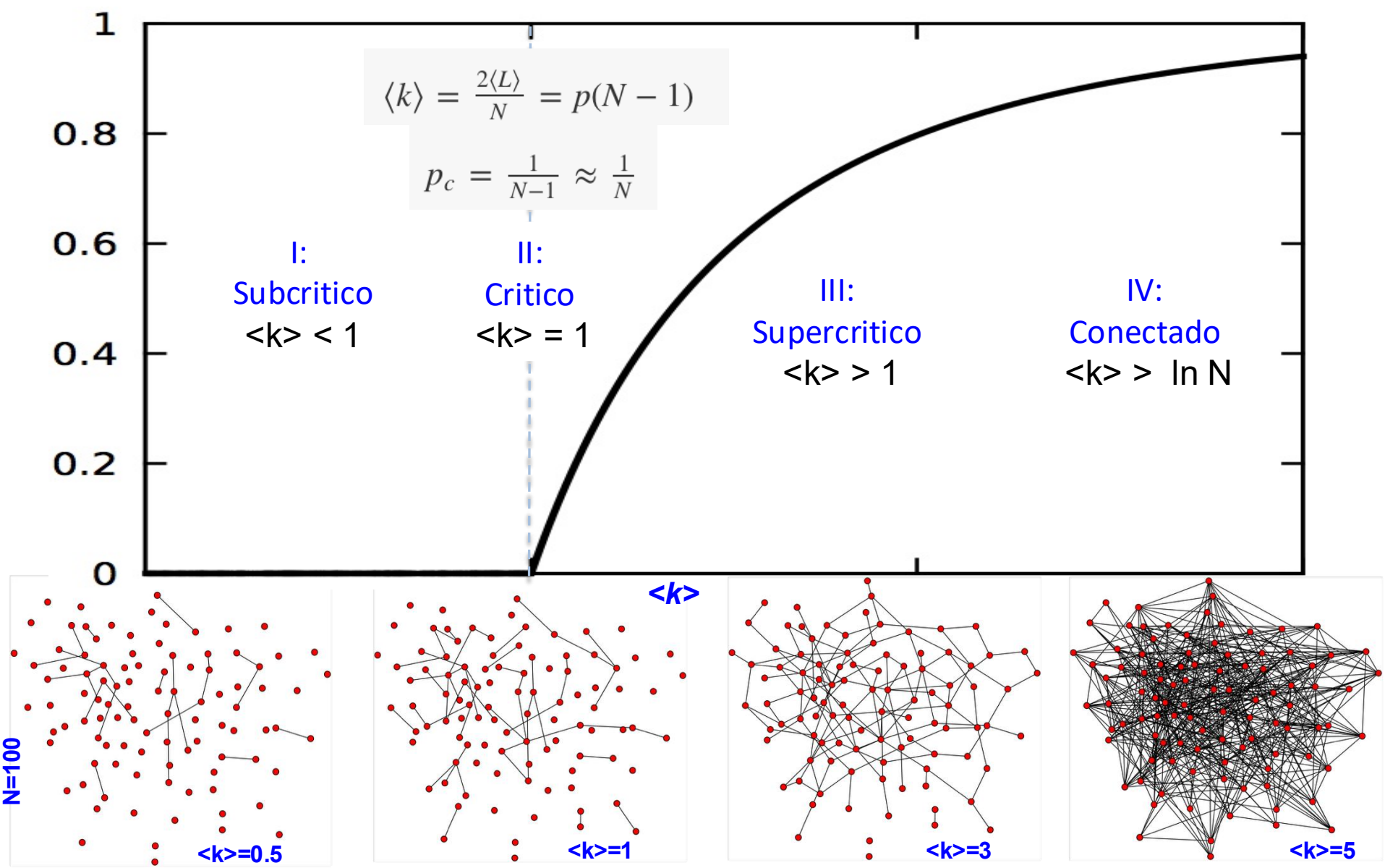
- Si $\langle k \rangle = 1$:

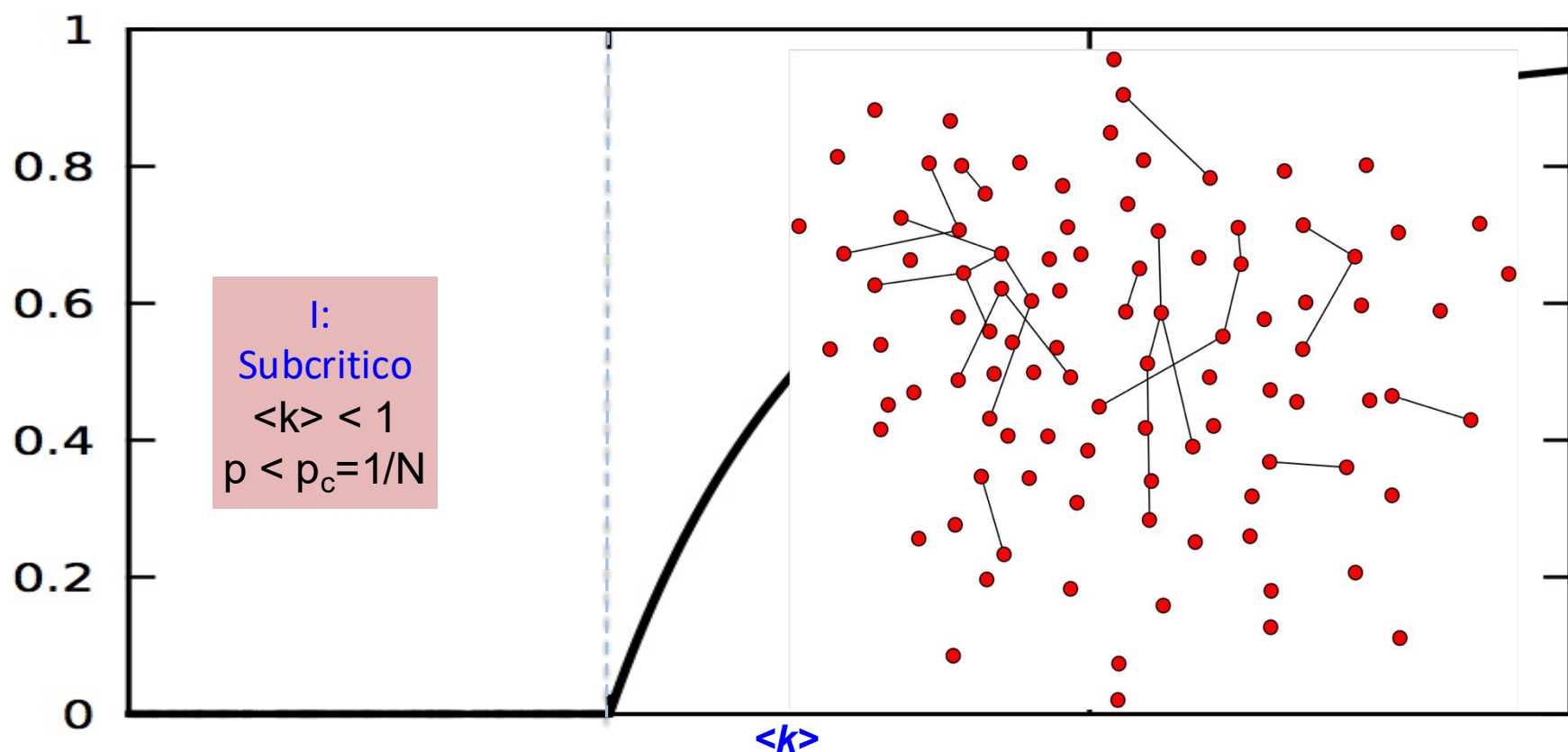
$$p_s \sim s^{-3/2}$$

La distribución no tiene escala típica: aparecen clústeres de muchos tamaños.

- Si $\langle k \rangle > 1$: sigue habiendo componentes finitos pequeños, pero además emerge un componente gigante.





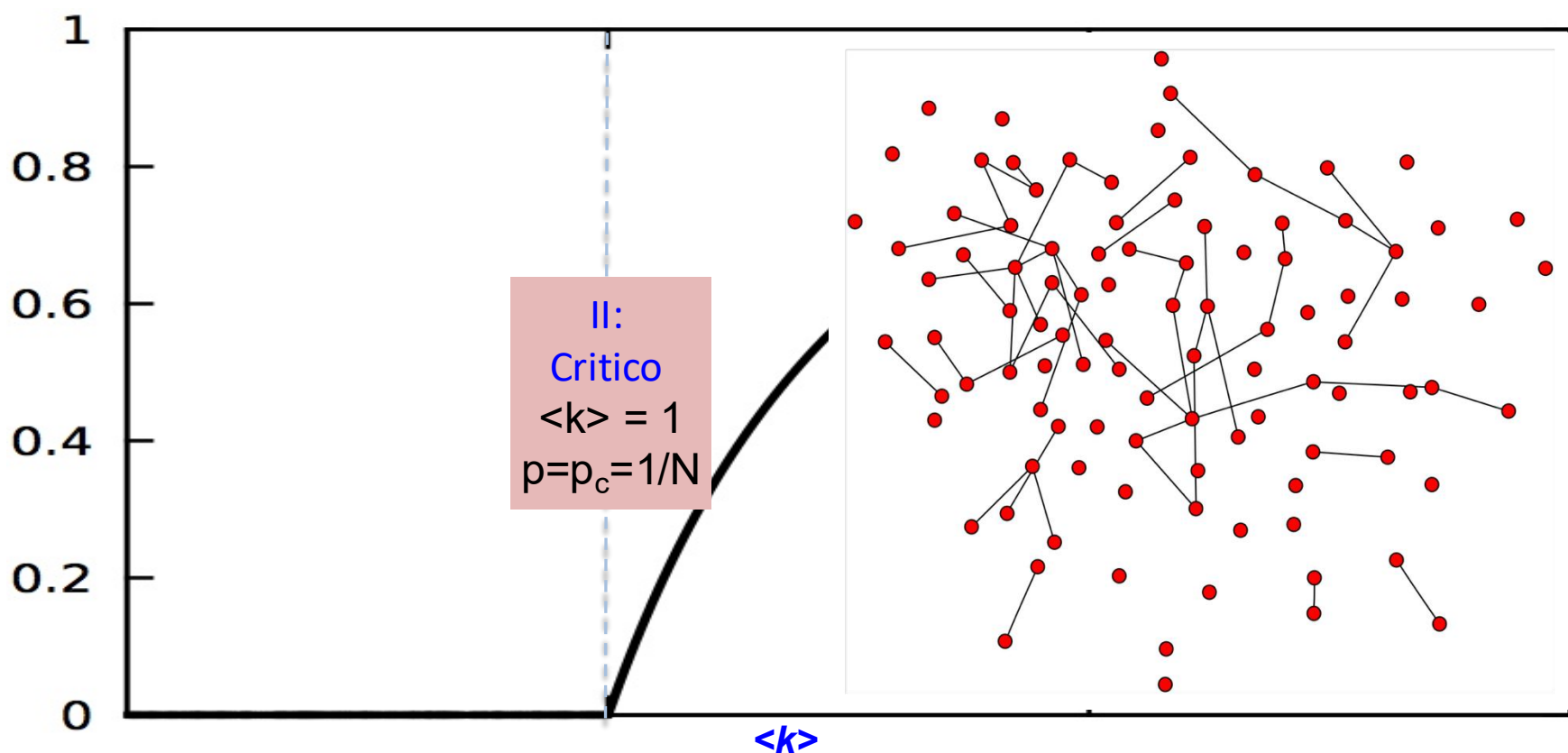


No hay componente gigante

$$p(s) \sim s^{-3/2} e^{-((k)-1)s + (s-1)\ln \langle k \rangle}$$

N-L clusters aislados, la distribución del tamaño del cluster es exponencial

El cluster mas grande es un árbol, su tamaño es $\sim \ln N$



Una sola componente gigante: $N_G \sim N^{2/3}$

Contiene una decreciente fracción de todos los nodos, $N_G/N \sim N^{-1/3}$

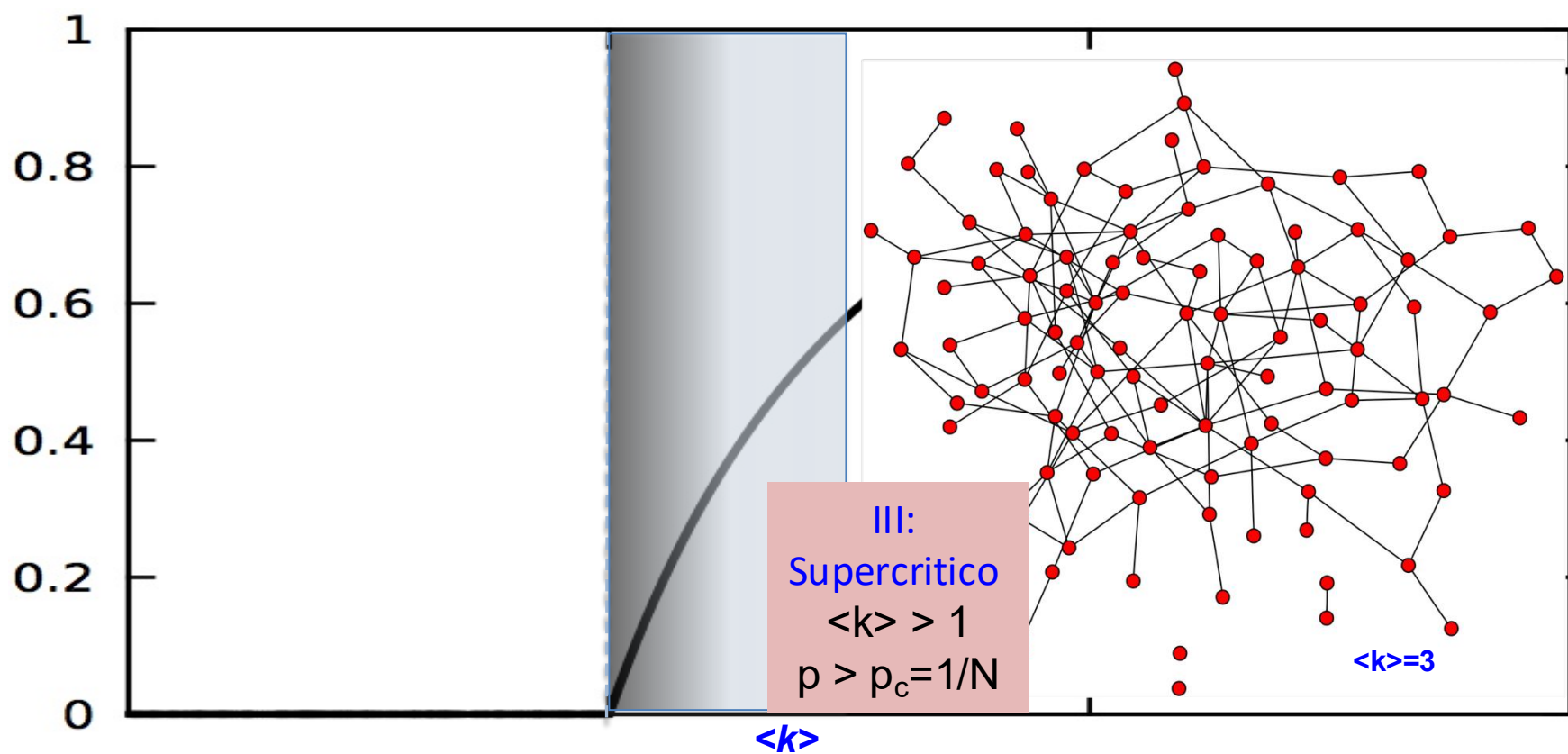
Componentes pequeños son árboles, GC tiene loops

Distribución del tamaño de cluster: $p(s) \sim s^{-3/2}$
(Power-law/ley de potencia)

Un salto en el tamaño del cluster:

$N=1,000 \rightarrow \ln N \sim 6.9; N^{2/3} \sim 95$

$N=7 \cdot 10^9 \rightarrow \ln N \sim 22; N^{2/3} \sim 3,659,250$

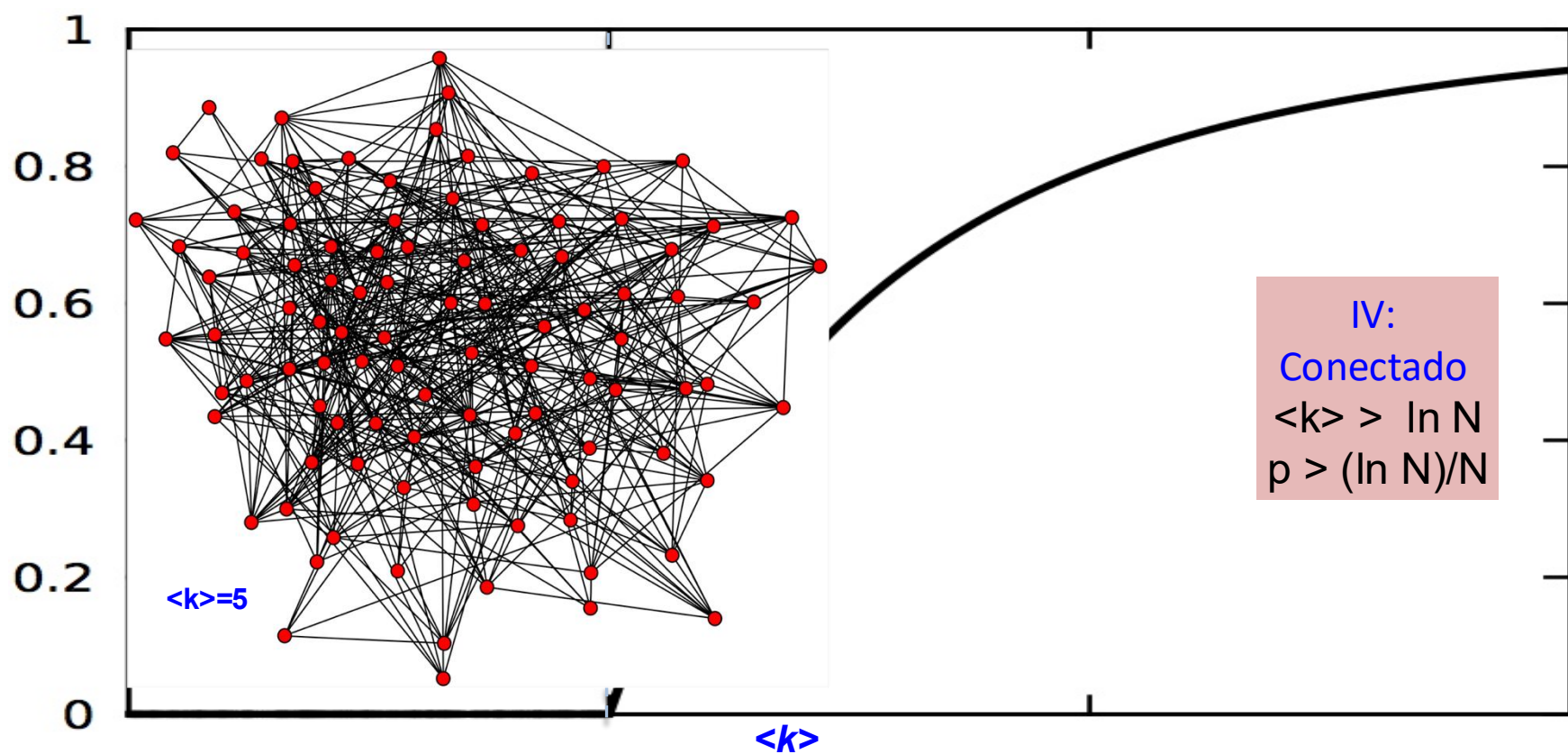


Una sola componente gigante: $N_G \sim (p - p_c)N$

GC tiene loops.

Distribución de tamaños de clusters: exponencial

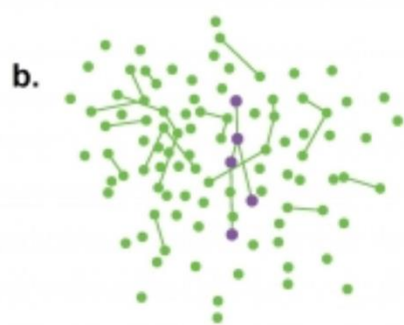
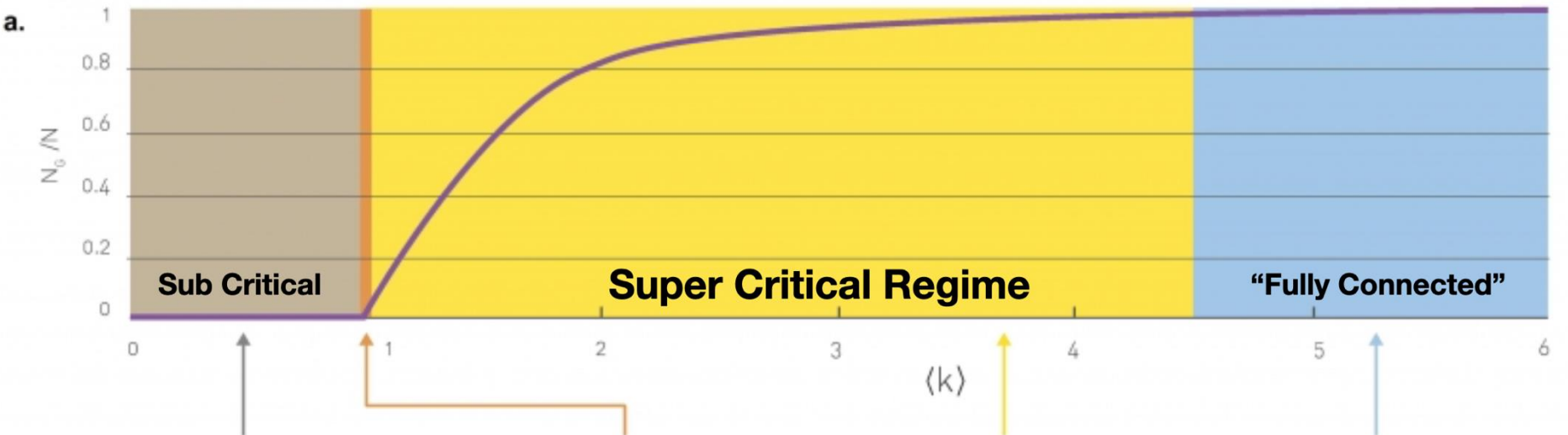
$$p(s) \sim s^{-3/2} e^{-((k)-1)s + (s-1)\ln\langle k \rangle}$$



Solo un cluster: $N_G = N$

GC es densa.

Distribución de tamaño de clusters: Ninguna



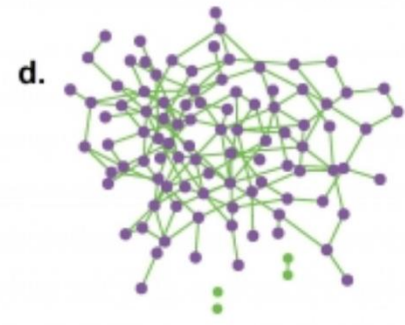
$\langle k \rangle < 1$

- (b) Subcritical Regime**
- No giant component
 - Cluster size distribution: $p_s \sim s^{-2} e^{-s}$
 - Size of the largest cluster $N_g \sim \ln N$
 - The clusters are trees



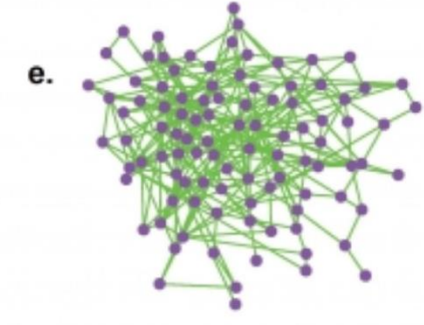
$\langle k \rangle = 1$

- (c) Critical Point**
- No giant component
 - Cluster size distribution: $p_s \sim s^{-3}$
 - Size of the largest cluster: $N_g \sim N^{2/3}$
 - The clusters may contain loops



$\langle k \rangle > 1$

- (d) Supercritical Regime**
- Single giant component
 - Cluster size distribution: $p_s \sim s^{-2} e^{-s}$
 - Size of the giant component: $N_g \sim (p - p_c)N$
 - The small clusters are trees
 - Giant component has loops



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

- (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.

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

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

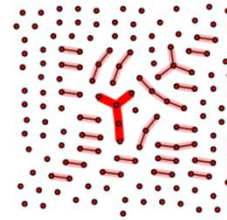
$\langle k_c \rangle = 1/N * (N-1) = 1$, para N grande
(Erdos y Renyi, 1959)

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.

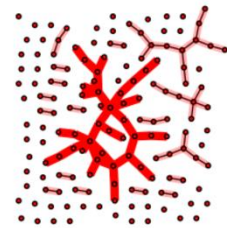
Es algo inesperado, sin embargo, que un enlace es **suficiente** para el surgimiento de un componente gigante.

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 $\langle k \rangle = 1$.

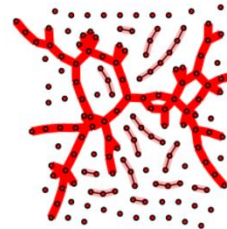
$p = 0.003$



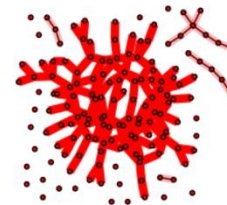
$p = 0.006$



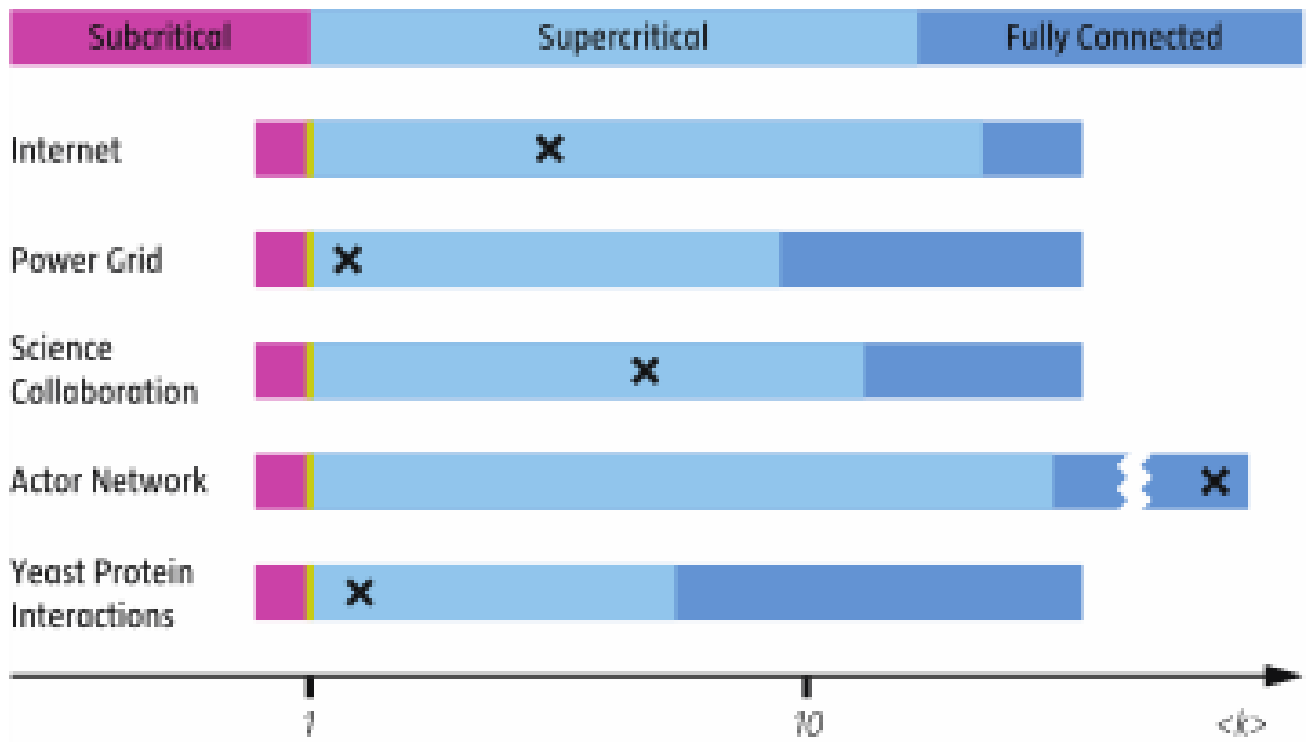
$p = 0.008$



$p = 0.015$



Las redes reales son supercríticas.



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
 $\langle k \rangle > \ln N$

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 compara 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:

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

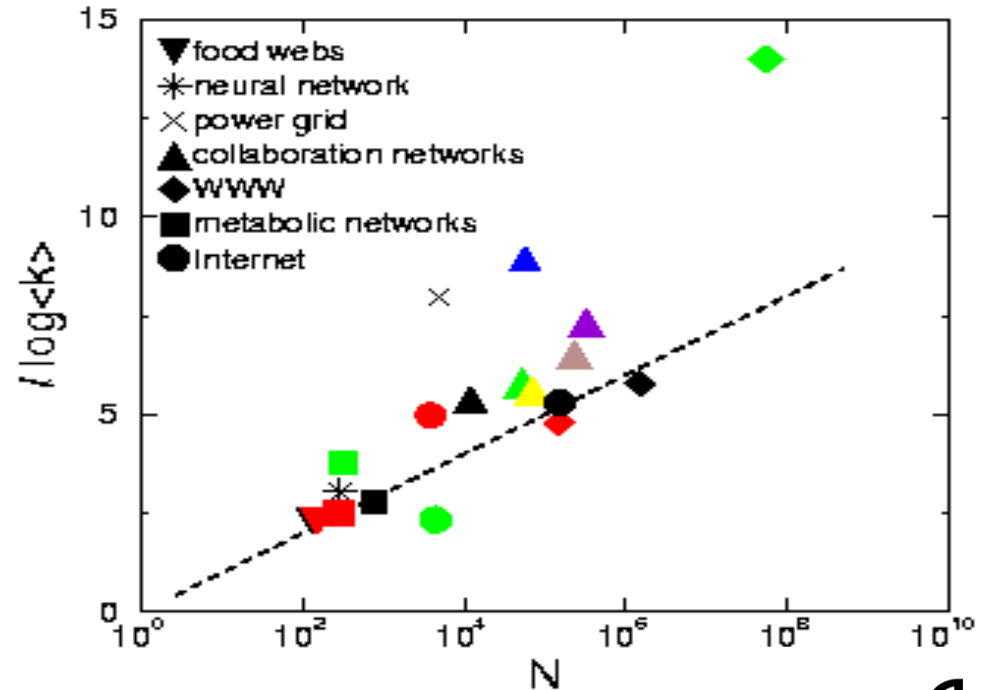
Distribución de Grados:

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

LONGITUD DE CAMINOS EN REDES REALES

Predicción:

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



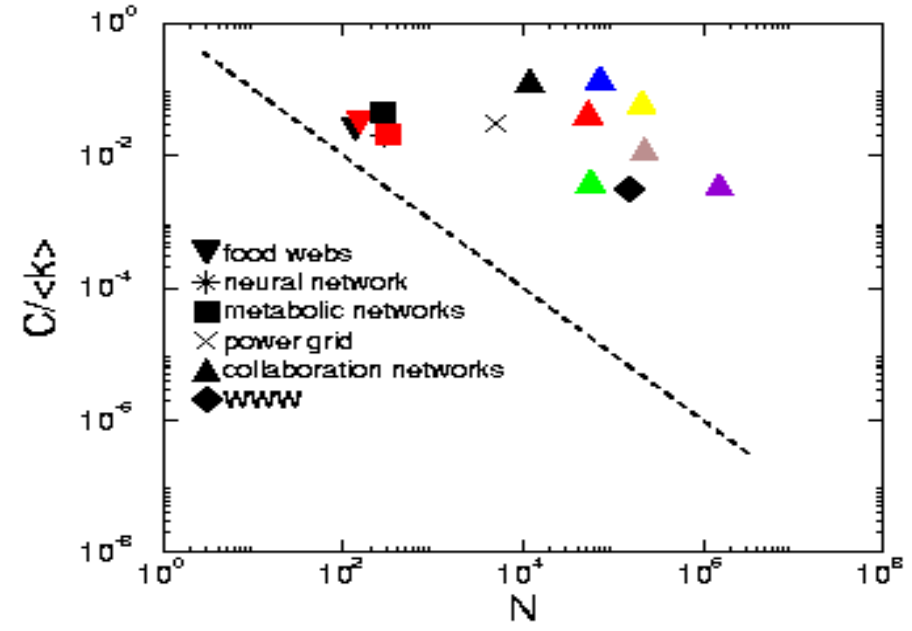
Las redes reales tienen distancias cortas como los grafos aleatorios.



COEFICIENTE DE CLUSTERING

Predicción:

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



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



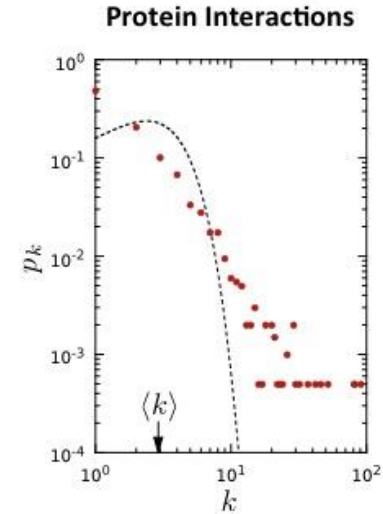
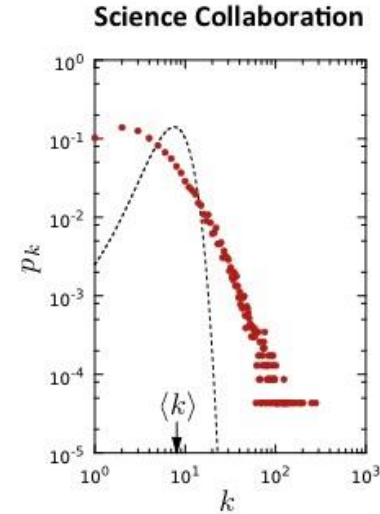
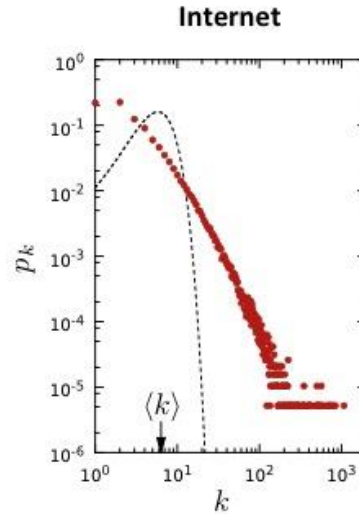
LA DISTRIBUCIÓN DE GRADO

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 compara 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:

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

Longitud media del camino:



Coeficiente de agrupamiento:

$$C_1 = \frac{2\langle L_1 \rangle}{k_1(k_1 - 1)} = p = \frac{\langle k \rangle}{N}$$



Distribución de Grados:

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



Preguntas básicas

- Qué representa la distribución de grado en una red y por qué es un concepto importante en análisis de redes?
- De qué manera se relaciona el concepto de centralidad de grado con distribución de grado de una red?
- Explica el concepto de una red aleatoria y cómo difiere de una red real.

Erdős–Rényi (ER): qué es

En $G(n, p)$, cada par de nodos se conecta independientemente con la misma probabilidad p .

Qué controla

- número de nodos
- densidad esperada o número total de enlaces
- conectividad aleatoria homogénea

Qué no controla

- heterogeneidad de grado
- clustering elevado
- comunidades
- crecimiento histórico
- atributos o restricciones espaciales/sociales.

ER como modelo nulo

Tipo: modelo nulo clásico.

Para qué sirve

ER responde la pregunta:

¿esta propiedad podría surgir simplemente por conexiones aleatorias homogéneas?

Qué suele reproducir razonablemente

- orden de magnitud de las distancias medias
- emergencia de componente gigante cuando el grado medio supera cierto umbral.

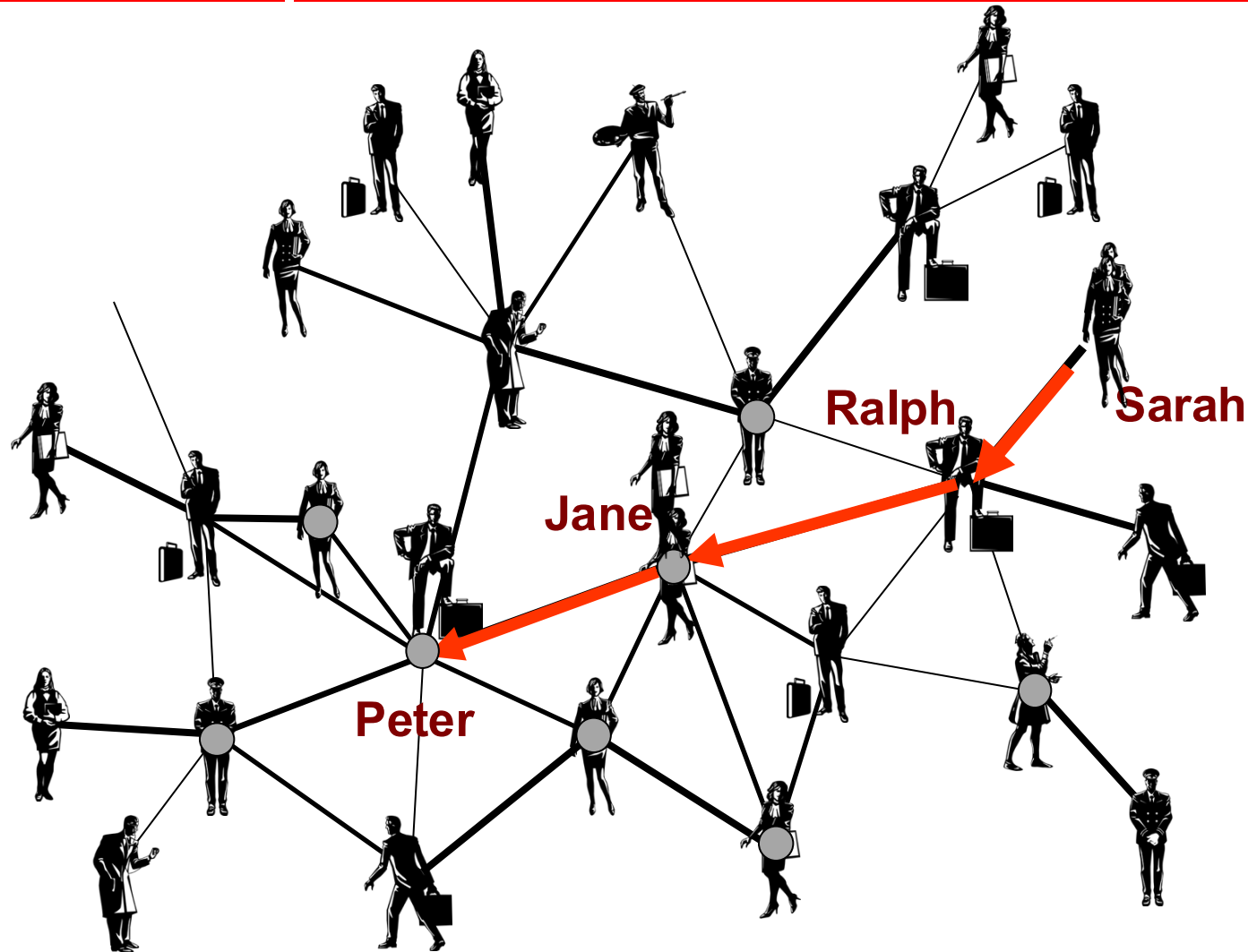
Entonces:

Si tu red real solo “supera” a ER en clustering, eso no implica automáticamente una explicación.

Solo implica que el azar homogéneo no basta.

Mundo Pequeño

(Small World, modelo Watts-Strogatz)

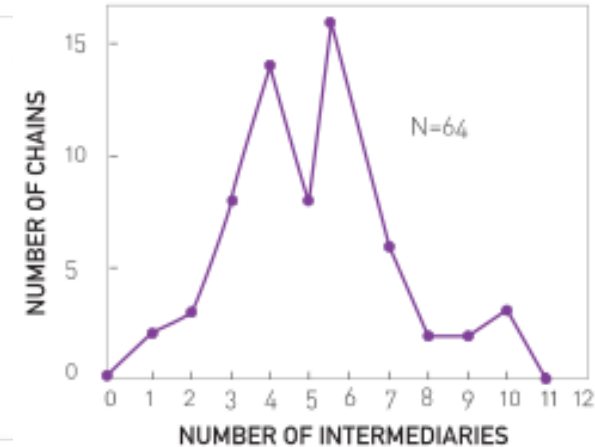
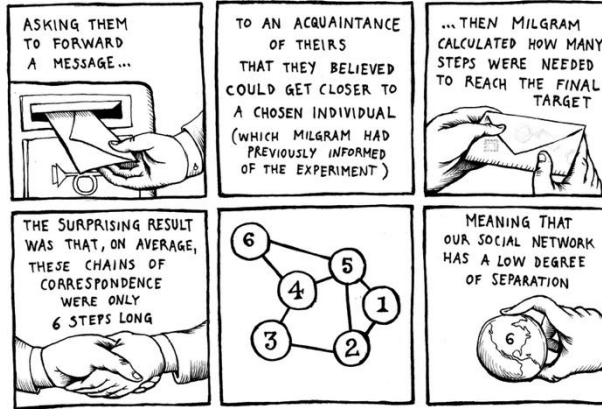
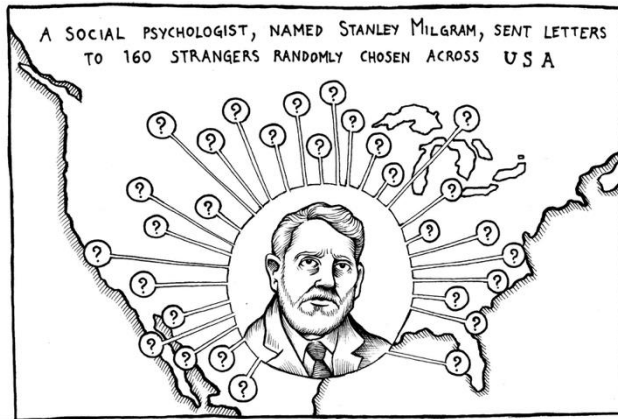


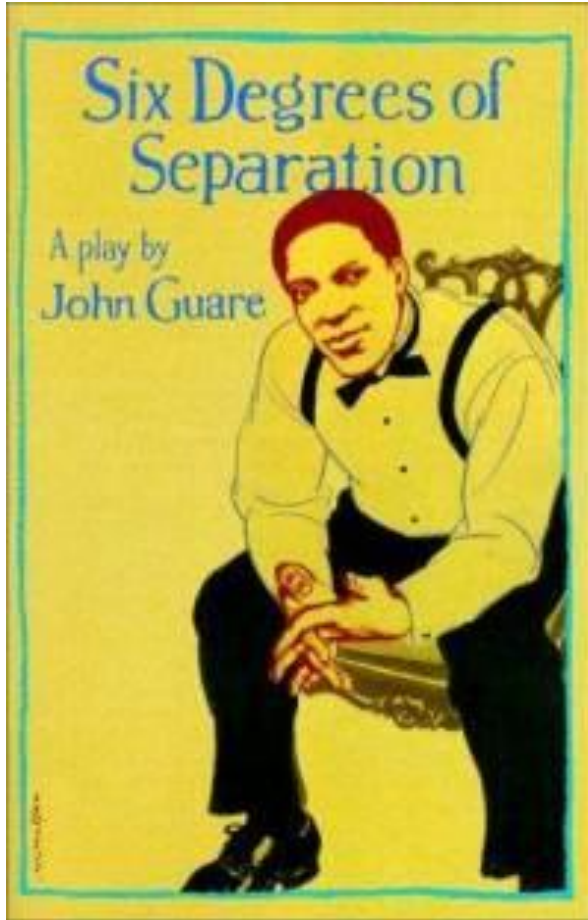
*Frigyes Karinthy, 1929
Stanley Milgram, 1967*

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?

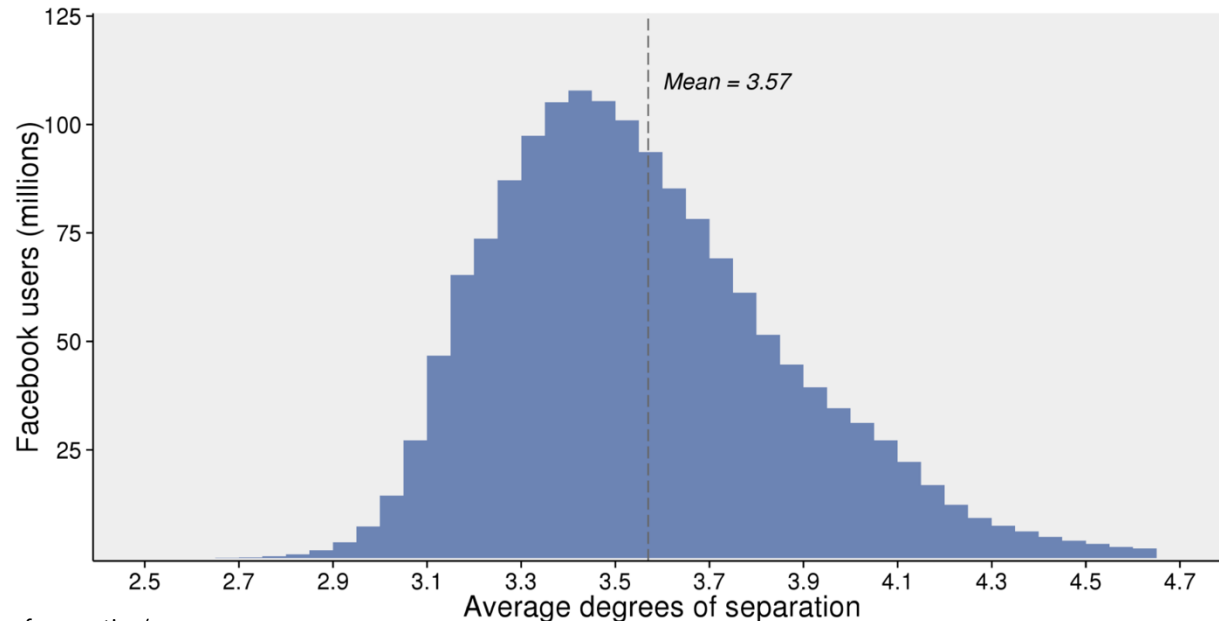
1967



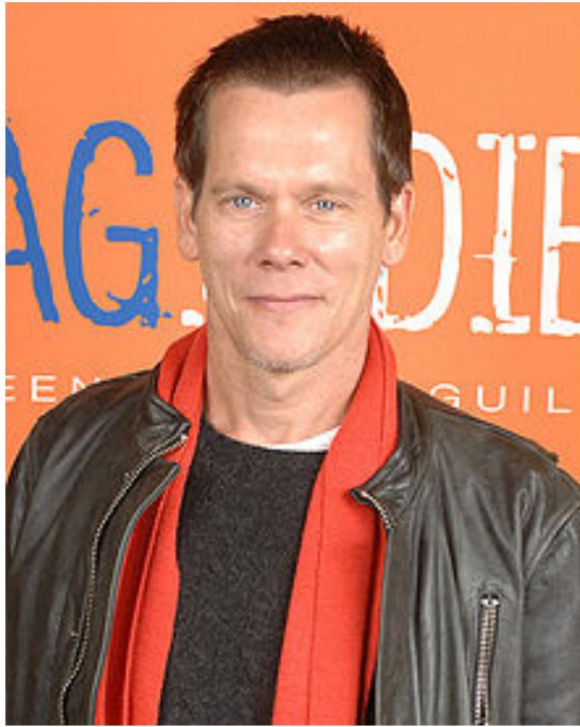


"Todos en este planeta están separados solo por otras seis personas. Seis grados de separación. Entre nosotros y todos los demás en este planeta. El presidente de los Estados Unidos. Un gondolero en Venecia ... No son solo los grandes nombres. Es cualquiera. Un nativo en una selva tropical. Un Tierra del Fuego. 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

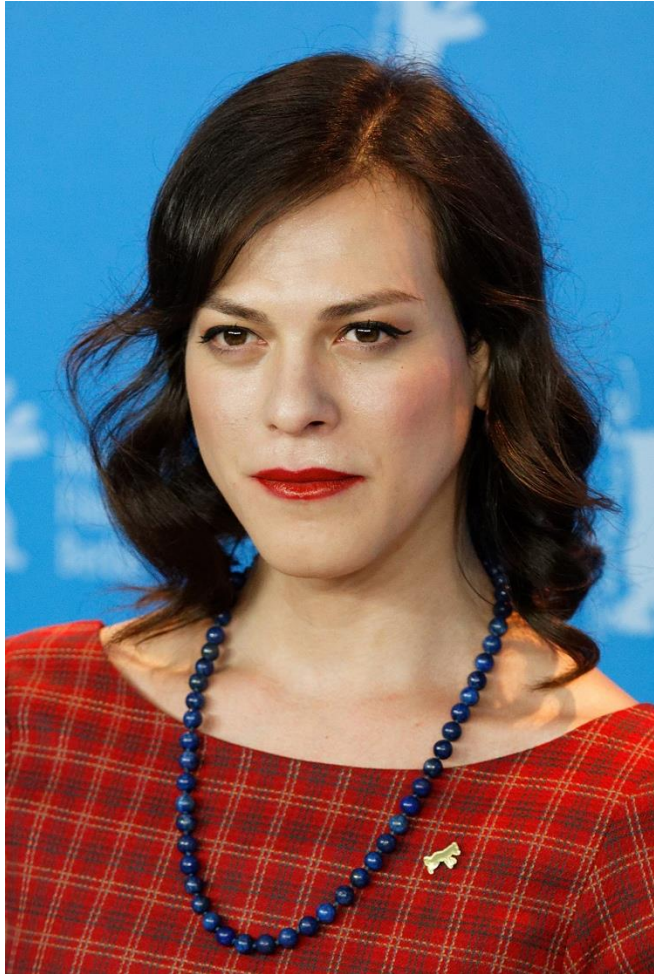
to

Lorenza Izzo

Find link

More o

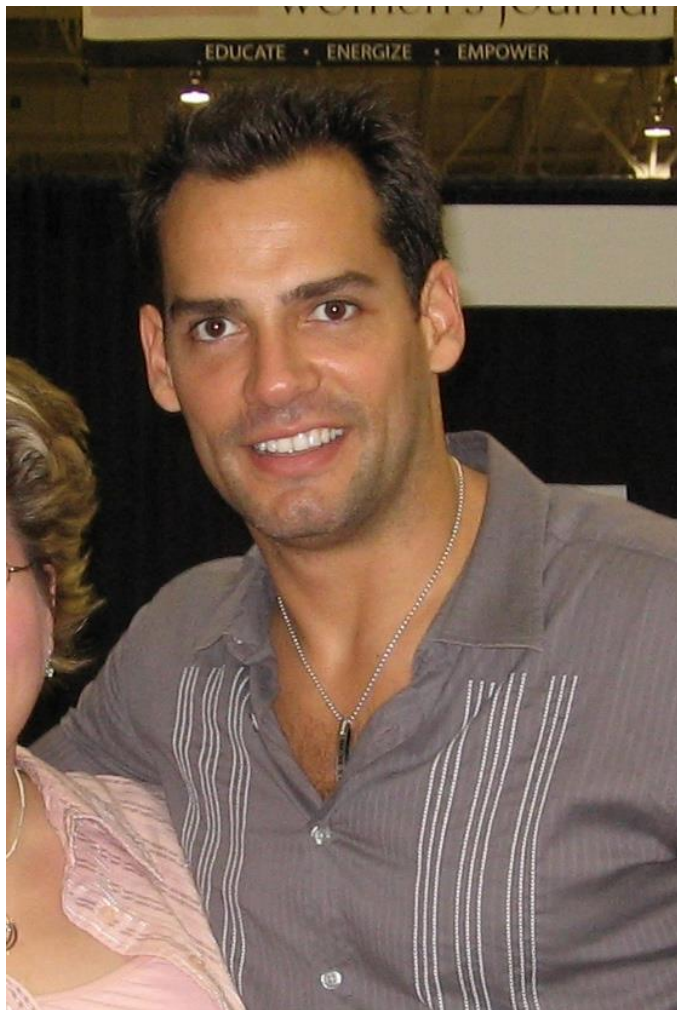
<https://oracleofbacon.org>



daniela vega has a Bacon number of 3.



Kevin Bacon to daniela vega Find link



cristian de la fuente has a Bacon number of 2.

Fin

Cristian de la Fuente

was in

Driven

with

Burt Reynolds

was in

Starting Over

with

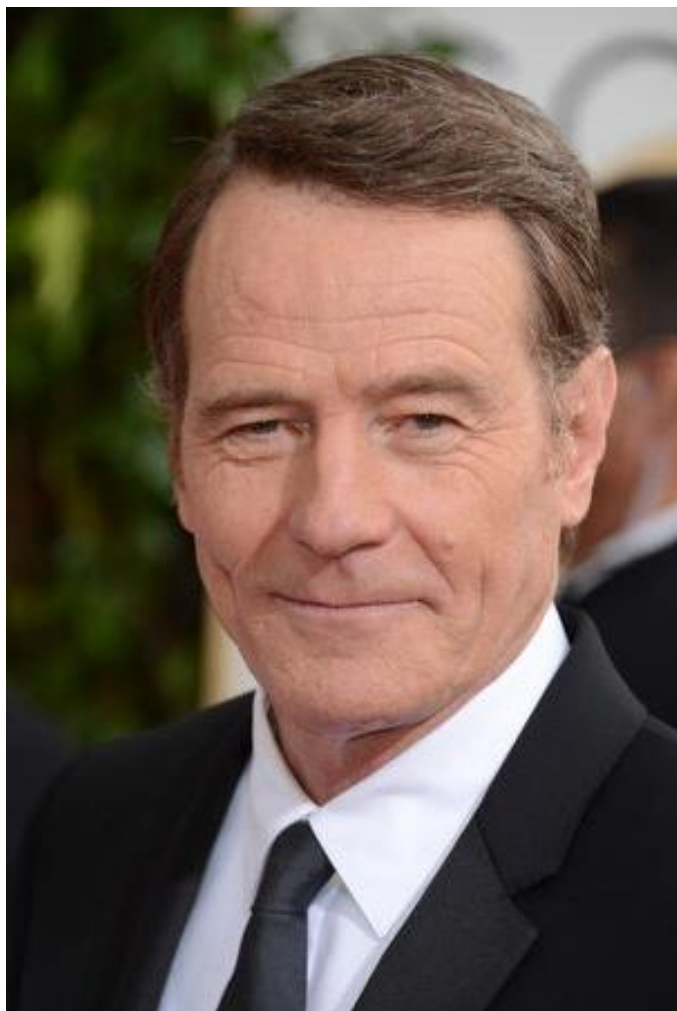
Kevin Bacon

Kevin Bacon

to cristian de la fuente

Find link

More opti



Bryan Cranston has a Bacon number of 2.

Bryan Cranston

was in

Little Miss Sunshine

with

Steve Carell

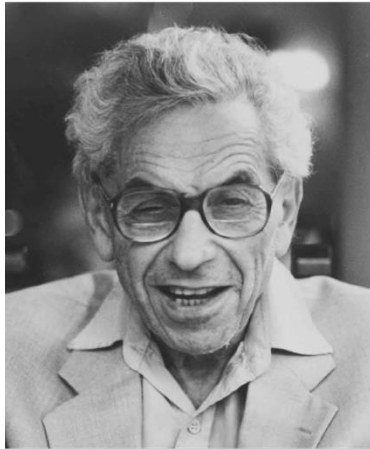
was in

Crazy, Stupid, Love

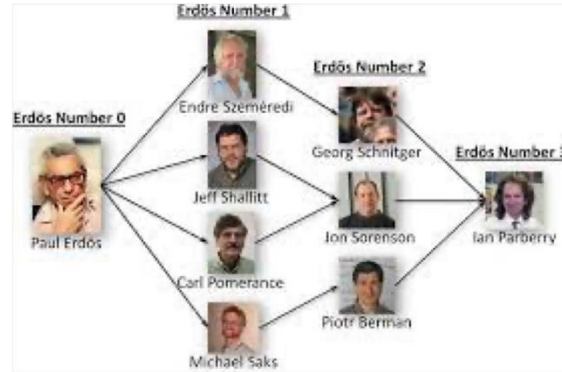
with

Kevin Bacon

Erdos Number

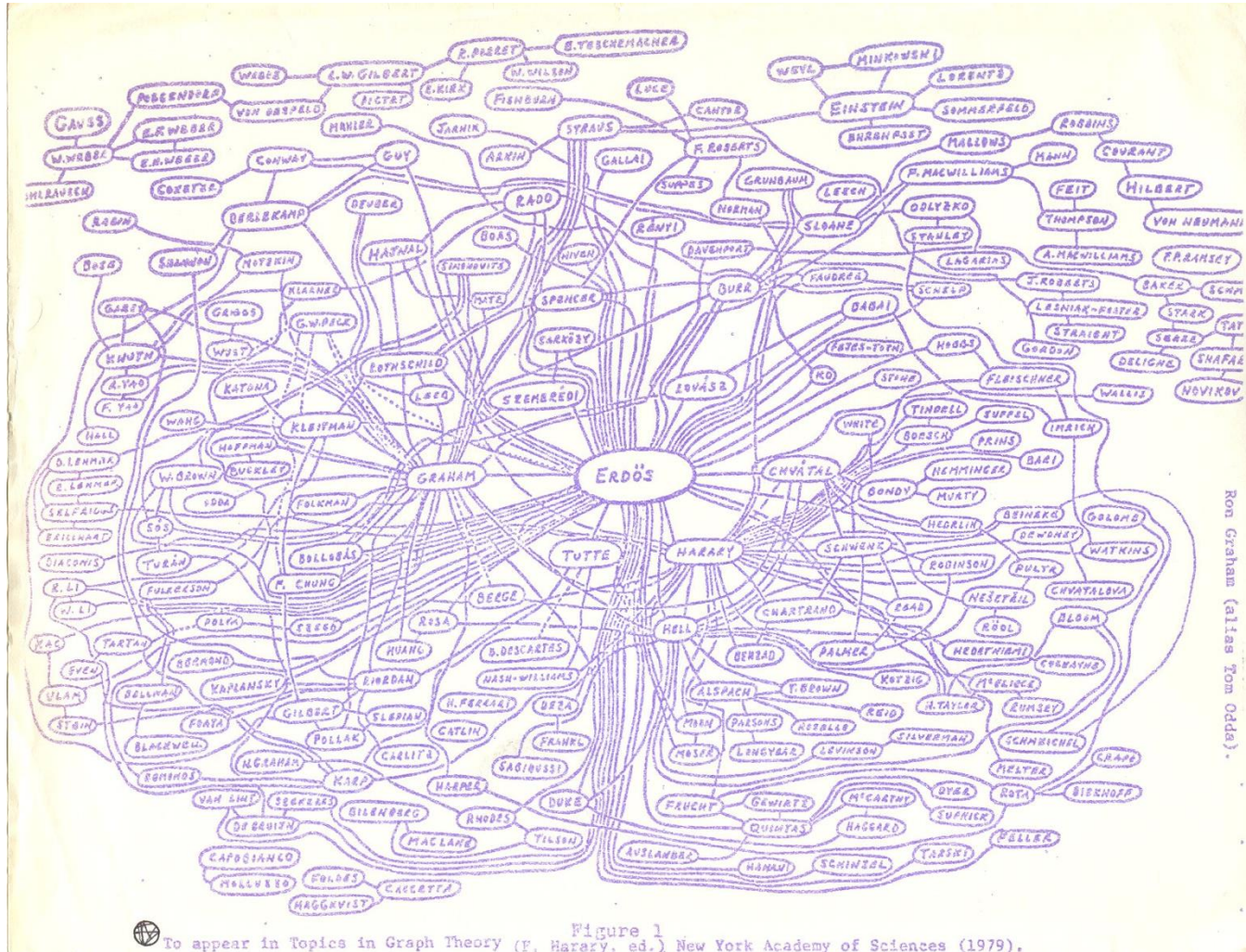


Paul Erdős



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



Erdős:

1,400 papers

507 coauthors

Einstein: $EN=2$

Paul Samuelson $EN=5$

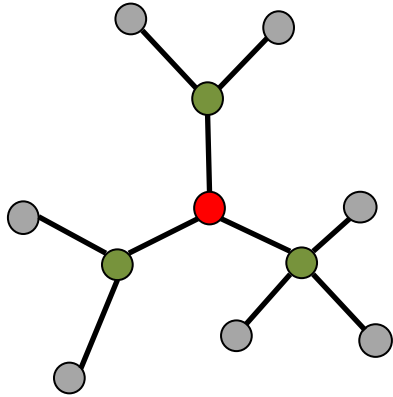
ALB: $EN=3$

CCV: $EN=4$

Ron Graham (alias Tom Odde).

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:

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

Ese es el origen de la propiedad small-world en grafos aleatorios.

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.

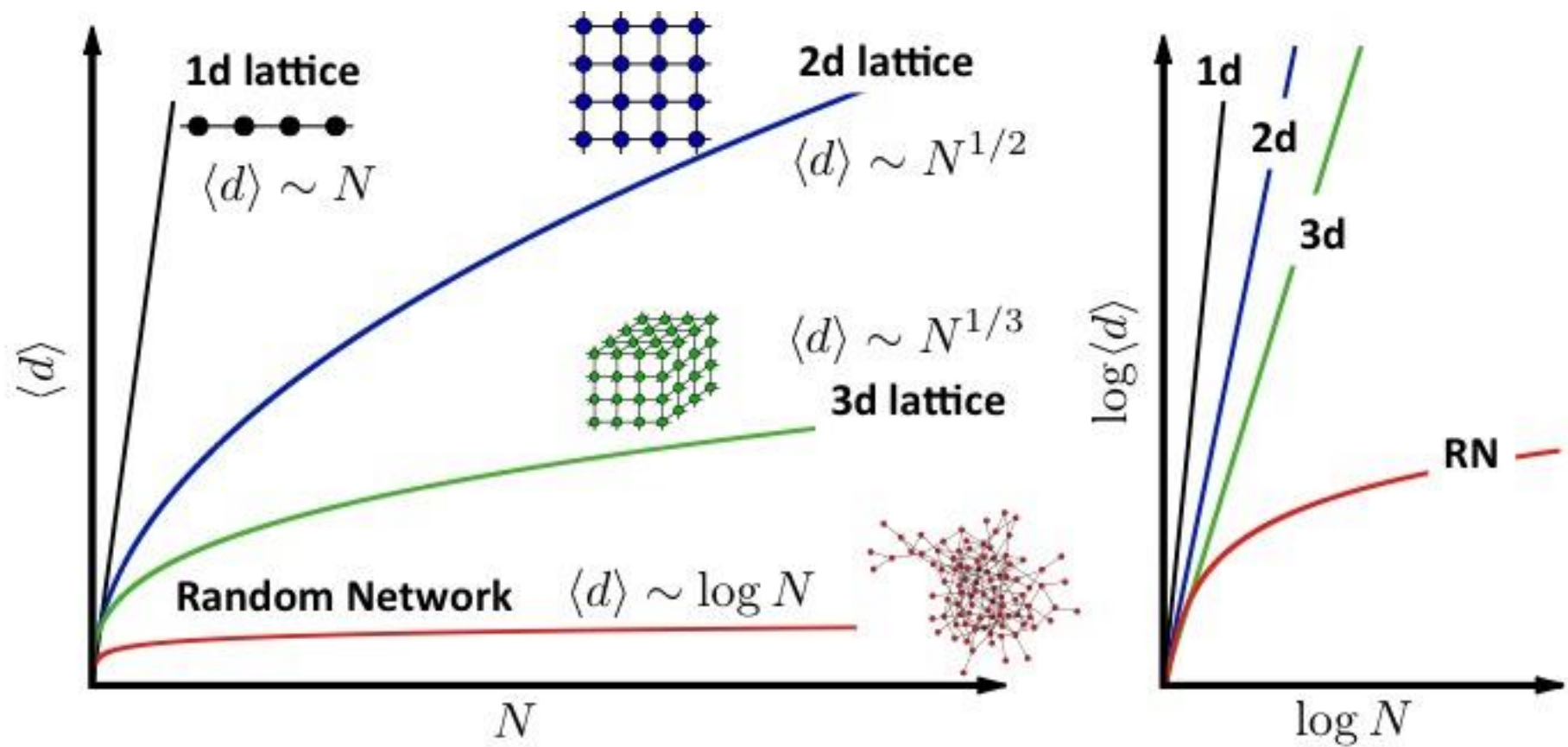
DISTANCIAS EN GRÁFOS ALEATORIOS

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

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?



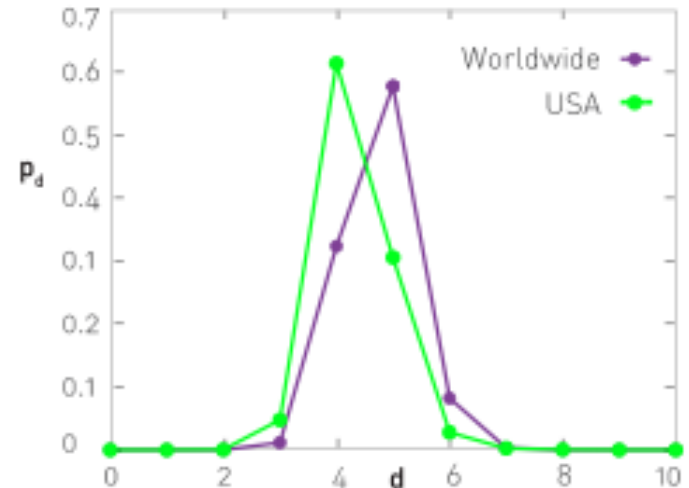
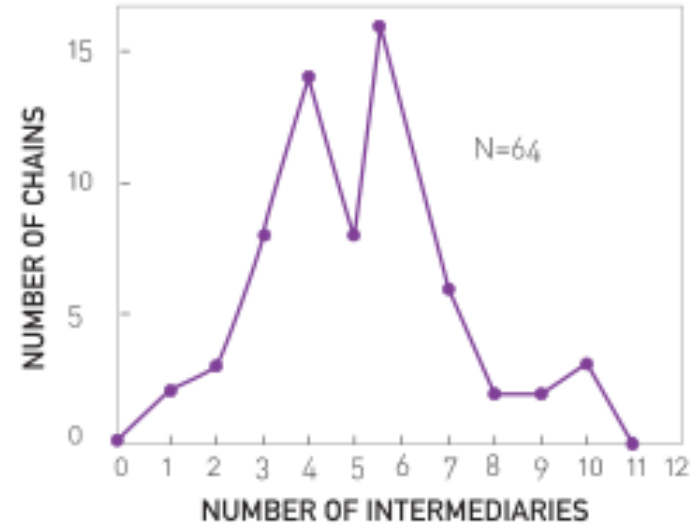
Tres, cuatro o seis grados?

Para las redes sociales del mundo:

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

$N \simeq 7 \times 10^9$ 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

MILESTONES

PUBLICATION DATE

1929

1935

1940

WWII

1945

1950

1958

1960

1967

1970

1978

1980

1985

1991

1998

XXI

2000

2005

2011



Frigyes Karinthy (1887-1938)

Hungarian writer, journalist and playwright, the first to describe the small world property. In his short story entitled 'Láncszemek' (Chains) he links a worker in Ford's factory to himself [23, 24].



Manfred Kochen



Ithiel de Sola Pool



Stanley Milgram

DISCOVERY

PUBLISHED 20 YEARS LATER

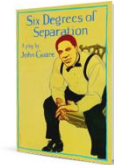
Manfred Kochen (1928-1989), Ithiel de Sola Pool (1917-1984)
Scientific interest in small worlds started with a paper by political scientist Ithiel de Sola Pool and mathematician Manfred Kochen. Written in 1958 and published in 1978, their work addressed in mathematical detail the small world effect, predicting that most individuals can be connected via two to three acquaintances. Their paper inspired the experiments of Stanley Milgram.

Stanley Milgram (1933-1984)
American social psychologist who carried out the first experiment testing the small-world phenomena. (BOX 3.6).



John Guare (1938)

The phrase 'six degrees of separation' was introduced by the playwright John Guare, who used it as the title of his Broadway play.



John Guare
6-DEGREE OF SEPARATION



Duncan J. Watts



Steven Strogatz



4-DEGREE OF SEPARATION

Duncan J. Watts (1971), Steven Strogatz (1959)
A new wave of interest in small worlds followed the study of Watts and Strogatz, finding that the small world property applies to natural and technological networks as well.

The Facebook Data Team measures the average distance between its users, finding "4 degrees" (BOX 3.6).

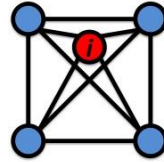
"Everybody on this planet is separated by only six other people. Six degrees of separation. Between us and everybody else on this planet. The president of the United States. A gondolier in Venice. It's not just the big names. It's anyone. A native in a rain forest. A Tierra del Fuegan. An Eskimo. I am bound to everyone on this planet by a trail of six people. It's a profound thought. How every person is a new door, opening up into other worlds."

Guare, 1991

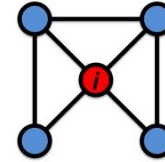
Coeficiente de Agrupamiento (Clustering)

COEFICIENTE DE AGRUPAMIENTO

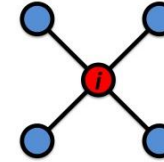
$$C_i \equiv \frac{2 \langle L_i \rangle}{k_i(k_i - 1)}$$



$$C_i = 1$$



$$C_i = 1/2$$



$$C_i = 0$$

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

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



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

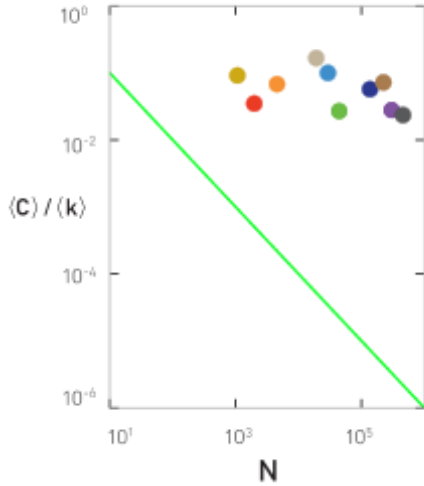
k_i =número de vecinos del nodo i

L_i =Número de enlaces que pasan por el nodo i

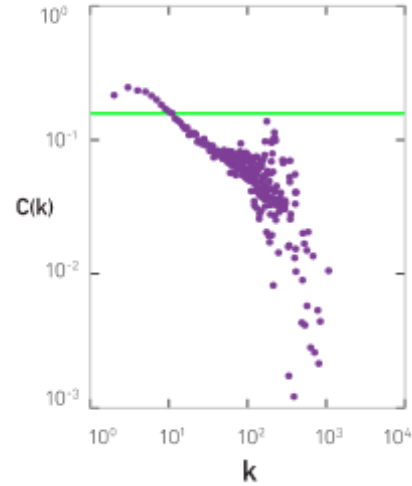
- El coeficiente de agrupamiento de los grafos aleatorios es pequeño.
- Para grados fijos, C disminuye con el tamaño del sistema N .
- C es independiente del grado k de nodos individuales.

COEFICIENTE DE AGRUPAMIENTO

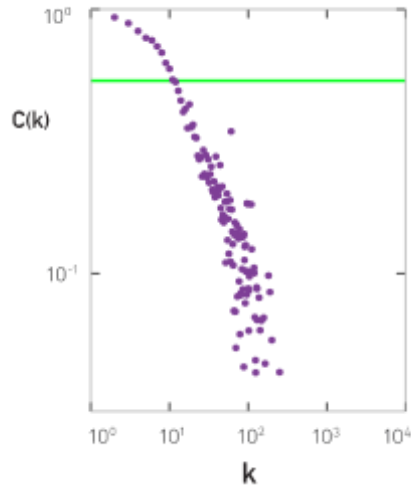
(a) All Networks



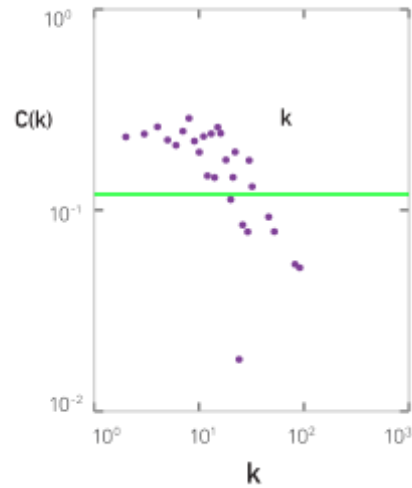
(b) Internet



(c) Science Collaboration



(d) Protein Interactions



$$C_1 = \frac{2\langle L \rangle}{k_1(k_1 - 1)} = p = \frac{\langle k \rangle}{N}$$

C disminuye con el tamaño del sistema N .

C es independiente del grado k de los nodos individuales.

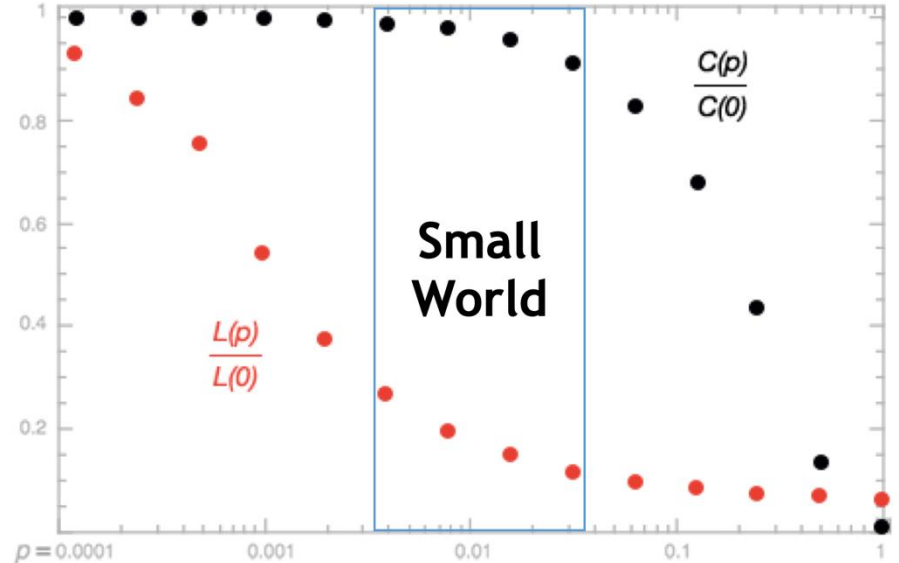
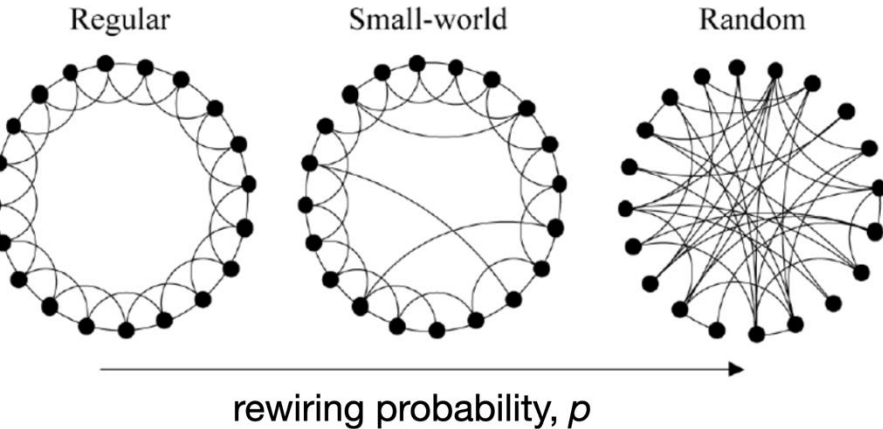
Small World Networks



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.



Watts–Strogatz (WS): qué es

WS parte de una red regular local y luego introduce rewiring o shortcuts aleatorios con probabilidad p .

La gracia del modelo es mostrar que un número pequeño de atajos puede reducir fuertemente las distancias sin destruir del todo la cohesión local. Newman muestra que en el small-world model el clustering disminuye con p , mientras la distancia media cae abruptamente.

Qué controla

- vecindad local
- estructura tipo lattice
- presencia gradual de shortcuts

Qué no controla bien

- hubs extremos
- crecimiento
- heterogeneidad de grado intensa
- colas de potencia robustas

WS: encuadre correcto

Tipo: modelo generativo estilizado, no nulo puro.

Qué explica

La coexistencia de:

- **alto clustering**
- **caminos cortos**

Es decir, el régimen **small-world**.

Por qué no es un nulo estricto

Porque no representa ausencia de mecanismo.

Ya incorpora una hipótesis fuerte: la red tiene una geometría local inicial y luego algunos atajos aleatorios.

Uso correcto

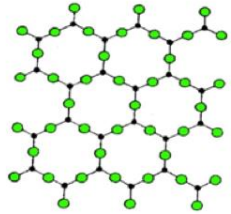
WS sirve como **baseline mecanístico** cuando la pregunta es:

¿basta una combinación de estructura local + pocos atajos para explicar mis distancias y clustering?

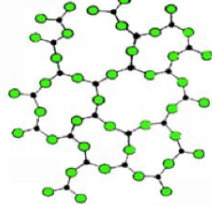
No sirve como nulo neutral de azar.

Scale-free Networks

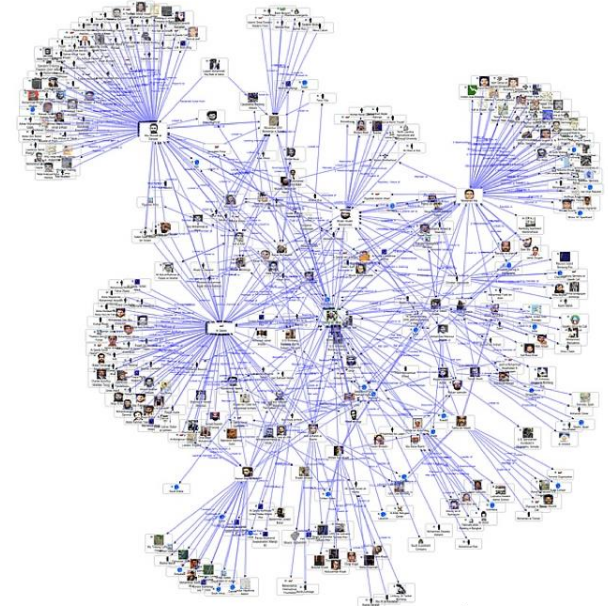
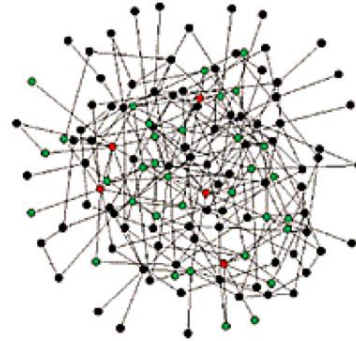
Desde sistemas físicos a sociales



Crystalline silica



Vitreous silica

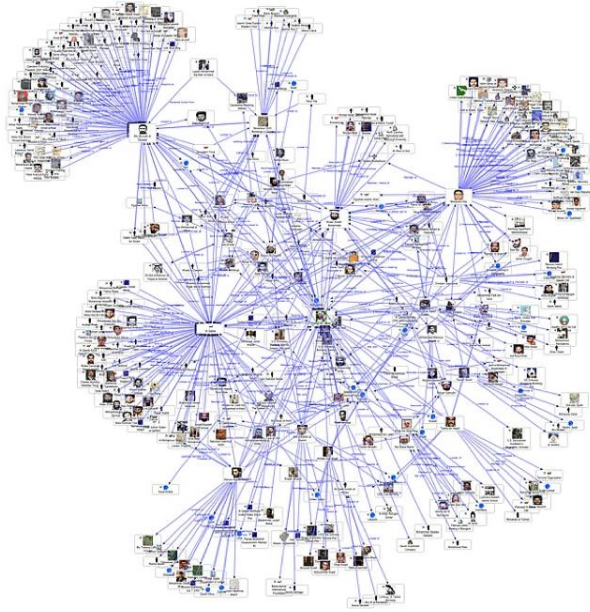


increasing network heterogeneity/complexity

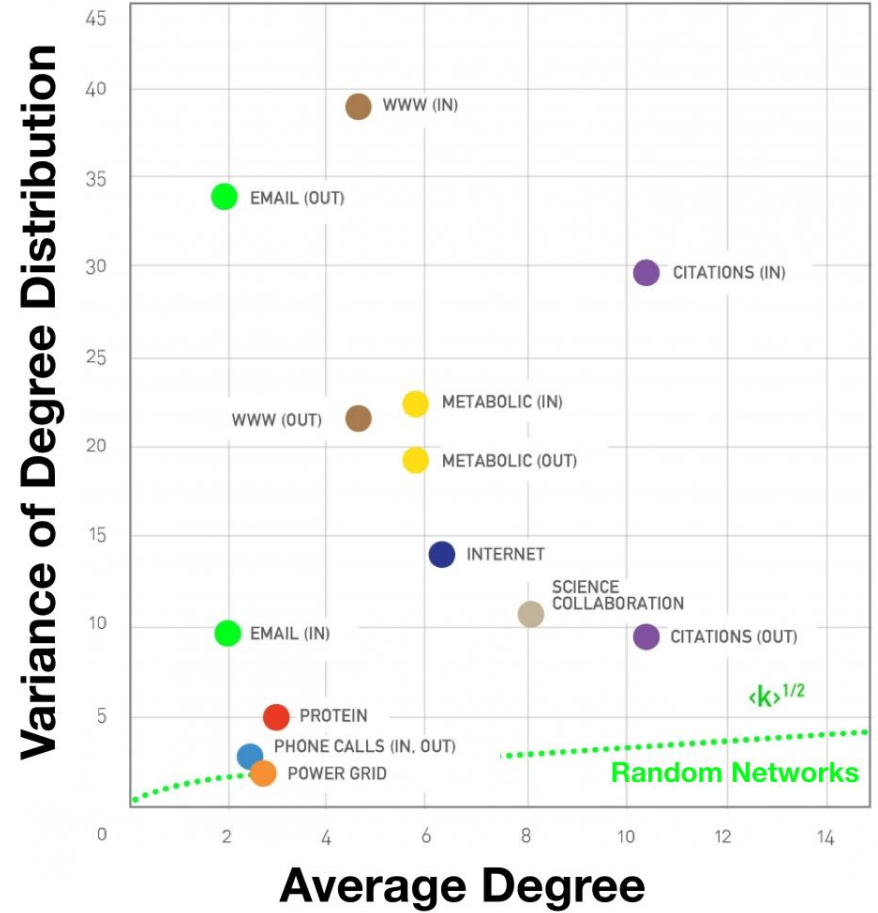
regular, lattices
physical systems

random structures
mathematics

complex structures
social systems



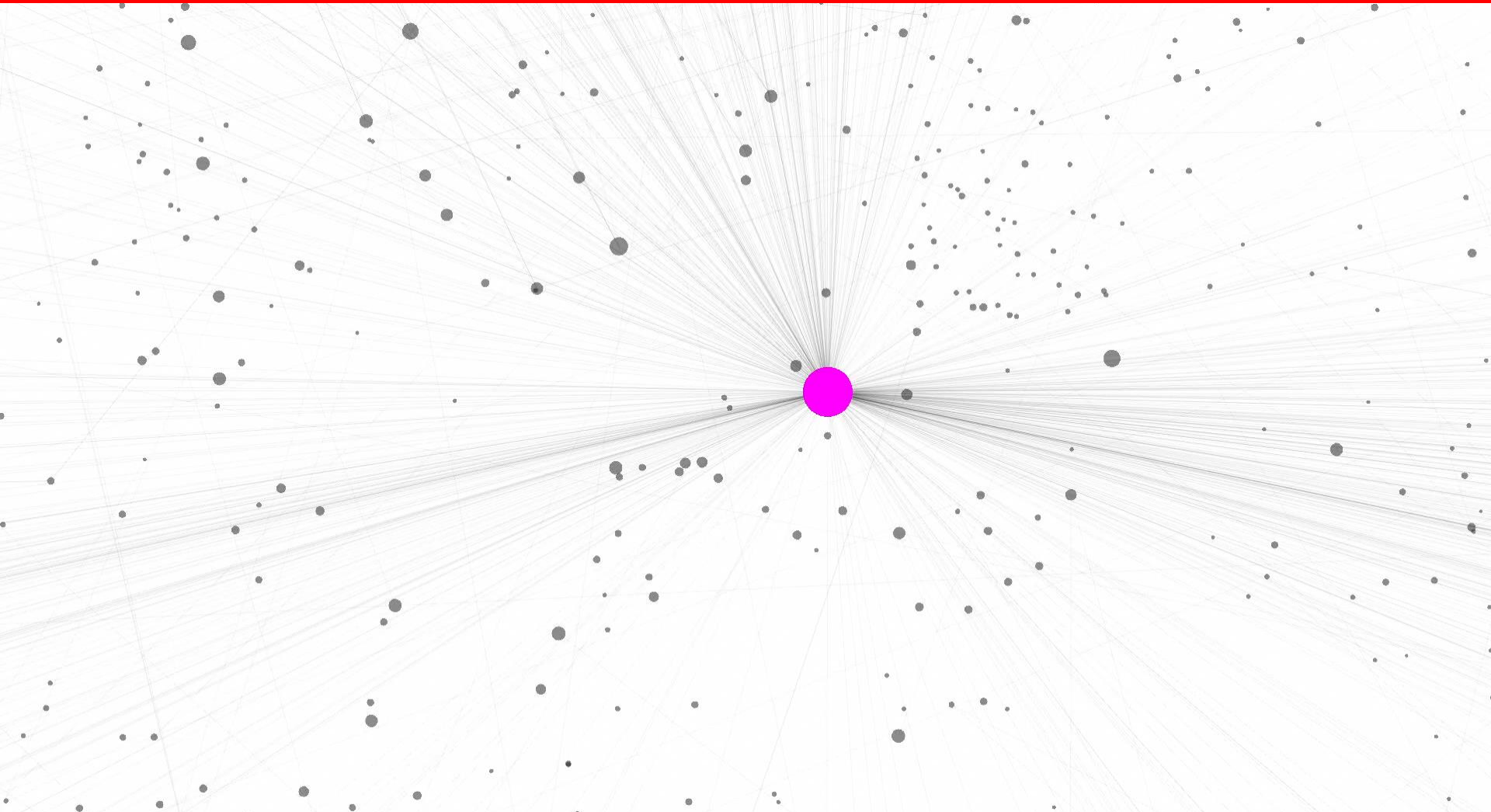
Las redes reales no son random son libre de escala!



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).

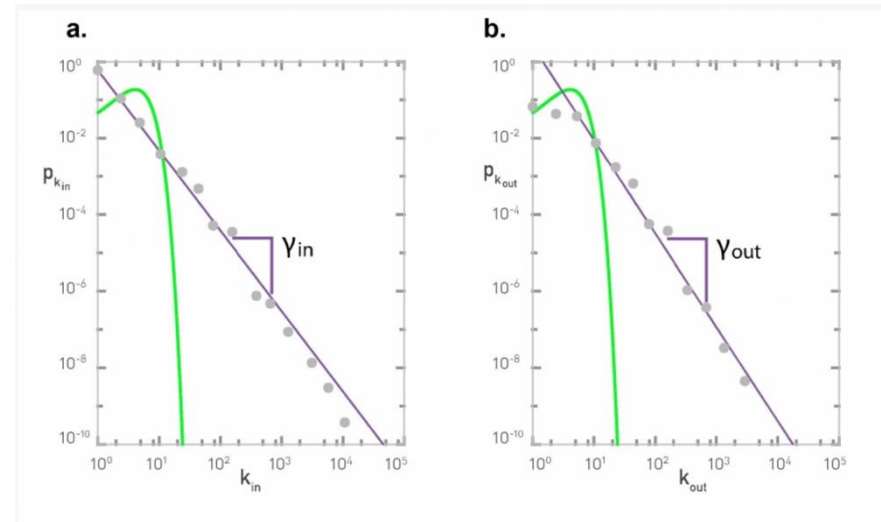
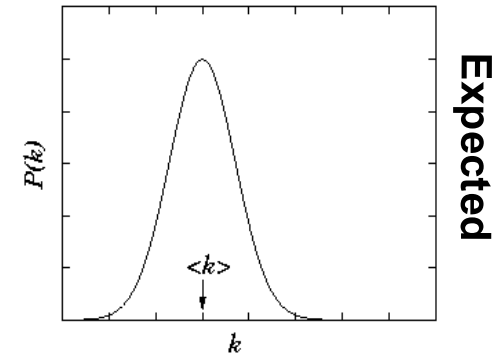
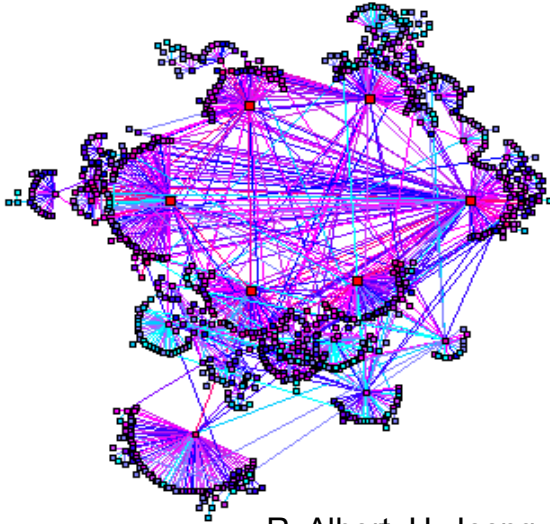
WORLD WIDE WEB

Nodos: **documentos WWW**

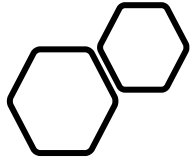
Enlaces: **links URL**

Más de 3 mil millones de documentos.

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

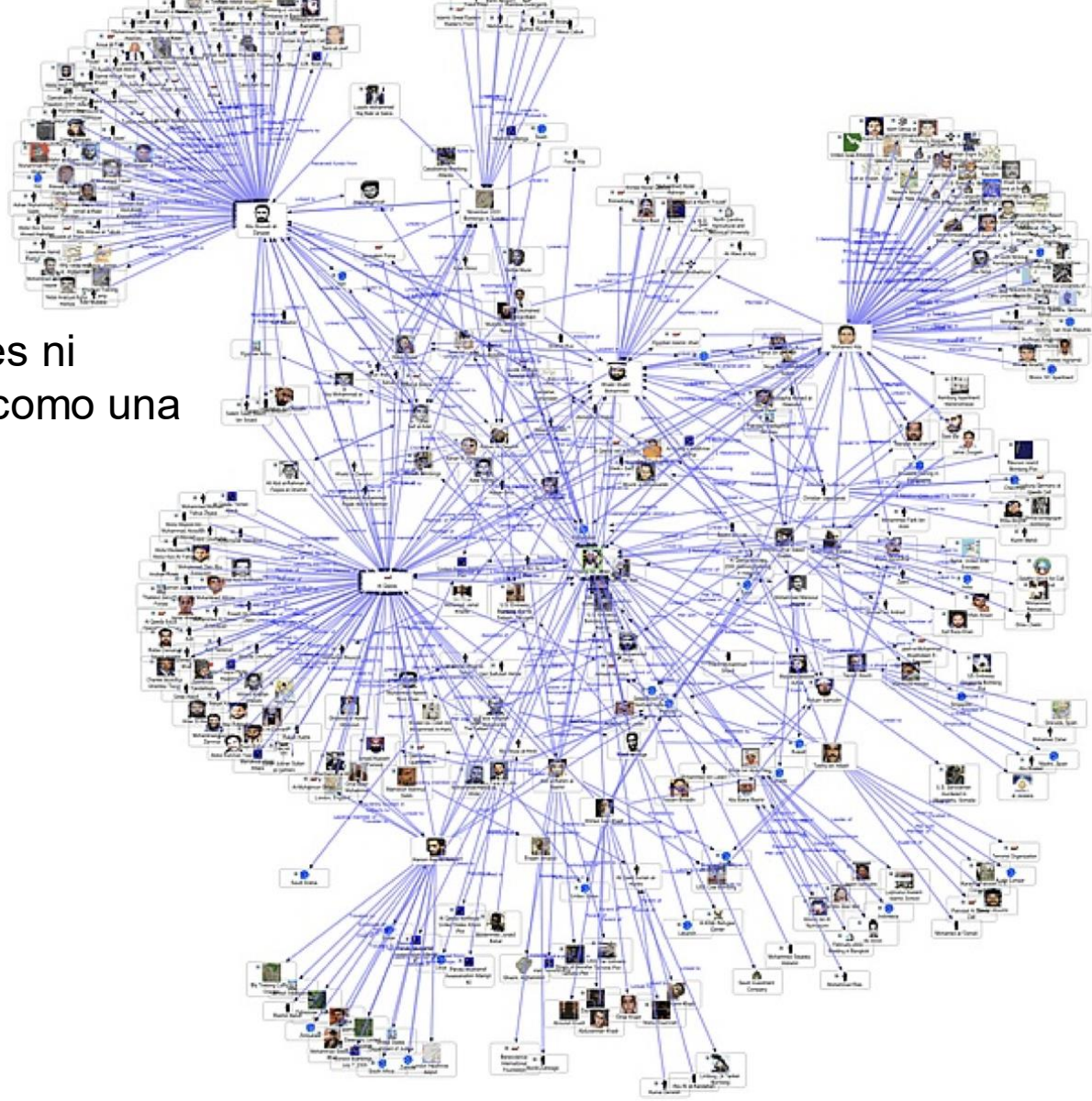


R. Albert, H. Jeong, A-L Barabasi, *Nature*, 401 130 (1999).



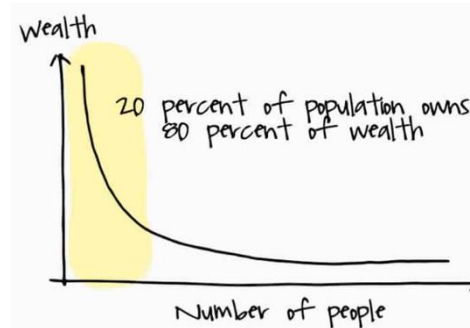
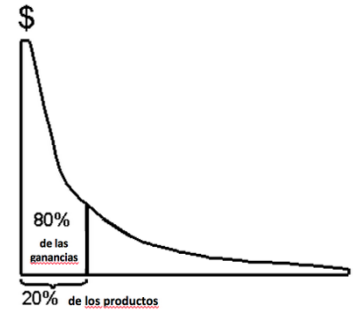
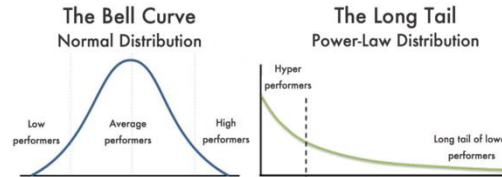
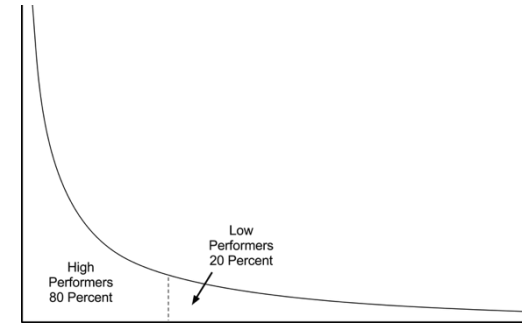
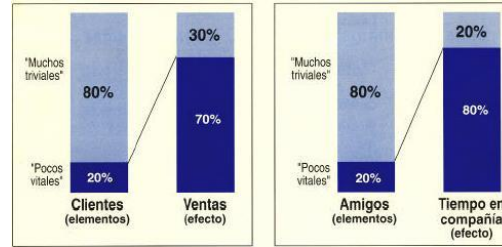
Las redes reales no son regulares ni aleatorias, ¡son mejor descritas como una red libre de escala!

- El nombre sin escala captura 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 ($\sigma = \langle k \rangle^{1/2}$)



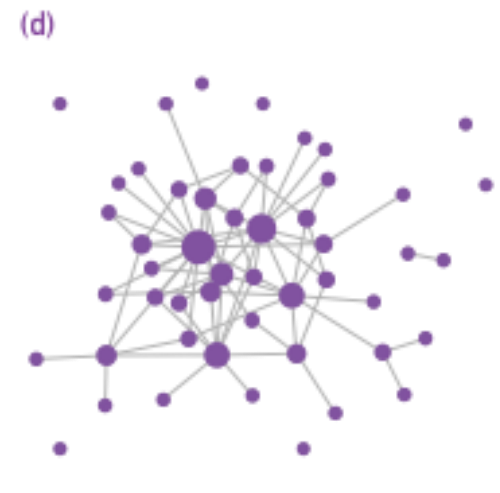
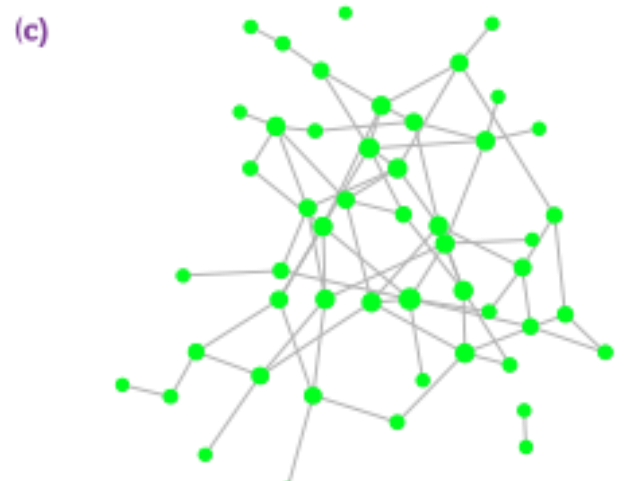
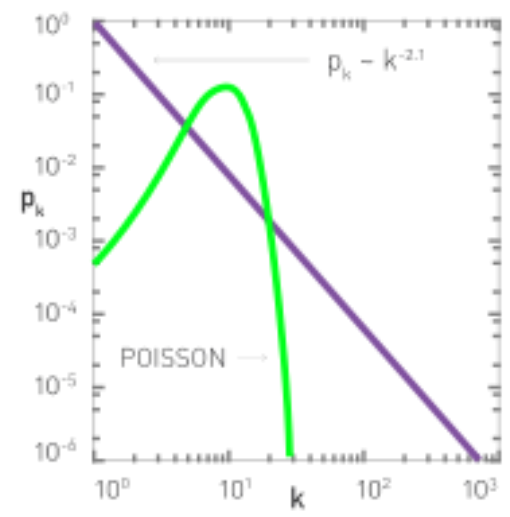
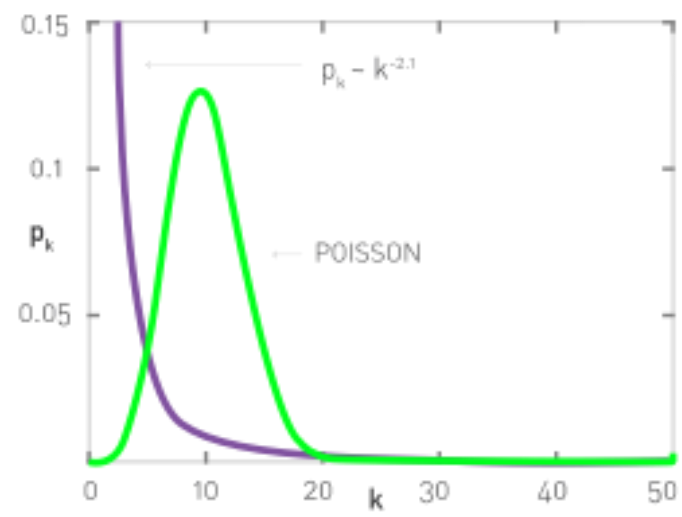
REGLA 80/20

- **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).



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.

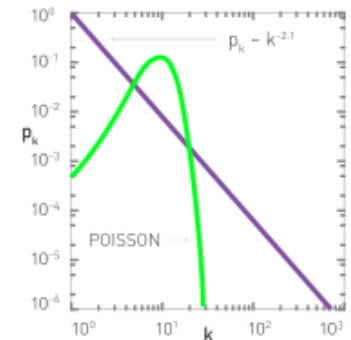
La probabilidad de tener un nodo con $k \sim 100$ es

• Cercana a $p_{100} \simeq 10^{-30}$ en una distribución de Poisson

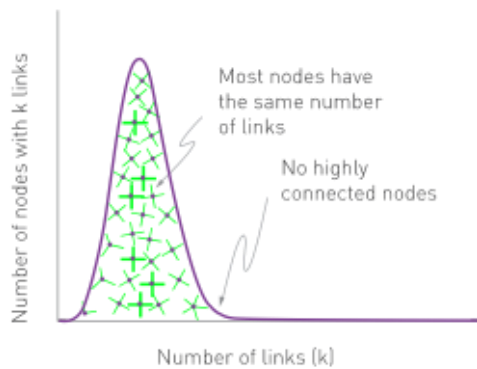
• Cercana a $p_{100} \simeq 10^{-4}$ si p_k sigue una ley de potencia

• 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.

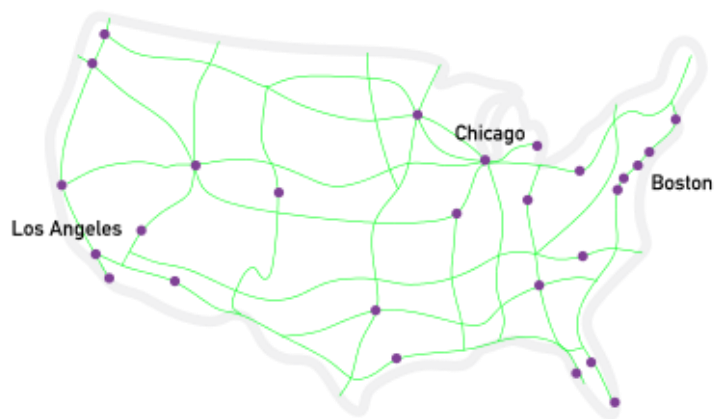
• Para una distribución de grado de ley de potencia, esperamos alrededor de $N_{k > 100} = 10^9$ nodos de grado $k > 100$



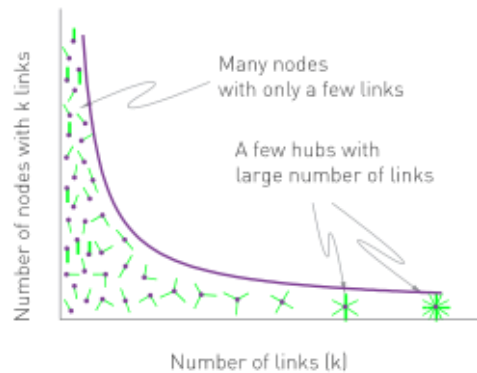
(a) POISSON



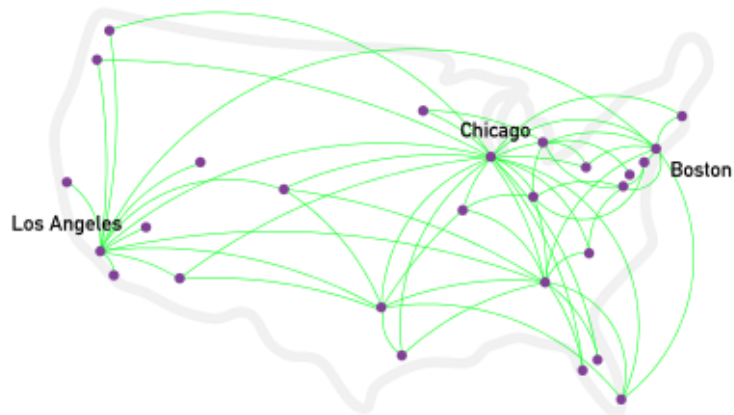
(b)



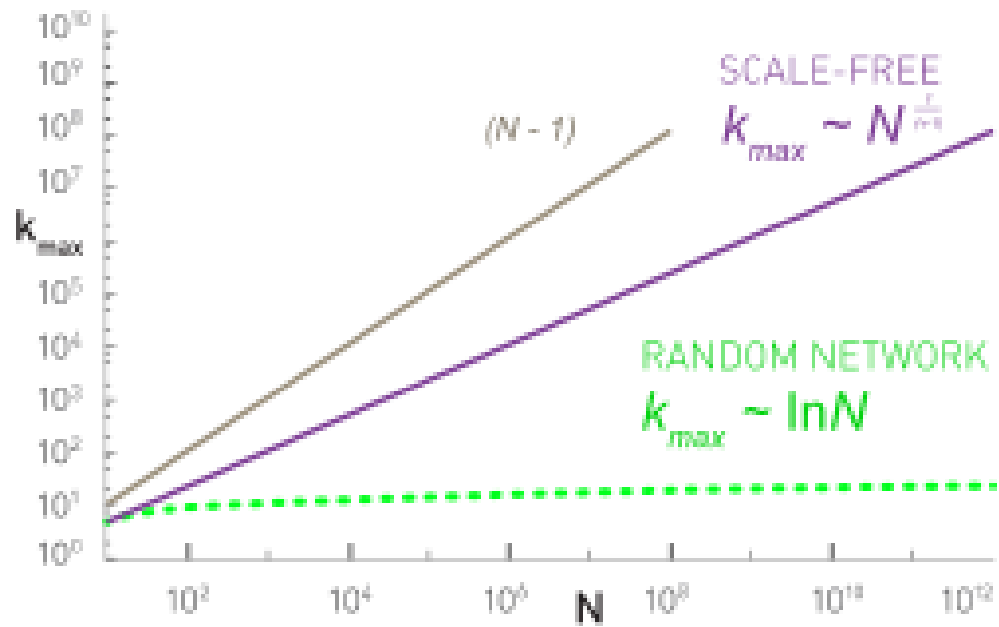
(c) POWER LAW



(d)



El tamaño del hub más grande.



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

DIVERGENCIA DE LOS MOMENTOS SUPERIORES

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

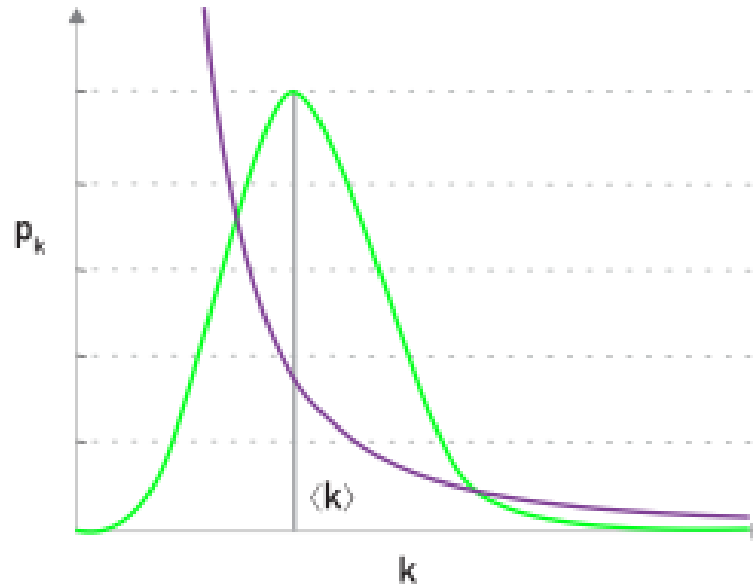
Para un γ fijo, esto significa que todos los momentos con $m > \gamma - 1$ divergen.

Network	Size	$\langle k \rangle$	κ	γ_{out}	γ_{in}
WWW	325 729	4.51	900	2.45	2.1
WWW	4×10^7	7		2.38	2.1
WWW	2×10^8	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	53×10^6	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	$\langle k \rangle$	$\langle k_{in}^2 \rangle$	$\langle k_{out}^2 \rangle$	$\langle k^2 \rangle$	γ_{in}	γ_{out}	γ
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 $N \rightarrow \infty$



Random Network

Nodo elegido al azar:

Scale: $\langle k \rangle$

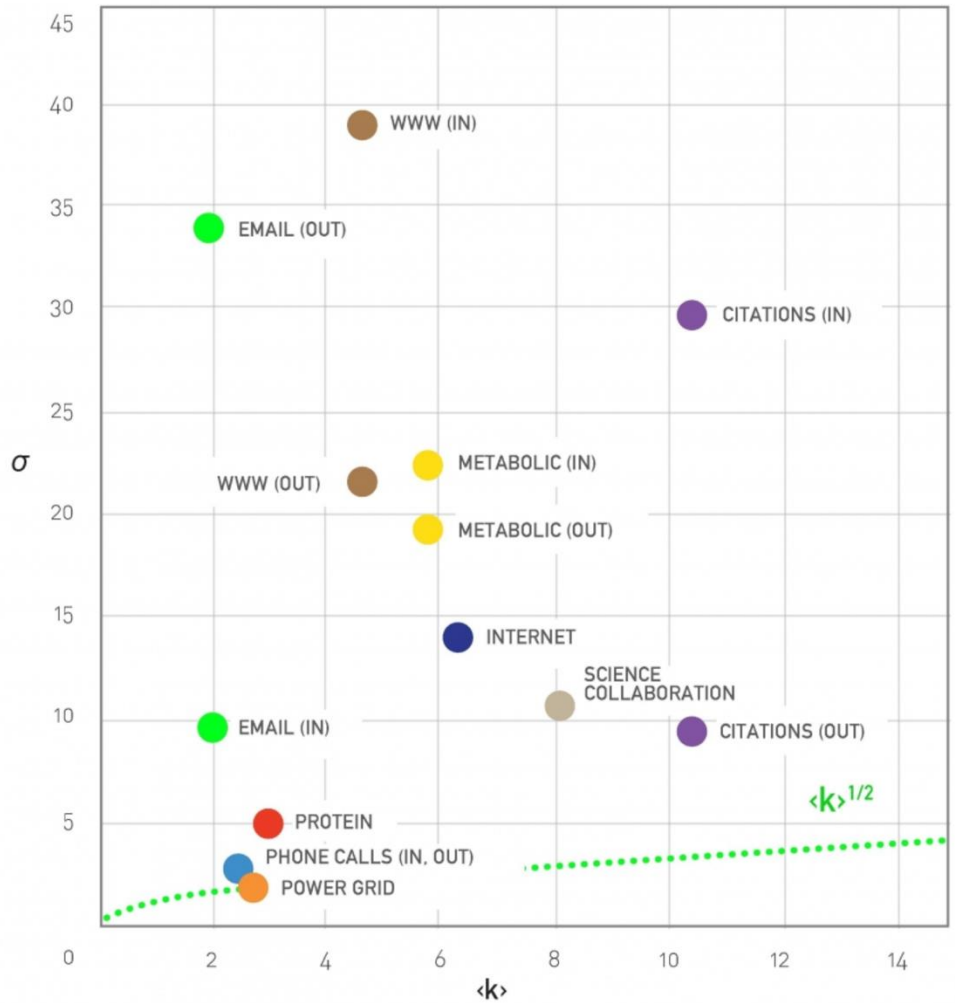
$$k = \langle k \rangle \pm \langle k \rangle^{1/2}$$

Scale-Free Network

Nodo elegido al azar:

Scale: none

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



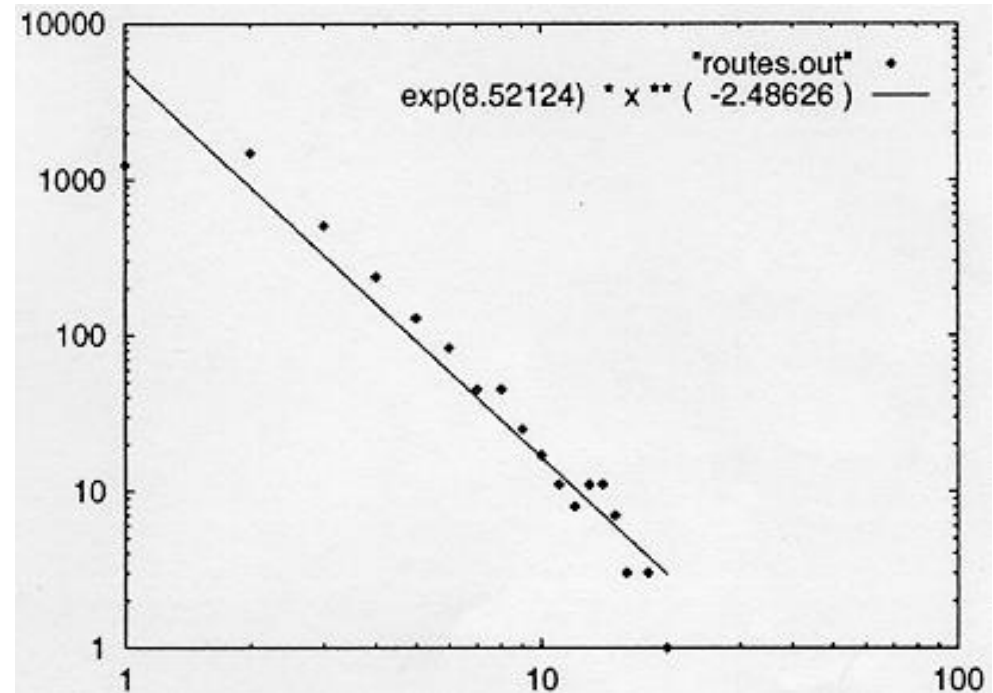
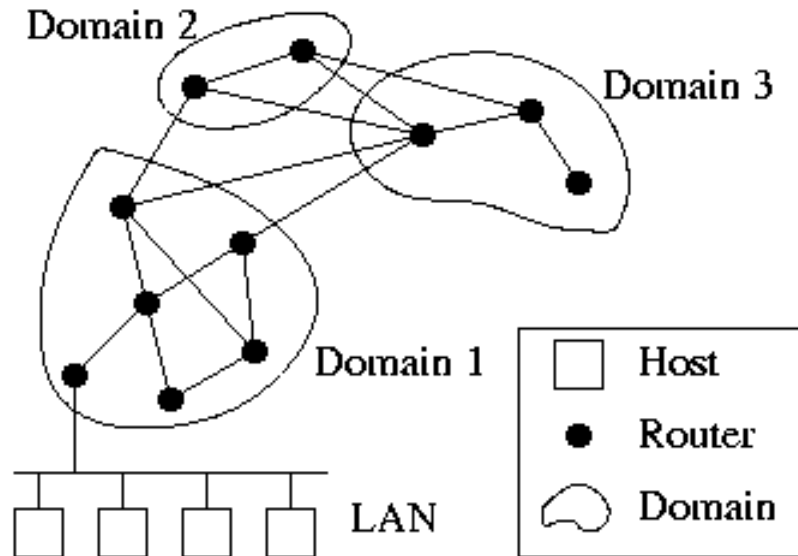
$$k = \langle k \rangle \pm \sigma_k$$

Universalidad

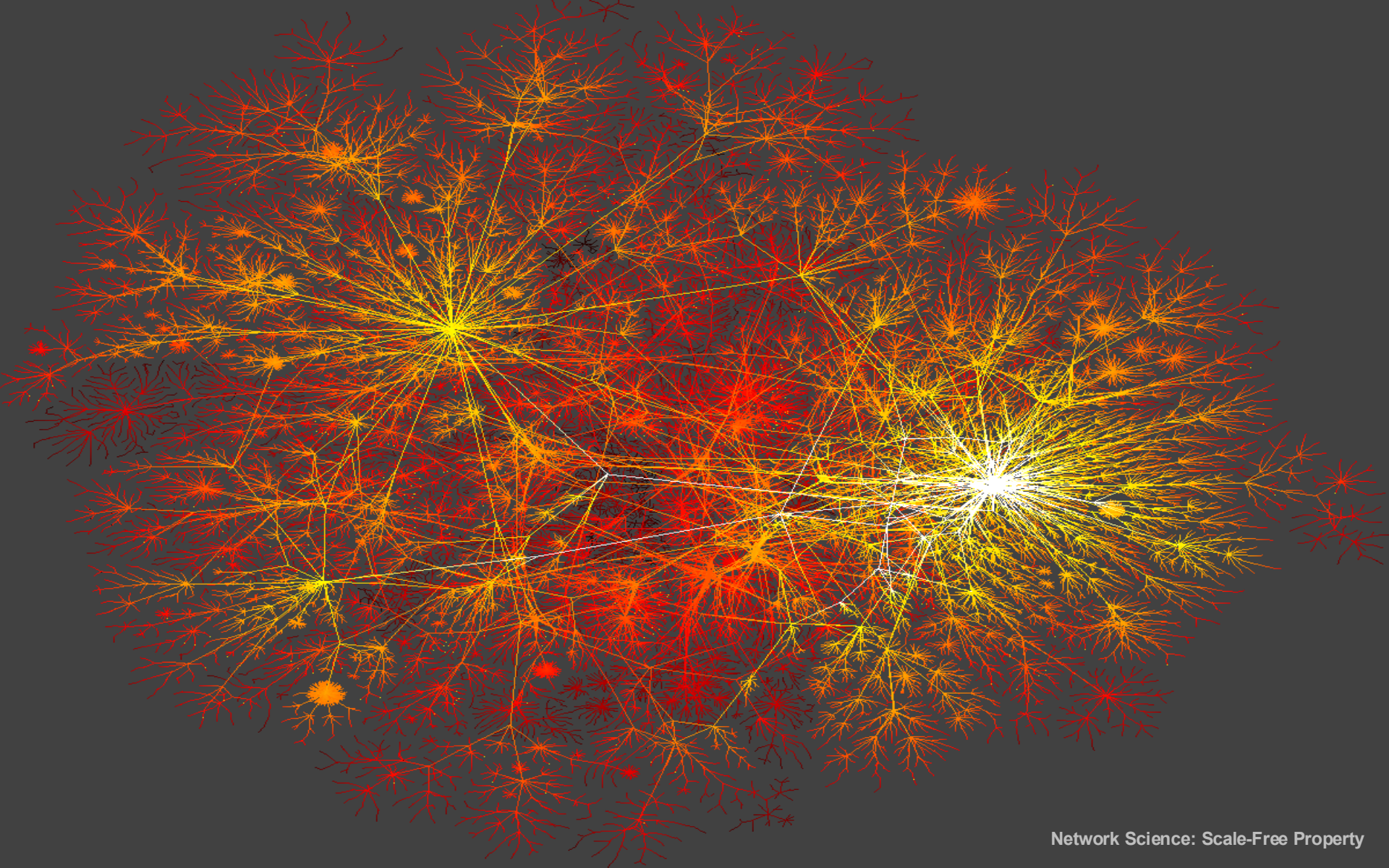
Backbone de internet

Nodos: computadores, routers

Enlaces: líneas físicas



(Faloutsos, Faloutsos and Faloutsos, 1999)

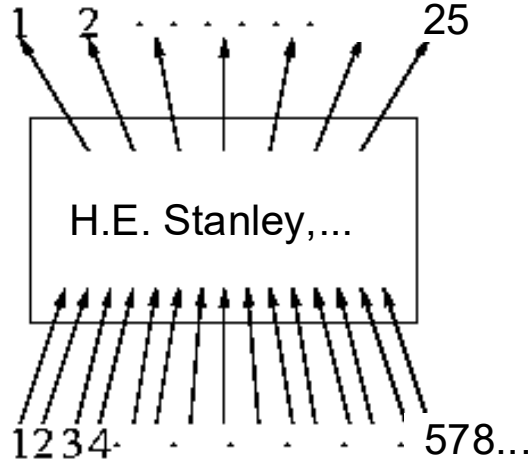


INDICE DE CITACIONES EN CIENCIA

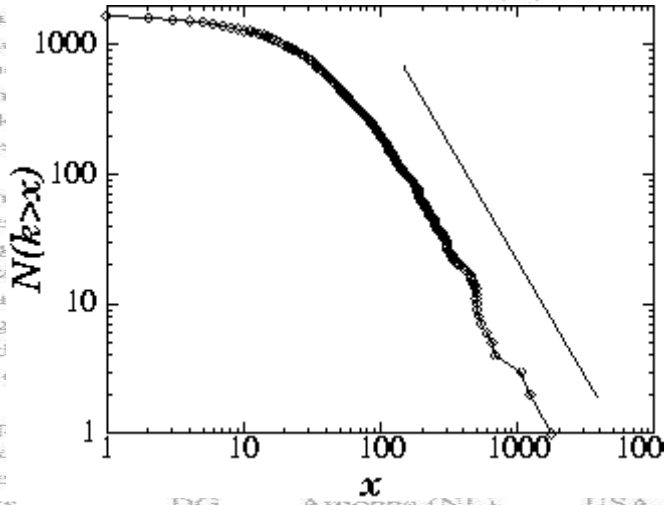
Out of over 500,000 Examined
(see <http://www.sst.nrel.gov>)

Nodos: papers
Enlaces: citaciones

Author	Institute	Country	Field	avg. cites	total art.	total cites	rank by total cit.
Witter	Princeton (U)	USA, NJ	High-energy (P)	168	138	23235	1
... ..	UCSB (U)	USA, CA	Semiconductors (E)				2
Cava	Bell Labs (I)	USA, NJ	Superconductivity (E)				3
Ballogg	Bell Labs (I)	USA, NJ	Superconductivity (E)				4
Ploog	Max-Planck (NL)	Germany	Semiconductors (E)				5
Ellis	Euro Nuclear Cent.	Switzerland	Astrophysics (E)				6
Fisk	Florida State (U)	USA, FL	Solid State (E)				7
Cardona	Max Planck (NL)	Germany	Semiconductors (E)				8
Nanopoulos	Texas A&M (U)	USA, TX	High-energy (P)				9
Heeger	UCSB (U)	USA, CA	Polymer Physics (E)				10
Lee*							11
Suzuki*							12
Anderson							13
Suzuki*							14
Freeman							15
Tana							16
Mull							17
Schn							18
Chen							19
Morl							20
Mille							21
Chu				44	213	9453	22
Bedn				49	85	9311	23
Cole				284	9311	9311	23
Metz				86	108	9300	25
Wass				57	162	9170	26
Shira				33	269	8841	27
Wieg				85	104	8822	28
Vand				67	129	8686	29
Uchi				28	301	8520	30
Hor				72	119	8512	31
Mur				111	76	8439	32
Birge				41	286	8375	33
Jorge				50	167	8298	34
Hinks	Argonne (NL)	USA, IL	Superconductivity (E)	37	223	8263	35



1736 PRL papers (1988)



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

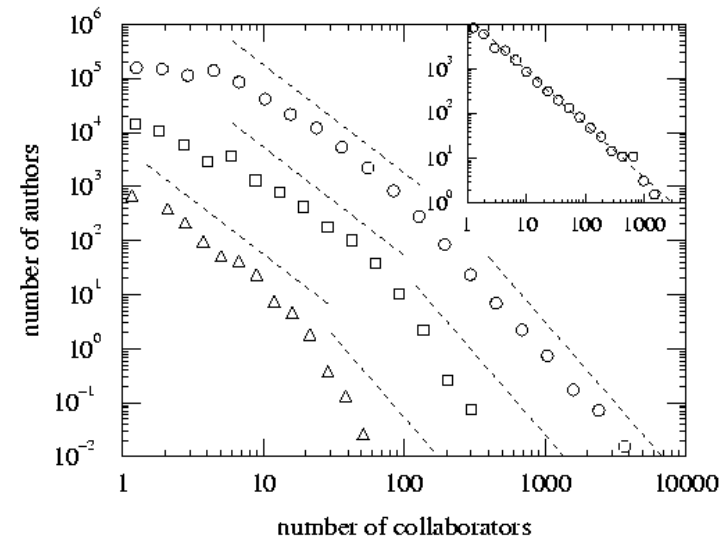
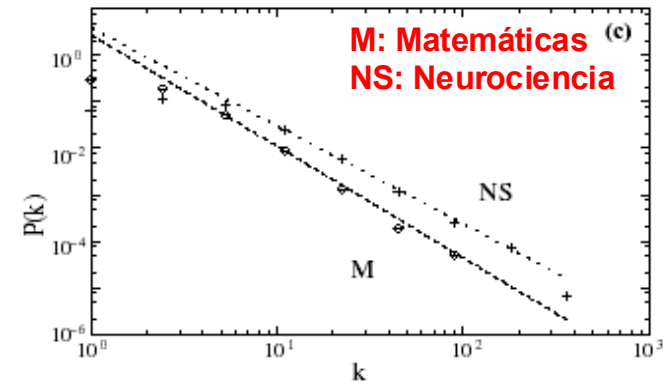
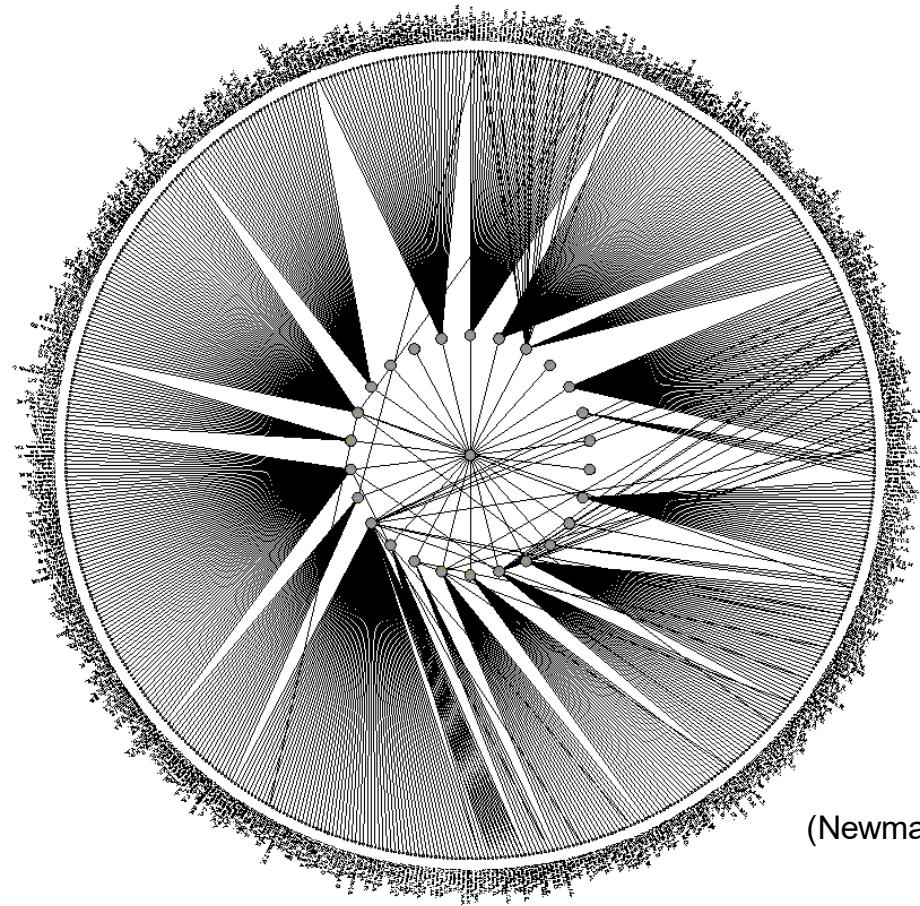
(S. Redner, 1998)

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

CO-AUTORIA EN CIENCIA

Nodos: Científicos (autores)

Links: Publicaciones conjuntas

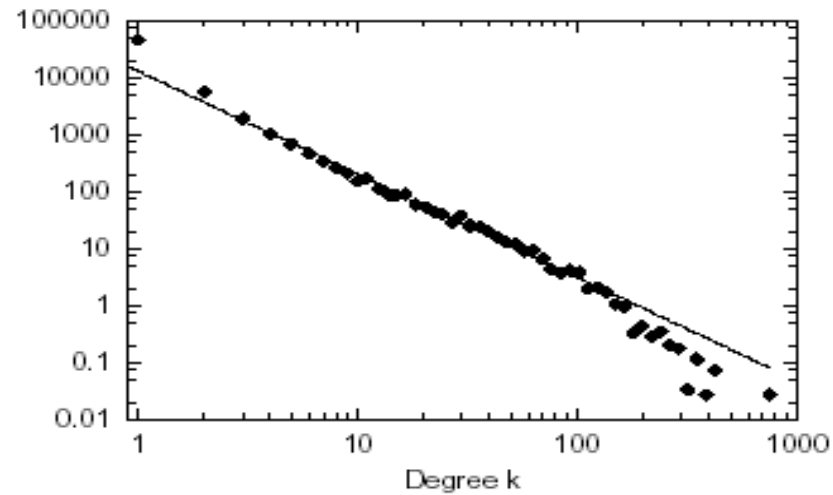


(Newman, 2000, Barabasi et al 2001)

COMUNIDADES ONLINE

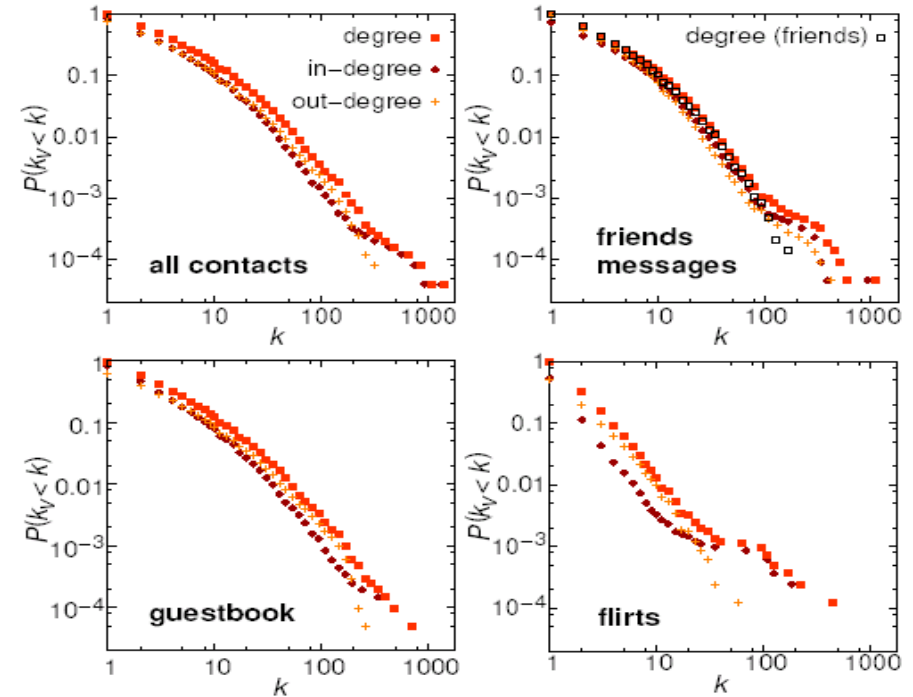
Nodos: Usuarios online
Enlaces: contactos email

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



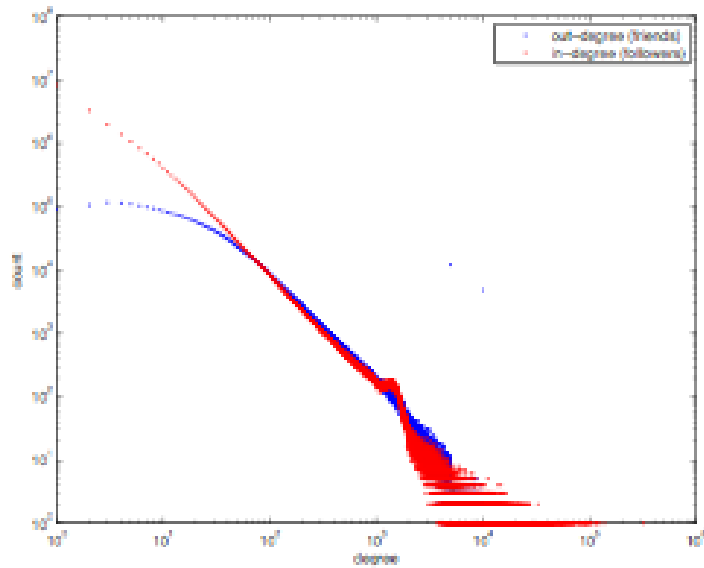
Ebel, Mielsch, Bornholdtz, PRE 2002.

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

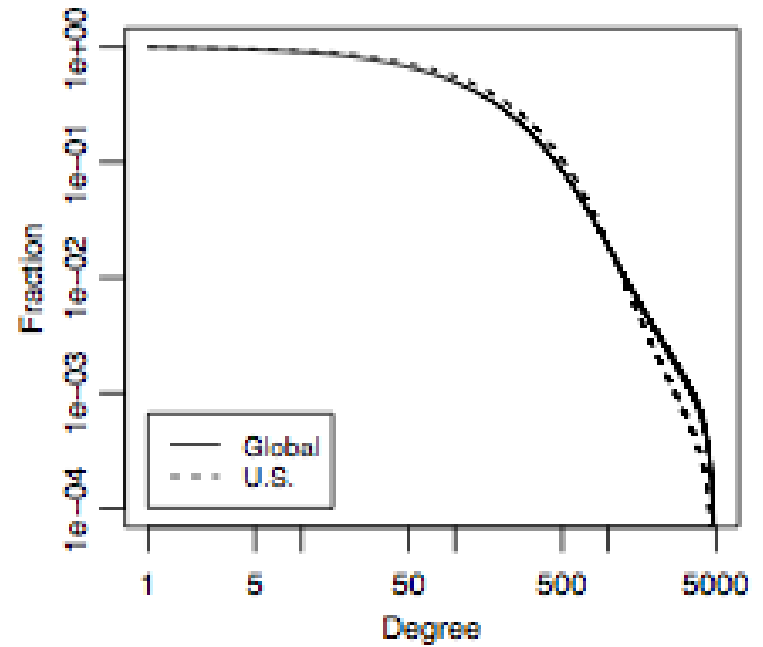


Holme, Edling, Liljeros, 2002.

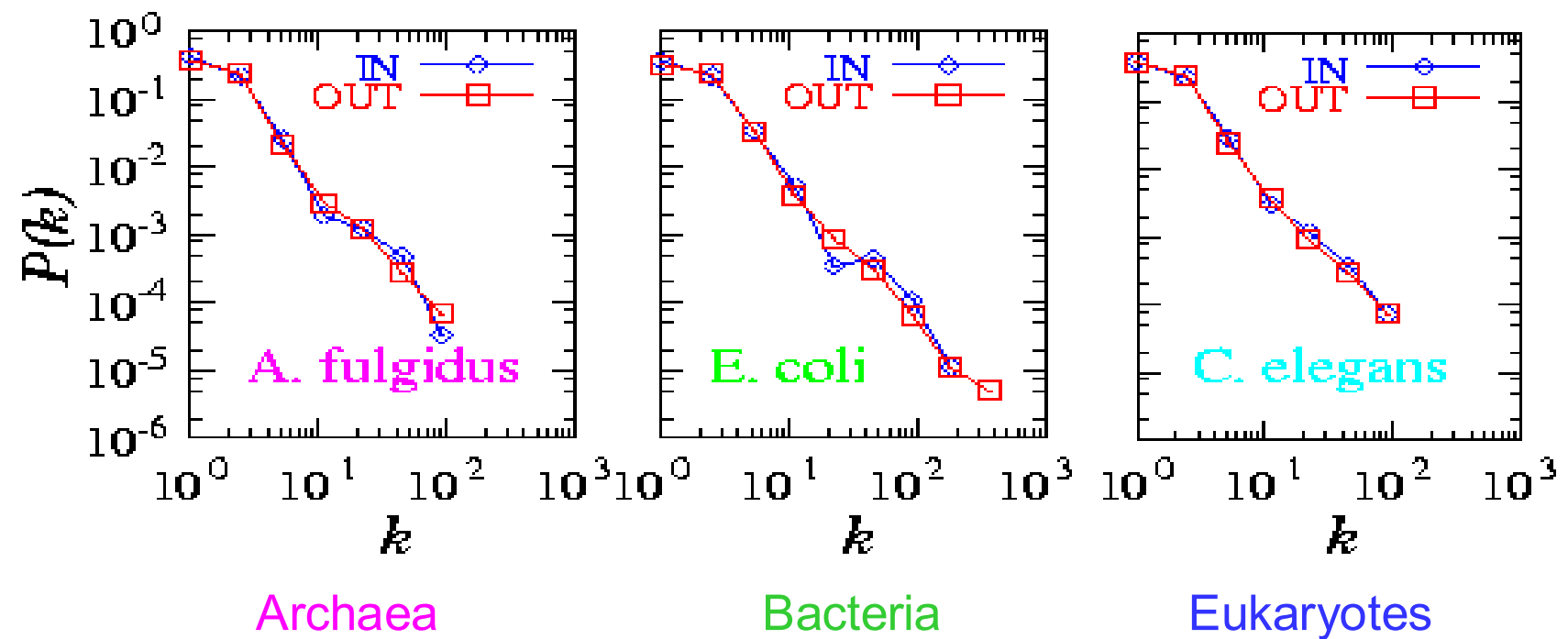
Twitter:



Facebook



RED METABOLICA

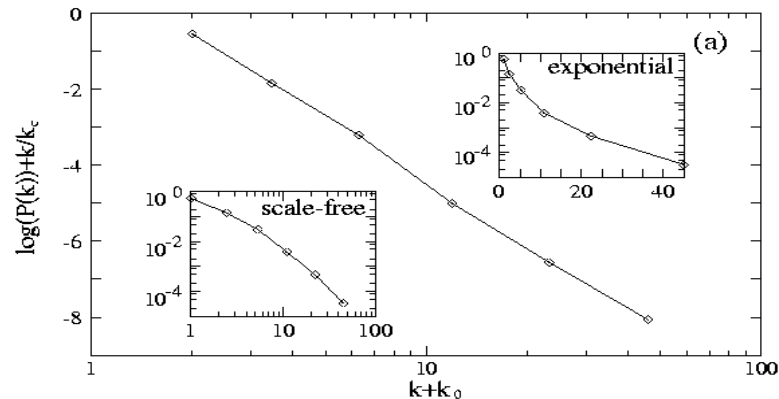


¡Los organismos de los tres dominios de la vida son **libres de escala!**

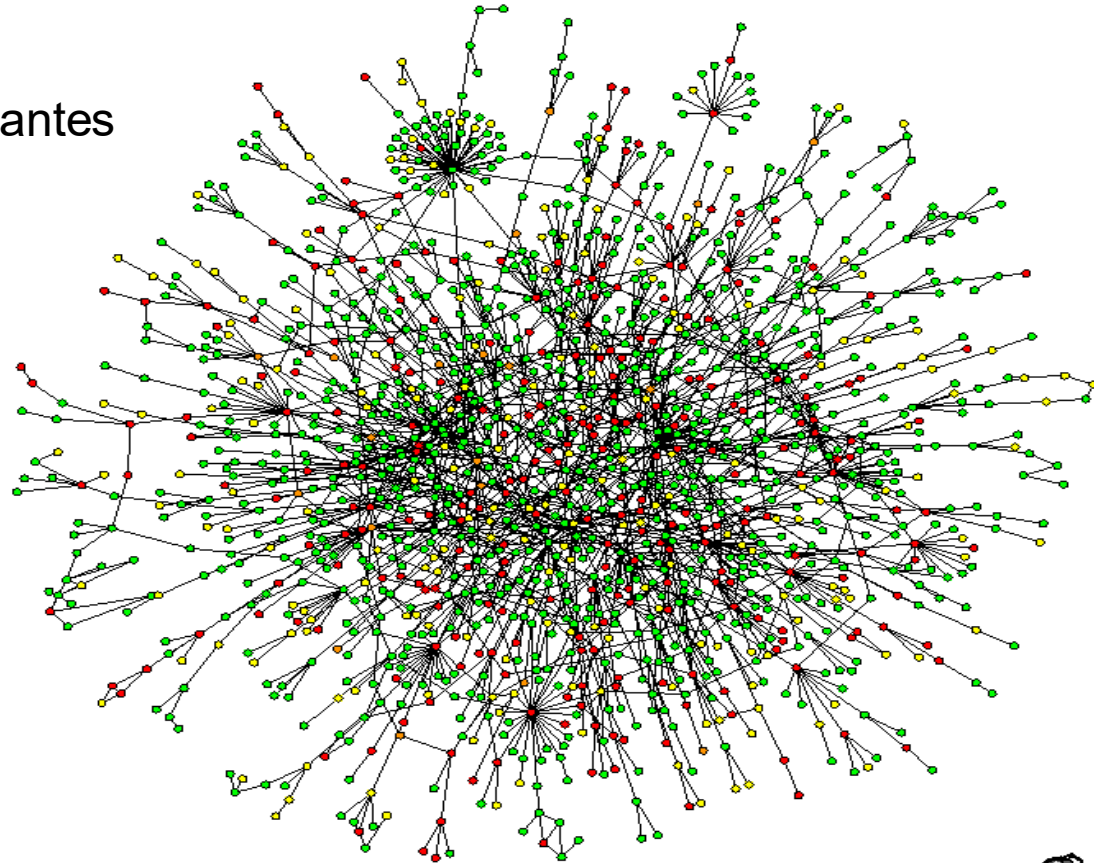
$$P_{in}(k) \approx k^{-2.2}$$
$$P_{out}(k) \approx k^{-2.2}$$

TOPOLOGÍA DE LA RED DE PROTEÍNAS

Nodos: proteínas
Enlaces: interacciones físicas vinculantes



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

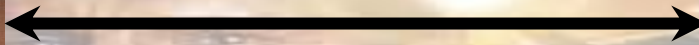
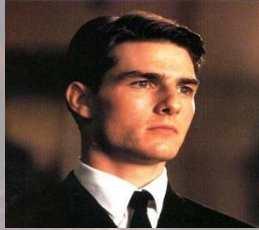


Paieck

RED DE ACTORES

Nodos: actores

Enlaces: película conjunta



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

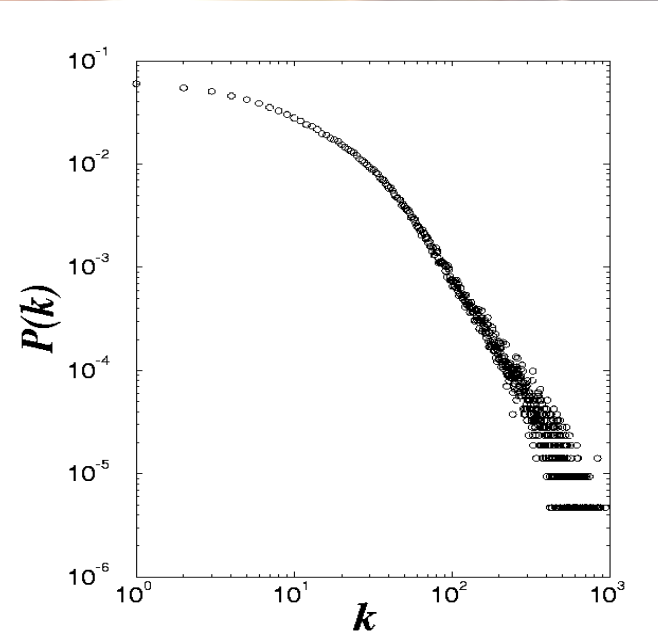


$N = 212,250$ actores

$\langle k \rangle = 28.78$

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

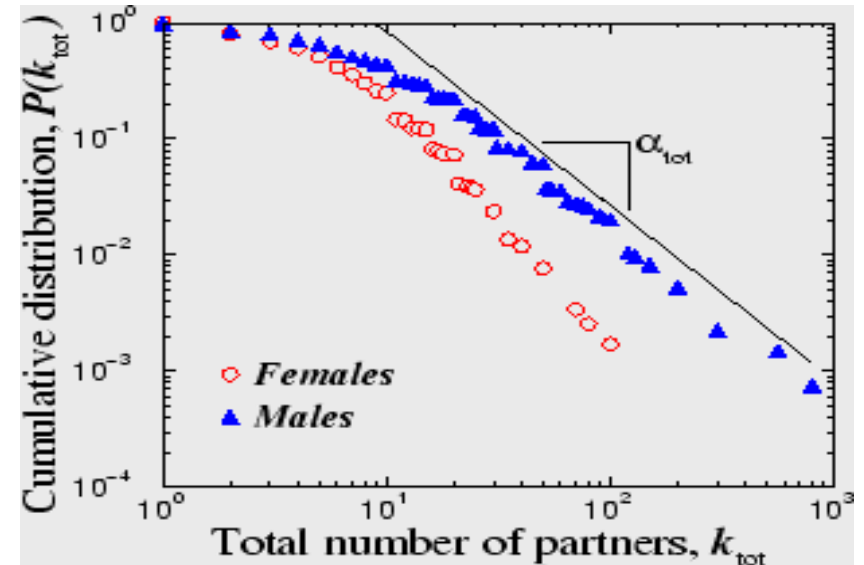
$\gamma = 2.3$





Nodos: Personas (Mujeres; Hombres)

Enlaces: Relaciones sexuales

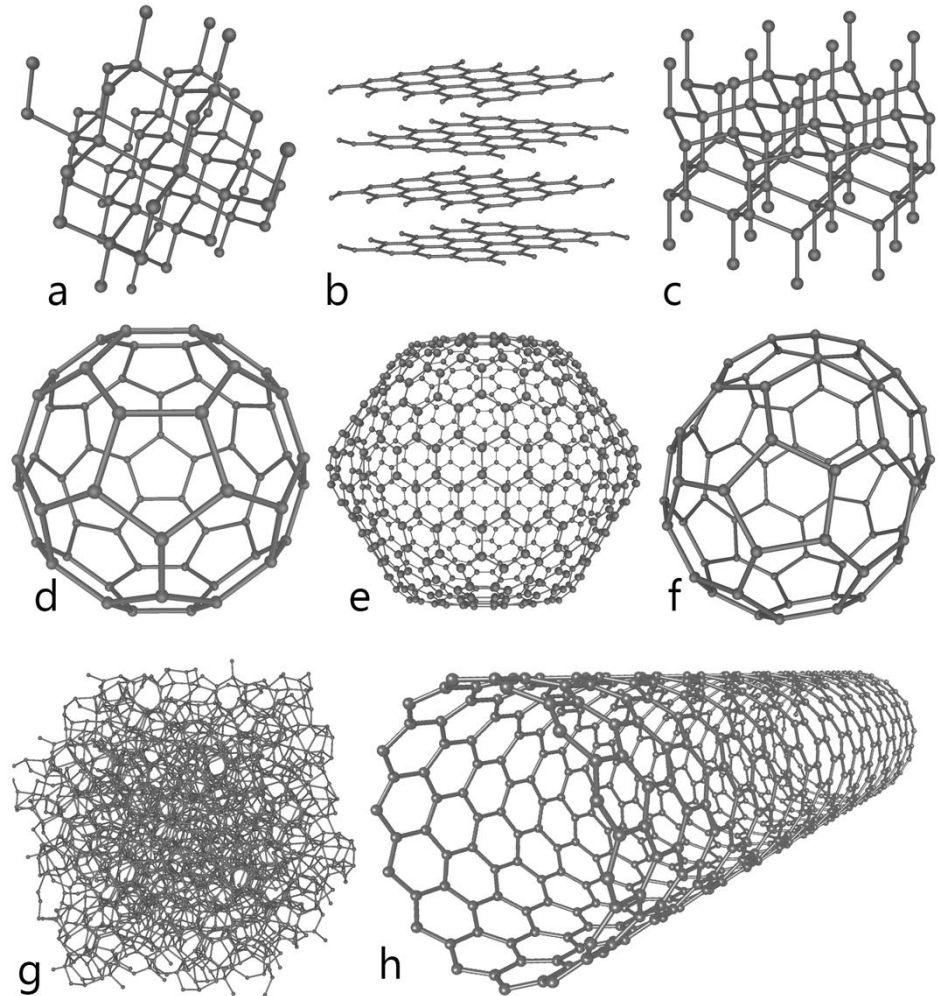


4781 Suizos; 18-74;
59% tasa de respuesta.

Liljeros et al. Nature 2001

No todas las redes pueden ser descritas como libres de escala

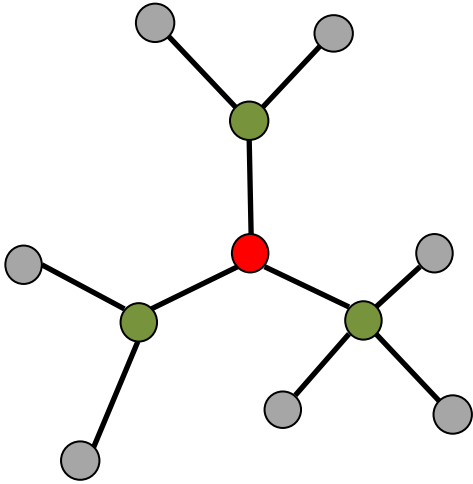
- *Redes que aparecen en la ciencia de los materiales, como la red que describe los enlaces entre los átomos en materiales cristalinos o amorfos, donde cada nodo tiene exactamente el mismo grado.*
- *La red neuronal del gusano *C.elegans*.*
- *La red eléctrica, compuesta por generadores e interruptores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



Propiedad ultra-pequeña

DISTANCIAS EN GRAFOS ALEATORIOS

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



•nr de primeros vecinos

$$N_1 \cong \langle k \rangle$$

•nr de segundos vecinos:

$$N_2 \cong \langle k \rangle^2$$

•nr de vecinos a distancia d:

$$N_d \cong \langle k \rangle^d$$

• estimar la distancia máxima:

$$1 + \sum_{l=1}^{l_{max}} \langle k \rangle^l = N \implies l_{max} = \frac{\log N}{\log \langle k \rangle}$$

COMPORTAMIENTO DE MUNDO PEQUEÑO EN REDES SIN ESCALA

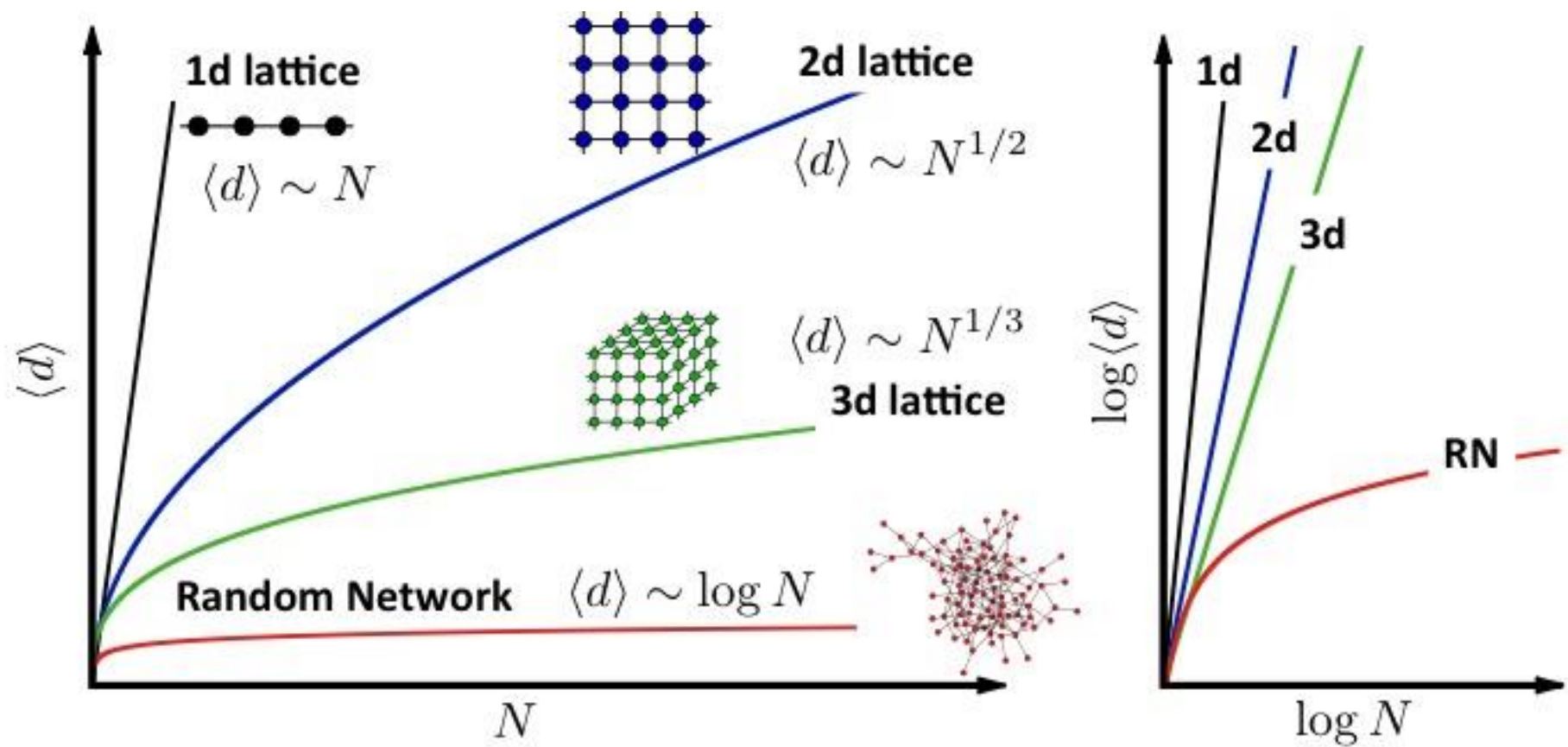
$$k_{\max} = k_{\min} N^{\frac{1}{\gamma-1}}$$

Ultra Small World	{	<i>const.</i>	$\gamma = 2$	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.
		$\frac{\ln \ln N}{\ln(\gamma - 1)}$	$2 < \gamma < 3$	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 libre de escala, la gran mayoría de la ruta pasa por unos pocos hubs de alto grado, lo que reduce las distancias entre los nodos.
Small World	{	$\frac{\ln N}{\ln \ln N}$	$\gamma = 3$	Algunos modelos clave producen $\gamma = 3$, 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.
		$\ln N$	$\gamma > 3$	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 aleatorio.

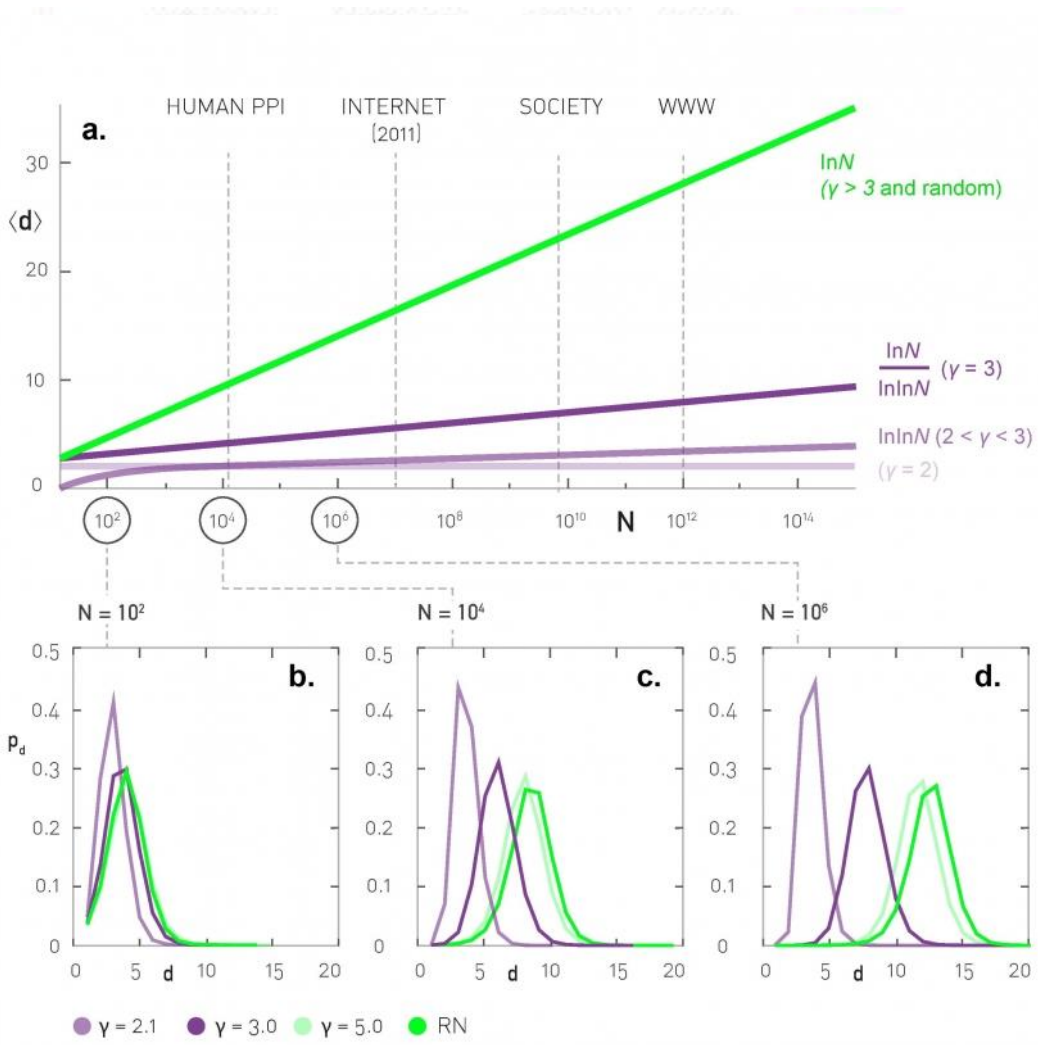
$\langle l \rangle \sim$

Cohen, Havlin Phys. Rev. Lett. 90, 58701(2003); Cohen, Havlin and ben-Avraham, in Handbook of Graphs and Networks, Eds. Bornholdt and Shuster (Wiley-VCH, NY, 2002) Chap. 4; Confirmed also by: Dorogovtsev et al (2002), Chung and Lu (2002); (Bollobas, Riordan, 2002; Bollobas, 1985; Newman, 2001

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

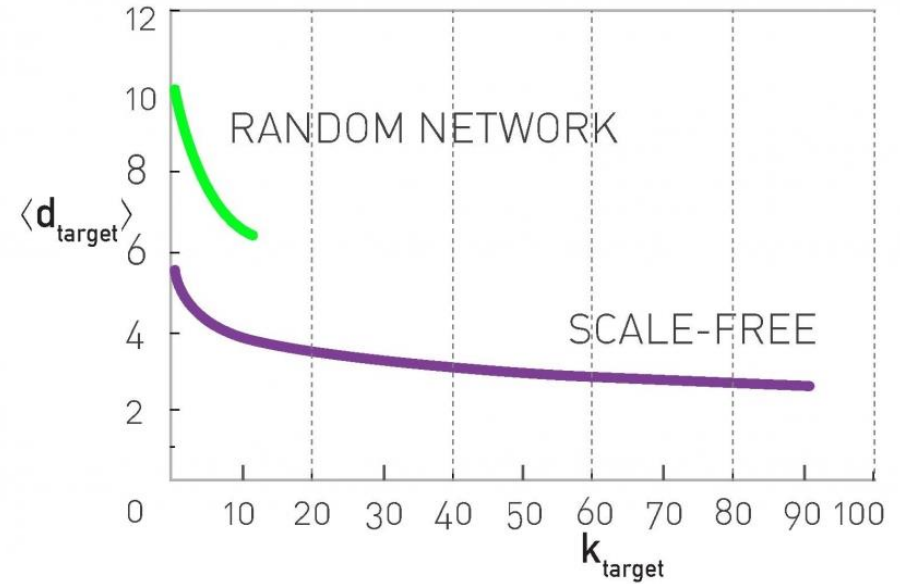


COMPORTAMIENTO DE MUNDO PEQUEÑO EN REDES SIN ESCALA



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

“Siempre es más fácil encontrar a alguien que conozca a una figura famosa o popular que a una persona insignificante.”
(Frigyes Karinthy, 1929)





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

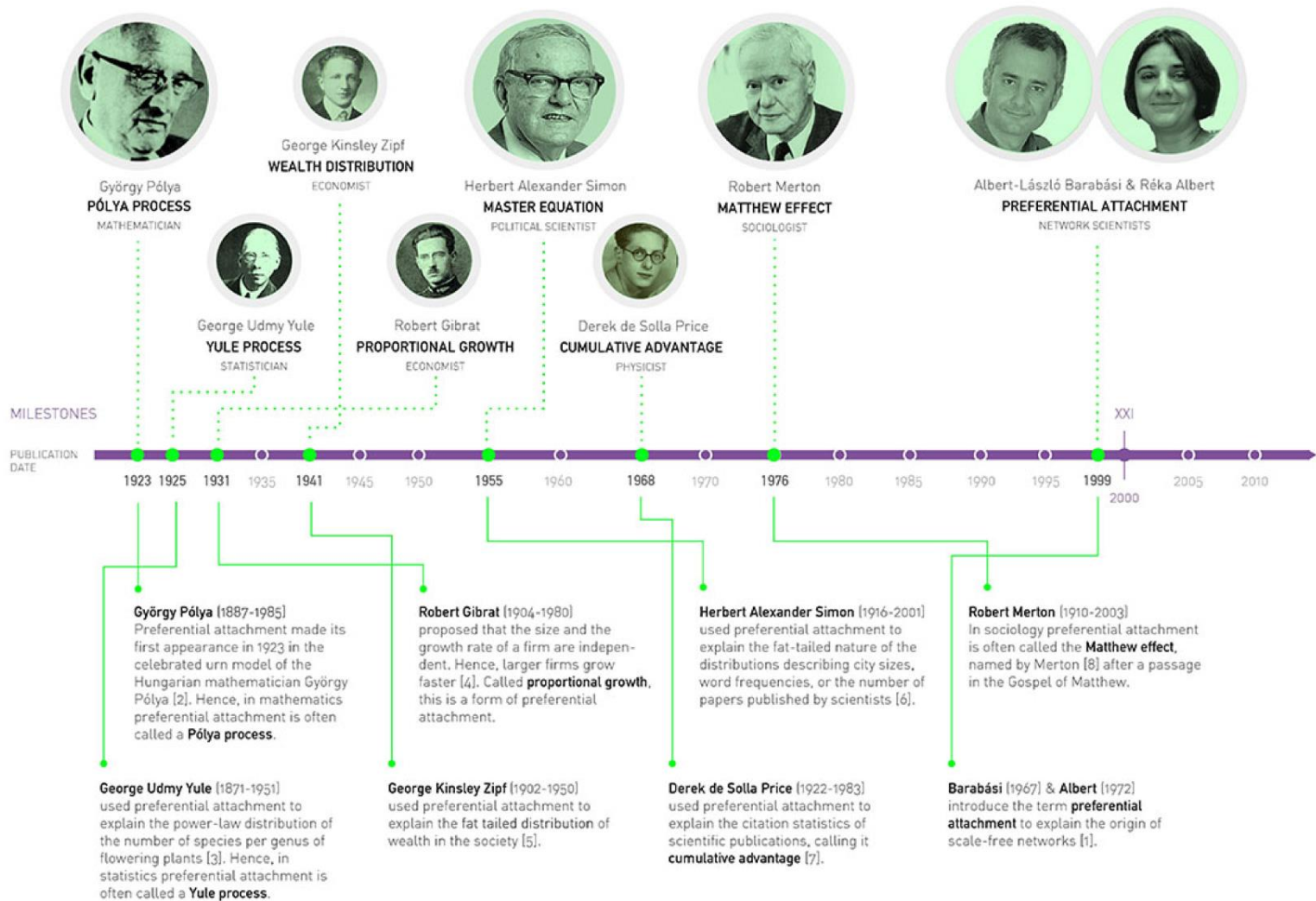


Crecimiento significa que tendremos una red en la que el número de nodos crecerá iterativamente en el tiempo hasta que alcancemos el tamaño de población deseado.



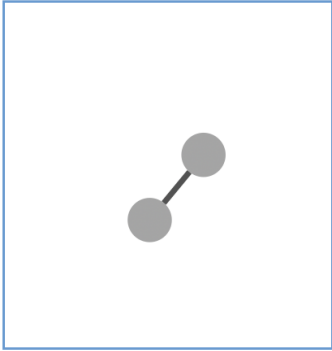
Los nuevos nodos se unirán a los nodos existentes, pero lo harán preferentemente a los más conectados.

Preferential Attachment



Growth and Preferential Attachment

Step 0



El algoritmo consta de los siguientes pasos:

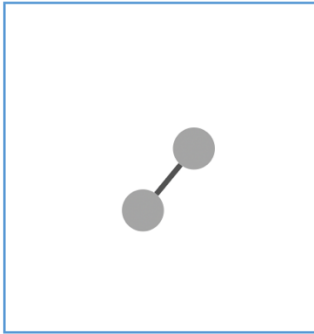
- Comenzando con m_0 nodos, los enlaces se eligen arbitrariamente, siempre que cada nodo tenga al menos un enlace.
- Para las $N - m_0$ iteraciones, agregue un nodo nuevo que se conecte con m ($< m_0$) enlaces a nodos preexistentes con probabilidad:

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

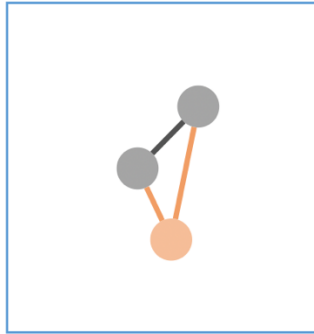
la probabilidad de que un enlace del nuevo nodo se conecte al nodo i depende del grado k_i como

Growth and Preferential Attachment

Step 0



Step 1



El algoritmo consta de los siguientes pasos:

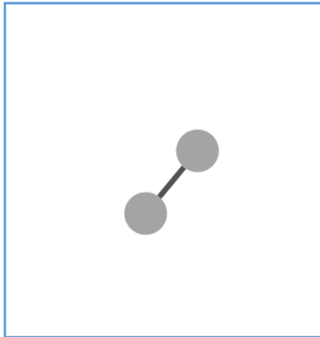
- Comenzando con m_0 nodos, los enlaces se eligen arbitrariamente, siempre que cada nodo tenga al menos un enlace.
- Para las $N - m_0$ iteraciones, agregue un nodo nuevo que se conecte con m ($< m_0$) enlaces a nodos preexistentes con probabilidad:

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

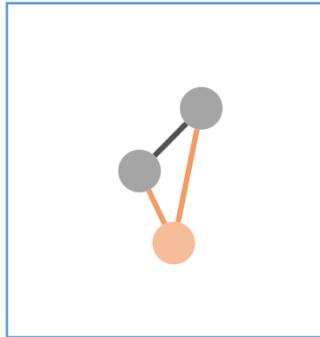
la probabilidad de que un enlace del nuevo nodo se conecte al nodo i depende del grado k_i como

Growth and Preferential Attachment

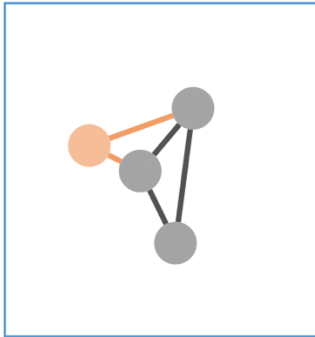
Step 0



Step 1



Step 2



El algoritmo consta de los siguientes pasos:

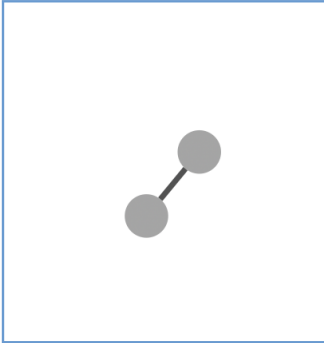
- Comenzando con m_0 nodos, los enlaces se eligen arbitrariamente, siempre que cada nodo tenga al menos un enlace.
- Para las $N - m_0$ iteraciones, agregue un nodo nuevo que se conecte con m ($m < m_0$) enlaces a nodos preexistentes con probabilidad:

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

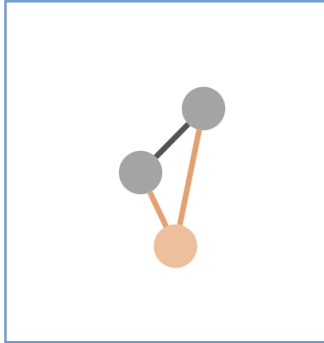
la probabilidad de que un enlace del nuevo nodo se conecte al nodo i depende del grado k_i como

Growth and Preferential Attachment

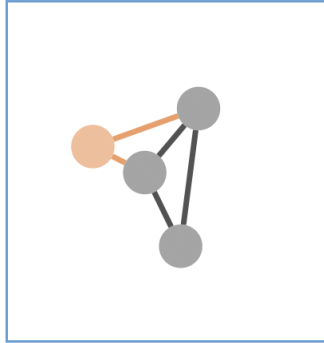
Step 0



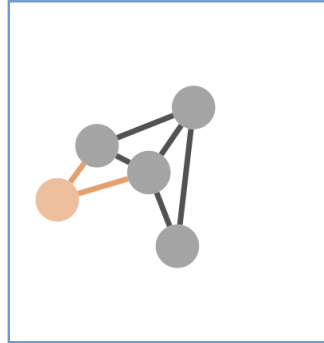
Step 1



Step 2



Step 3



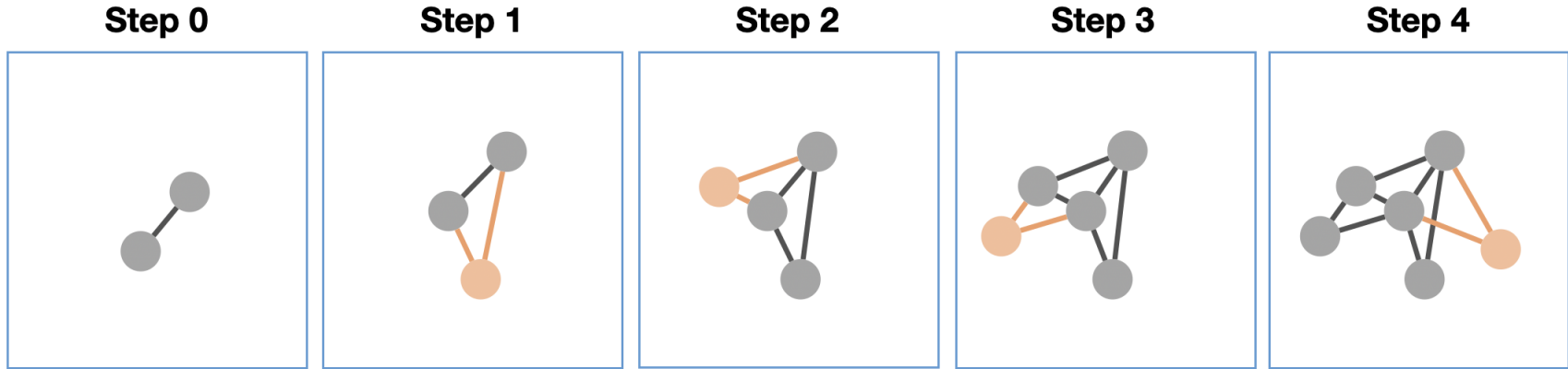
El algoritmo consta de los siguientes pasos:

- Comenzando con m_0 nodos, los enlaces se eligen arbitrariamente, siempre que cada nodo tenga al menos un enlace.
- Para las $N - m_0$ iteraciones, agregue un nodo nuevo que se conecte 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:

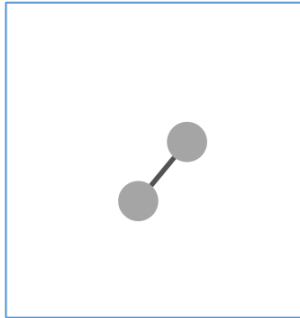
- Comenzando con m_0 nodos, los enlaces se eligen arbitrariamente, siempre que cada nodo tenga al menos un enlace.
- Para las $N - m_0$ iteraciones, agregue un nodo nuevo que se conecte con m ($< m_0$) enlaces a nodos preexistentes con probabilidad:

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

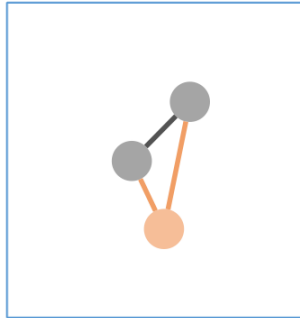
la probabilidad de que un enlace del nuevo nodo se conecte al nodo i depende del grado k_i como

Growth and Preferential Attachment

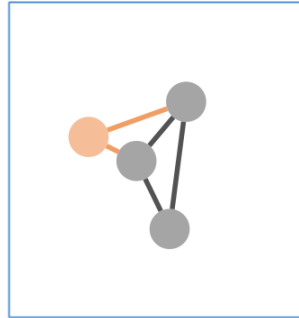
Step 0



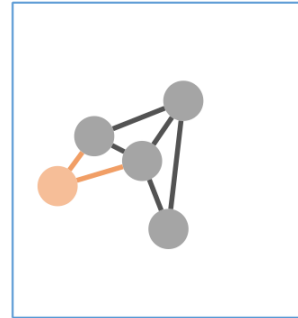
Step 1



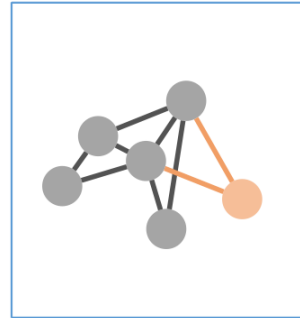
Step 2



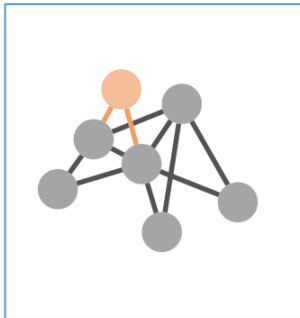
Step 3



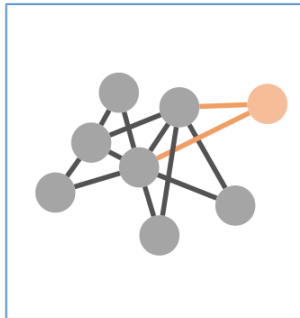
Step 4



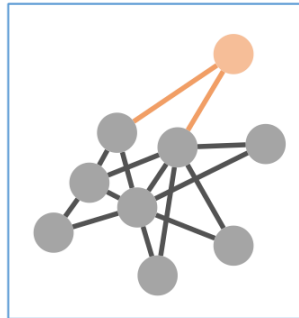
Step 5



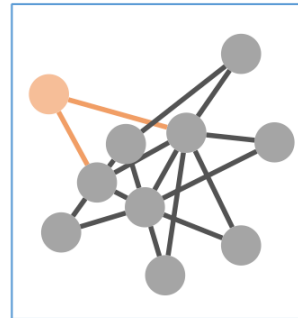
Step 6



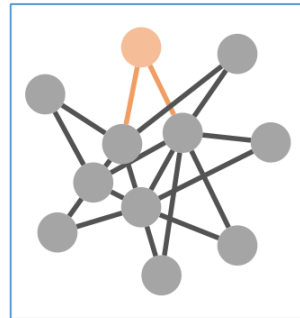
Step 7



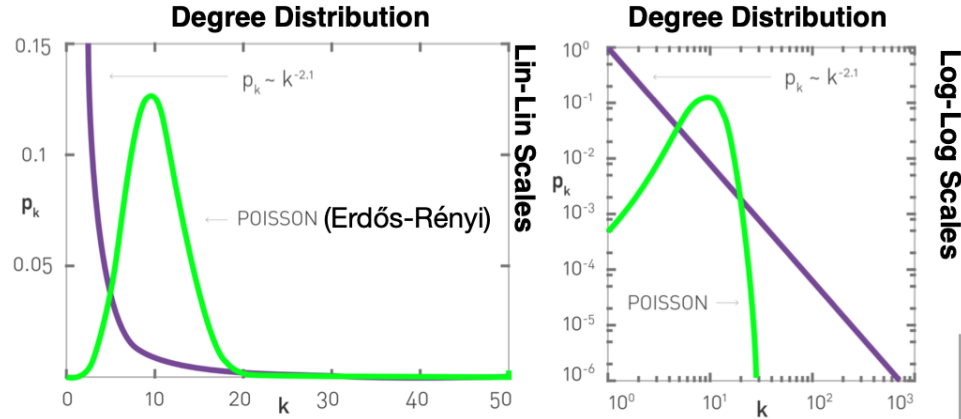
Step 8



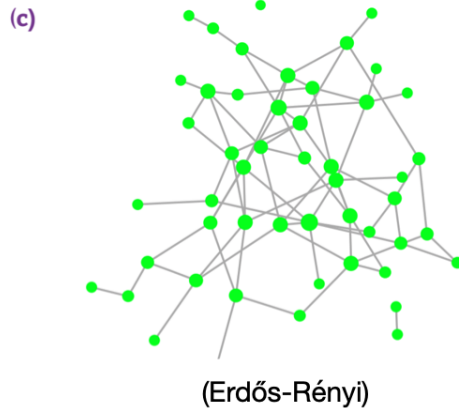
Step 9



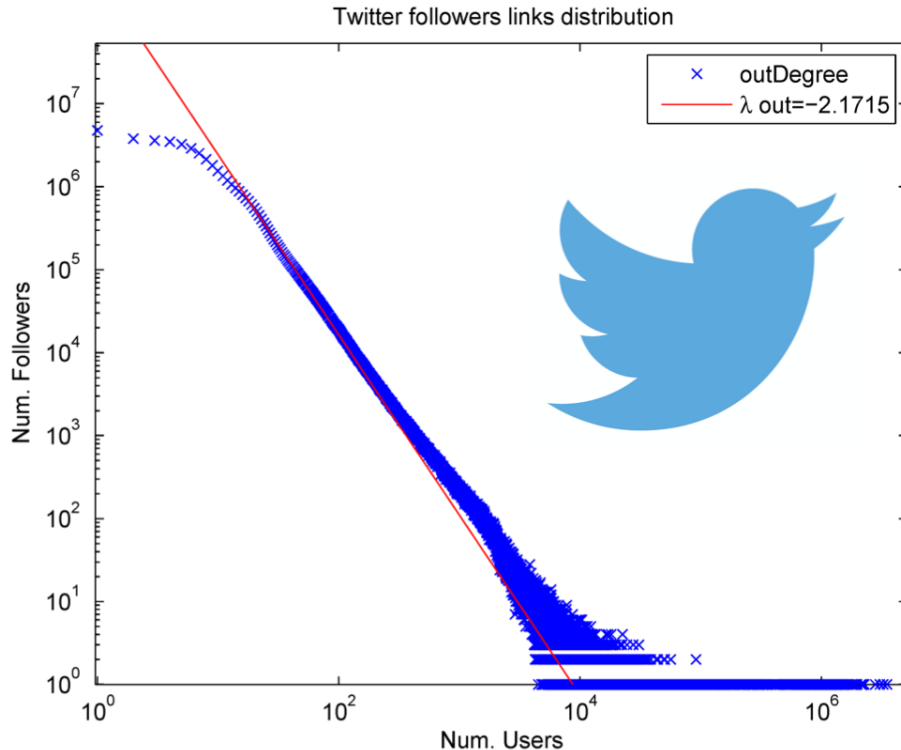
Scale-Free and Random Networks



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

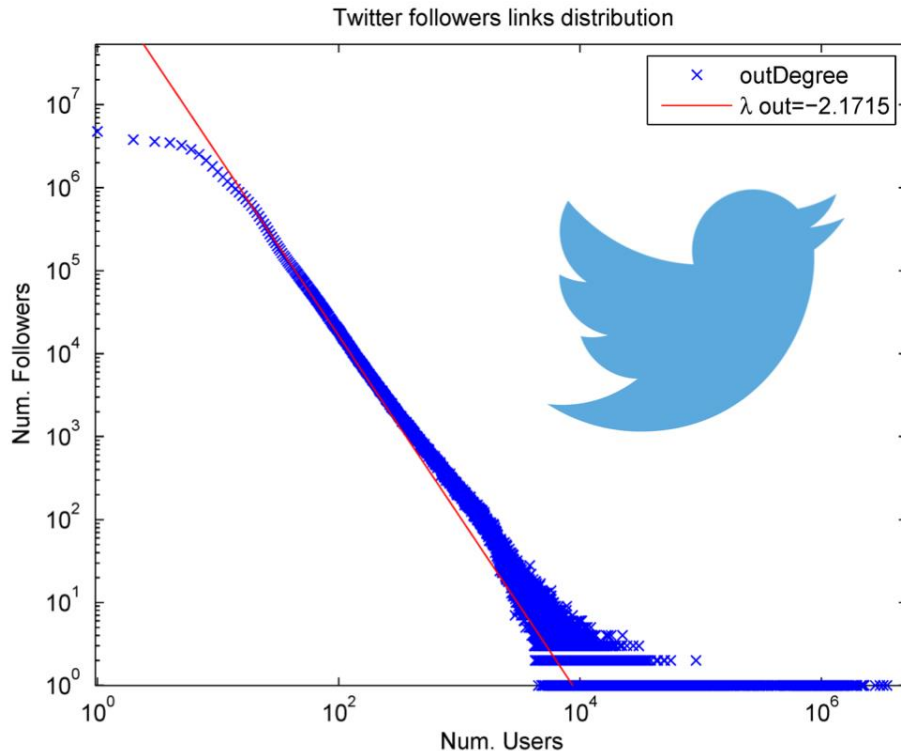


Scale Free Networks



- 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”

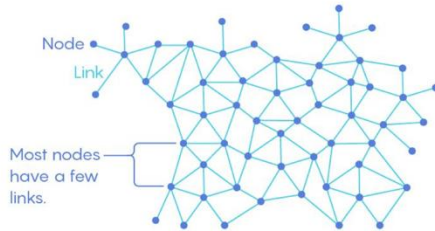
To Be or Not to Be Scale-Free

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

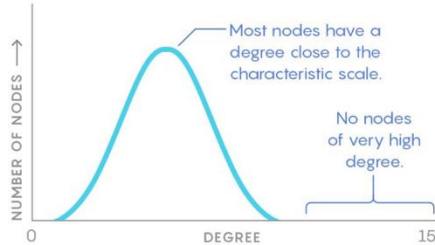
Some experts see so-called scale-free networks everywhere. But a new study suggests greater diversity in real-world networks.

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.

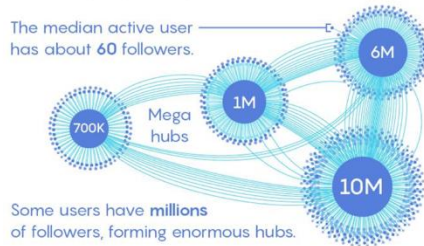


The distribution of degrees is shaped roughly like a bell curve that peaks at the network’s “characteristic scale.”

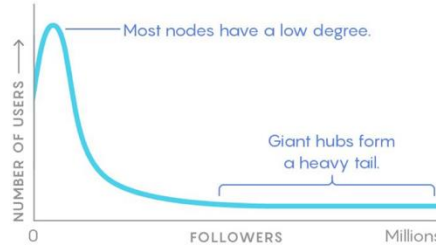


Twitter’s Scale-Free Network

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

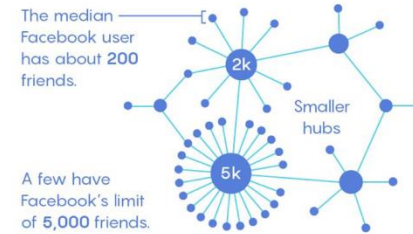


The degrees roughly follow a power law distribution that has a “heavy tail.” The distribution has no characteristic scale, making it scale-free.

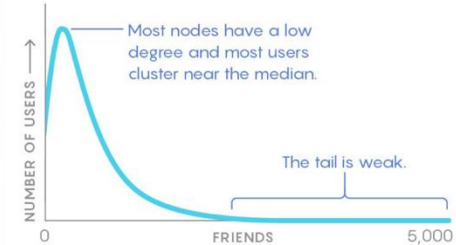


Facebook’s In-Between Network

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.



Barabási–Albert (BA): qué es

BA introduce dos mecanismos:

- **growth**: la red crece agregando nodos;
- **preferential attachment**: los nodos nuevos se conectan con mayor probabilidad a nodos que ya tienen muchos enlaces.

El resultado clásico:

$$P(k) = \frac{2m^2}{k^3}$$

con exponente $\gamma = 3$, asociado al principio “los ricos se hacen más ricos”.

BA: encuadre correcto

Tipo: modelo generativo explicativo de red en evolución.

Qué explica

- hubs
- heterogeneidad fuerte de grado
- colas pesadas
- desigualdad estructural acumulativa

Barabási insiste en que el mensaje central del modelo es que **estructura y evolución son inseparables**. A diferencia de ER, aquí la red no se entiende como un conjunto fijo de nodos al que luego se le “ponen links”.

Por qué no es un modelo nulo

Porque no controla mínimamente: impone una narrativa mecánica fuerte sobre cómo nace la red.

Uso correcto

BA sirve como baseline si la pregunta es:

¿mis hubs pueden explicarse por crecimiento + preferential attachment?

Preguntas básicas

- Explica cómo se alcanza la propiedad de mundo pequeño en el modelo Watts-Strogatz.
- Qué es lo que explica el modelo Barabási-Albert en ciencia de redes?
- Compara y contrasta las distribuciones de grado de los tres modelos. De qué manera difieren y como se relacionan con la estructura de la red?

ER vs WS vs BA

Modelo	Tipo	Qué preserva / impone	Qué patrón explica mejor
ER	nulo aleatorio básico	(n), densidad esperada	camino corto bajo azar homogéneo
WS	generativo estilizado	localidad + shortcuts	clustering alto + distancias bajas
BA	generativo evolutivo	crecimiento + apego preferencial	hubs y cola pesada

No hay un “mejor modelo” universal.
Hay modelos adecuados para distintas preguntas.

Qué comparar en la práctica

Cuando contrastamos **red real vs baseline**, no basta con una sola métrica. La comparación mínima debería incluir:

- tamaño n , número de enlaces m , densidad
- grado medio y distribución de grados
- clustering global y local
- distancia media y diámetro
- componente gigante
- opcional: assortativity, motifs, modularidad, rich-club

Newman discute explícitamente que la interpretación de clustering y motifs depende de comparar contra valores esperados bajo conexiones aleatorias; valores mayores, menores o similares pueden tener significados distintos.

Criterio de interpretación: qué significa “señal no explicable”

Diremos que una propiedad de la red real contiene **señal no explicable** por un baseline si:

1. el baseline preserva controles relevantes
2. la propiedad observada cae fuera del rango típico simulado
3. la diferencia es sistemática, no una realización aislada

Ejemplos

- clustering real muy superior a ER → densidad sola no basta
- cola pesada ausente en ER pero presente en datos → azar homogéneo no basta
- hubs similares a BA pero clustering muy distinto → preferential attachment solo no basta
- clustering proyectado alto en una red bipartita → puede ser artefacto de proyección, no señal sustantiva.

Advertencia central: no confundir ajuste parcial con explicación

Un modelo puede parecerse a la red real en una propiedad y fallar en otras.

Ejemplos

- ER puede aproximar distancias pero fallar totalmente en clustering.
- BA puede aproximar la cola de grados pero no el clustering o la modularidad.
- WS puede dar small-world pero no hubs.

Entonces

Un buen contraste requiere declarar **qué propiedad está siendo testeada** y **qué controles incorpora el baseline**.

Más allá de ER: nulos útiles para redes tradicionales

1. Configuration model

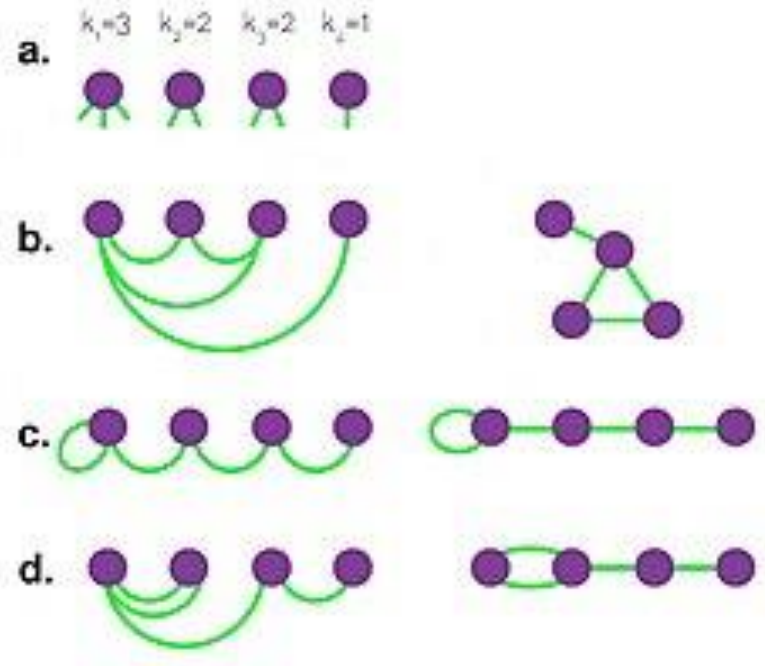
Preserva la **secuencia de grados** y aleatoriza el resto.

Es mucho más útil que ER cuando el grado es altamente heterogéneo.

Newman dedica un capítulo completo a random graphs with general degree distributions and al configuration model como referencia para redes con distribución de grado no Poisson.

Pregunta que responde

¿lo observado excede lo esperable dado el grado de cada nodo?



Más allá de ER: nulos útiles para redes tradicionales

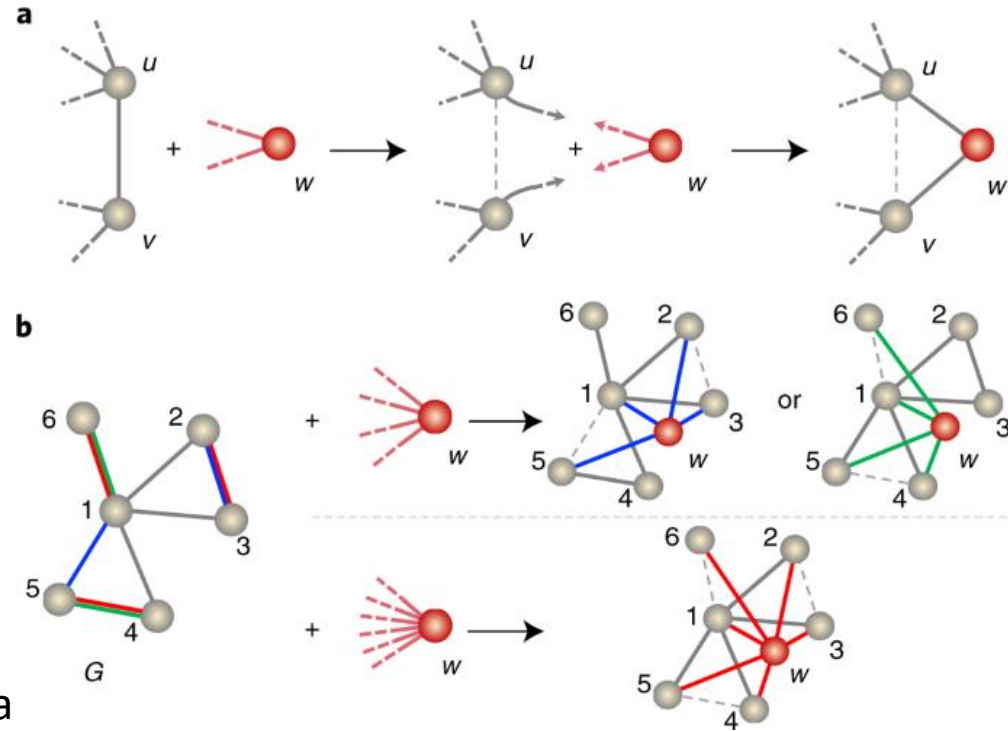
2. Degree-preserving rewiring

Rewire aleatorio que conserva el grado de cada nodo. Útil para:

- motifs
- clustering residual
- assortativity
- rich-club
- modularidad residual

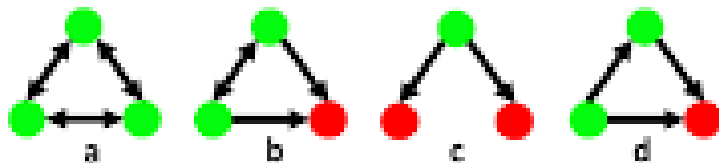
Ventaja

Mantiene la heterogeneidad nodal observada, evitando atribuir a “estructura compleja” algo que en realidad viene solo del degree sequence.

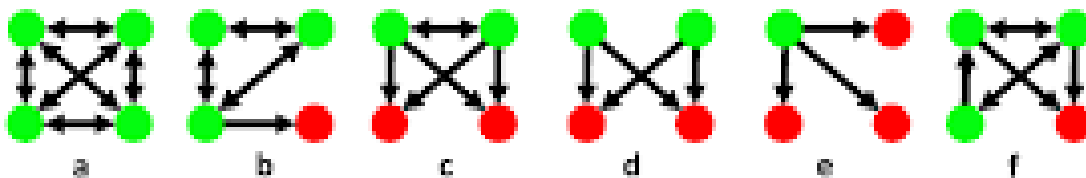


Motifs?

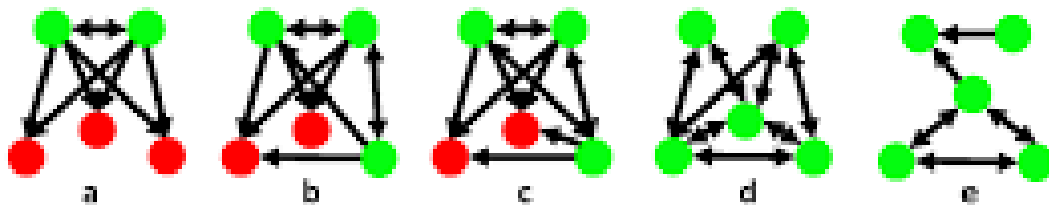
3-node motifs



4-node motifs



5-node motifs



Más allá de ER: nulos útiles para redes tradicionales

3. Directed nulls

Para redes dirigidas, el control correcto suele preservar:

- in-degree
- out-degree
- a veces reciprocidad

4. Weighted nulls

En redes ponderadas, puede ser necesario preservar:

- topología binaria
- distribución de pesos
- o fuerza nodal

5. Stochastic Block Model (SBM)

No es un nulo básico, pero sí un baseline estadístico útil cuando la hipótesis relevante es de **equivalencia estocástica** o de bloques. Los nodos son equivalentes si sus probabilidades de conexión dependen solo de su clase.

Pregunta que responde

¿la organización relevante es comunitaria o de bloques, más que individual?

6. Nulos espaciales o atributivos

Si la red depende fuertemente de geografía, tamaño institucional o atributos, un nulo serio debería preservar esos condicionantes.

Redes bipartitas: por qué necesitan nulos propios

Las redes bipartitas $G = (V, U, E)$ tienen dos tipos de nodos y enlaces solo entre modos distintos. Latapy et al. enfatizan que analizar estas redes mediante proyecciones puede perder mucha información y generar artefactos.

Problema central

La proyección 1-mode puede inflar enlaces y clustering de forma mecánica.

Latapy et al. muestran explícitamente que el clustering alto en proyecciones puede ser simplemente consecuencia de la estructura bipartita subyacente, no una propiedad sustantiva del sistema.

Nulos útiles para redes bipartitas

1. Random bipartite graph con mismo tamaño

Preserva:

- número de nodos top y bottom
- número total de enlaces o densidad bipartita
- Útil como control básico, análogo a ER

2. Bipartite degree-preserving null

Preserva:

- secuencia de grados del modo top
- secuencia de grados del modo bottom

Este es normalmente el control más útil para afiliación, coautoría, empresa-licitación, acto-película, etc.

Latapy et al. usan explícitamente random bipartite graphs con el mismo tamaño y las mismas degree distributions para comparar clustering, redundancia y otras estadísticas

Qué comparar en bipartitas

En vez de proyectar de inmediato, comparar sobre la bipartita original:

- tamaño de ambos modos
- densidad bipartita
- distribuciones de grado top y bottom
- distancias
- overlap de vecindarios
- clustering bipartito
- redundancy

Latapy et al. enfatizan que su contribución central es justamente un conjunto de estadísticas rigurosas para análisis de redes 2-mode sin perder la naturaleza bipartita. También muestran que la redundancy puede ser más informativa que extensiones inmediatas del clustering clásico.

Advertencia fuerte sobre proyecciones bipartitas

No interpretar alto clustering en una proyección como señal fuerte sin baseline bipartito.

Razón:

cada evento o grupo induce cliques en la proyección. Newman lo explica claramente: al proyectar una bipartita, cada grupo genera una clique en la proyección 1-mode.

Latapy et al. refuerzan esto con una conclusión metodológica contundente: el expected clustering en las proyecciones puede ser grande solo por la estructura bipartita, por lo que no debe interpretarse automáticamente como propiedad significativa

Model checking

- Definir la red real y sus decisiones de modelamiento.
- Declarar explícitamente el baseline.
- Decir qué preserva y qué no preserva.
- Simular varias realizaciones.
- Comparar métricas clave y distribuciones.
- Documentar la señal residual.
- No sobrerreclamar.

Entonces debemos: Comparar contra **baselines estructurales bien establecidos bajo las mismas condiciones** y documentar explícitamente el procedimiento.

Regla práctica para elegir nulo o baseline

Usa ER si

quieres un primer control de tamaño y densidad.

Usa configuration / degree-preserving si

el grado importa y sospechas heterogeneidad estructural.

Usa WS si

tu hipótesis relevante es small-world: alta cohesión local con distancias cortas.

Usa BA si

tu hipótesis relevante es crecimiento con hubs por adjunción preferencial.

Usa nulos bipartitos si

tu red es 2-mode o viene de afiliaciones, eventos, compras, coautoría, etc.

Regla práctica para elegir nulo o baseline

Los modelos generativos y nulos no son un adorno teórico: son una defensa metodológica contra interpretaciones espurias.

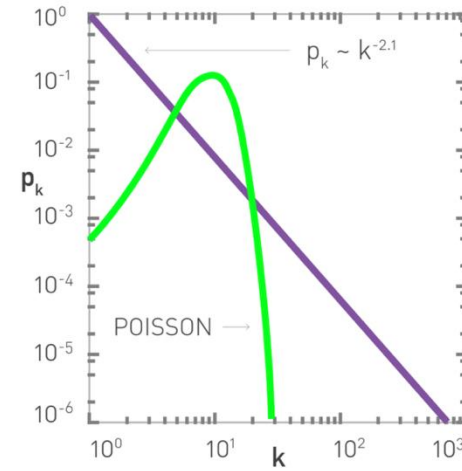
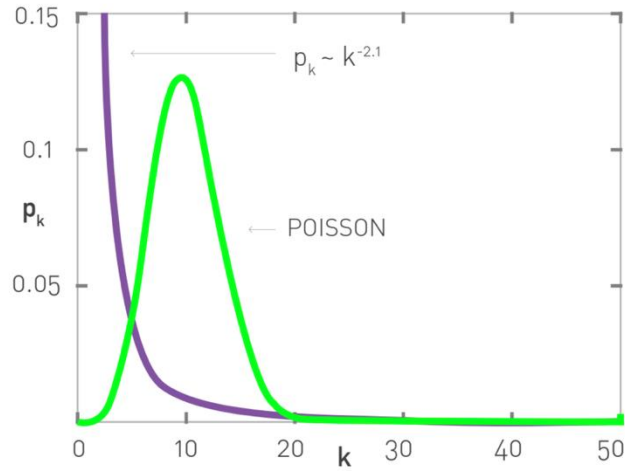
Finalmente

Antes de interpretar una red real, pregunten siempre:

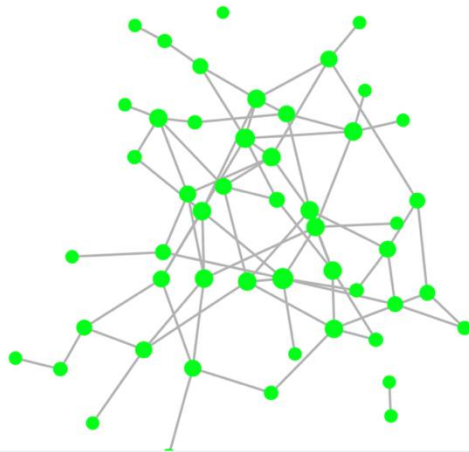
- ¿comparado con qué?
- ¿qué controla ese baseline?
- ¿qué rasgo sigue sin explicación después del control?
- Esa pregunta separa la descripción ingenua del análisis serio.

Apéndice

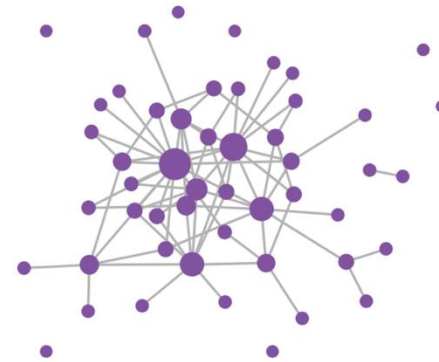
Diferencia entre red aleatoria y libre de escala



(c)

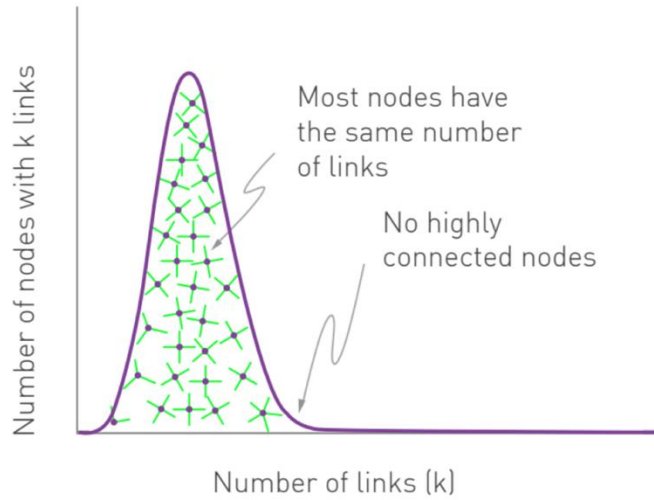


(d)

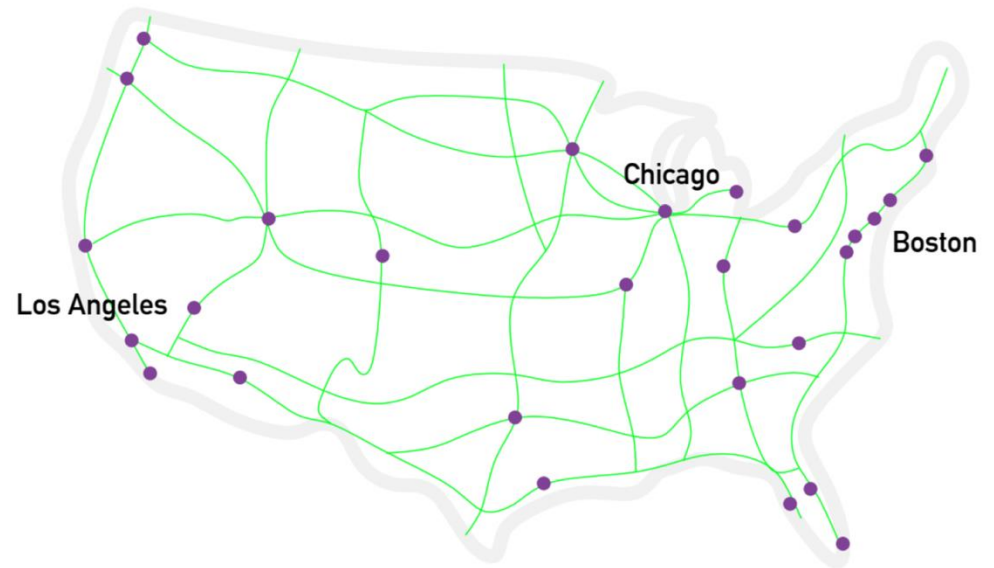


Diferencia entre red aleatoria y libre de escala

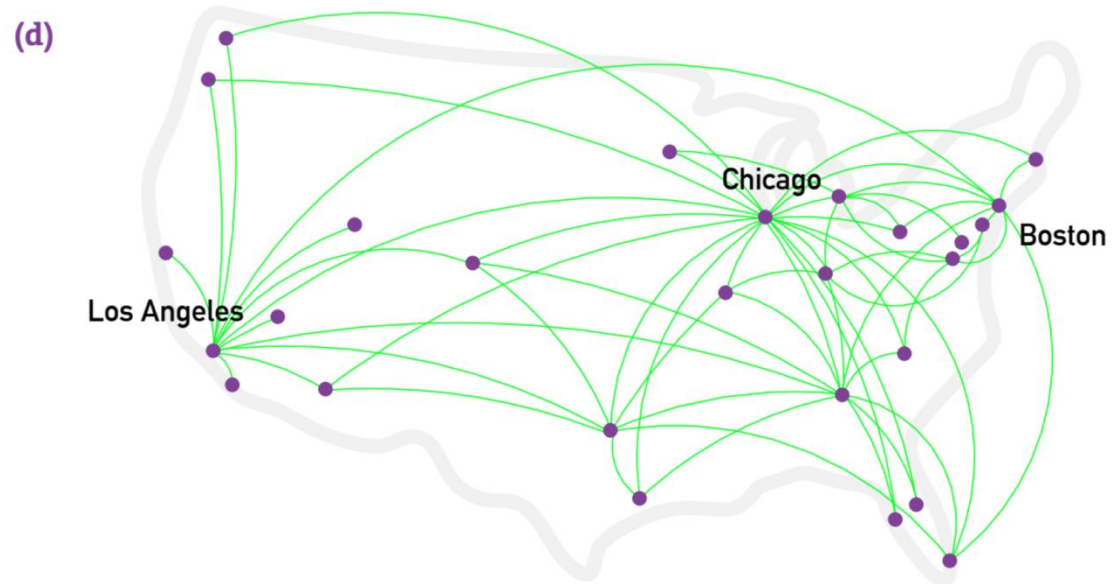
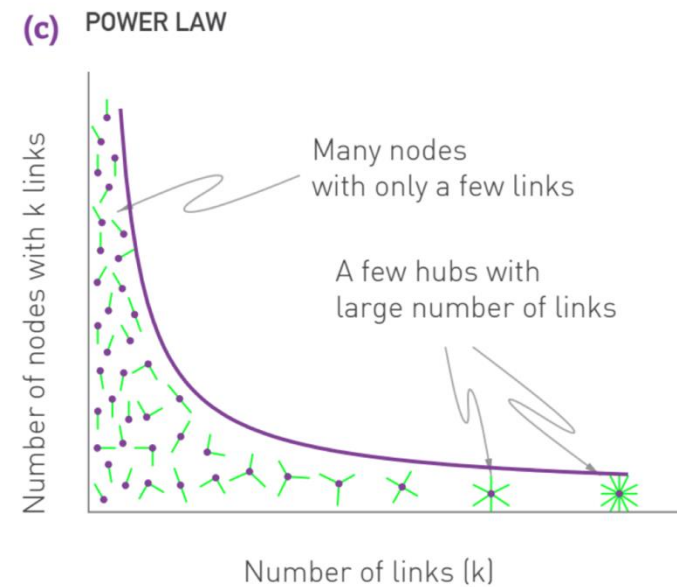
(a) POISSON



(b)



Diferencia entre red aleatoria y libre de escala



Estas estructuras tienen un impacto significativo en cómo se propaga la información en los sistemas físicos, biológicos, sociales, etc.

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

Dist. Binomial de "x"

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

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

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

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

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 :

$$P(L) = \underbrace{\binom{\binom{N}{2}}{L}}_{\text{Número de formas diferentes en que podemos elegir los enlaces } L \text{ entre todos los enlaces potenciales.}} p^L (1-p)^{\frac{N(N-1)}{2} - L}$$

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...

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

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

MODELO DE REDES ALEATORIAS

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

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

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

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

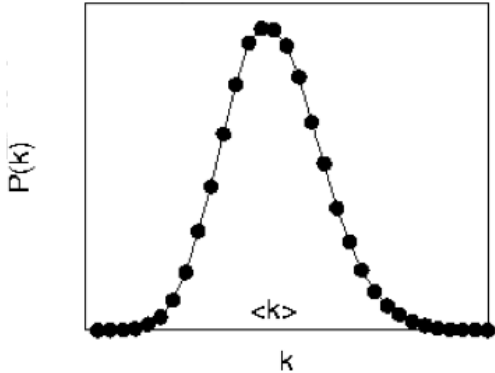
$$\langle x \rangle = \sum_{x=0}^N x p_x = Np$$

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

•La varianza

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

DISTRIBUCION DE GRADO DE UNA RED ALEATORIA



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

Selecciona k nodos de un conjunto de N-1 nodos

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

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

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

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

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

$$(1-p)^{(N-1)-k} = e^{-\langle k \rangle}$$

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

DISTRIBUCIÓN DE GRADO DE POISSON

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

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

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

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:

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

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

INTERPRETATION:

$$p_k$$

Formalismo continuo

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

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

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

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

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

El significado de la propiedad
libre de escala

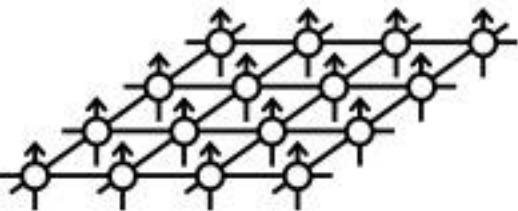
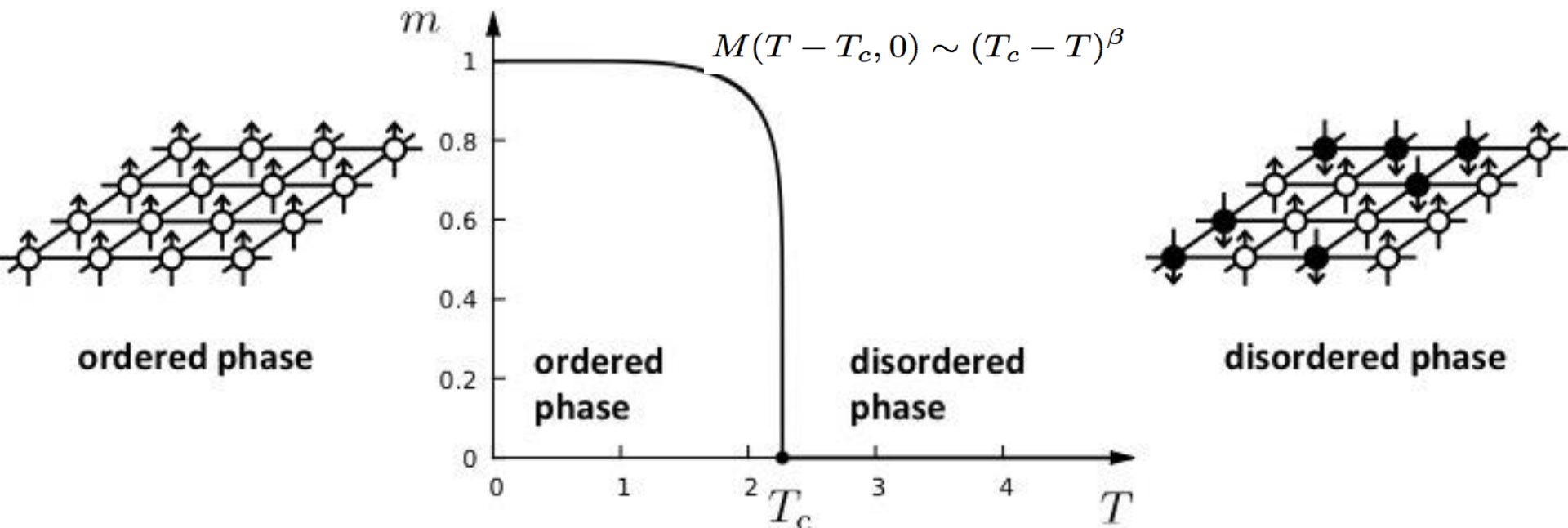
Definición :

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

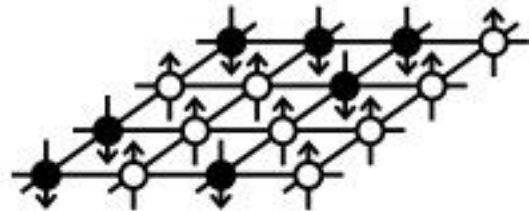
¿De dónde viene el nombre?

Fenómenos críticos e invariancia de escala.
(un desvío)

Transiciones de fase en sistemas complejos I: Magnetismo

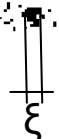


ordered phase

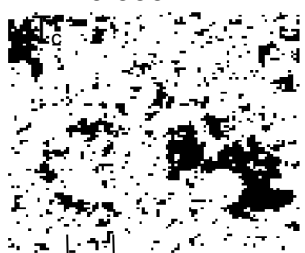


disordered phase

$T = 0.99 T_c$



$T = 0.999$



$T = T_c$



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

$T = 1.5 T_c$



$T = 2 T_c$



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



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

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.

$$P(k) = Ck^{-\gamma} \quad k = [k_{\min}, \infty) \quad \int_{k_{\min}}^{\infty} P(k) dk = 1 \quad C = \frac{1}{\int_{k_{\min}}^{\infty} k^{-\gamma} dk} = (\gamma - 1)k_{\min}^{\gamma-1}$$

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

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

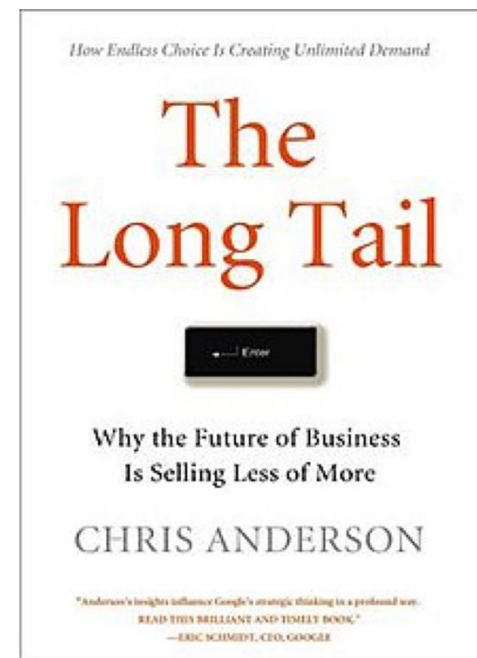
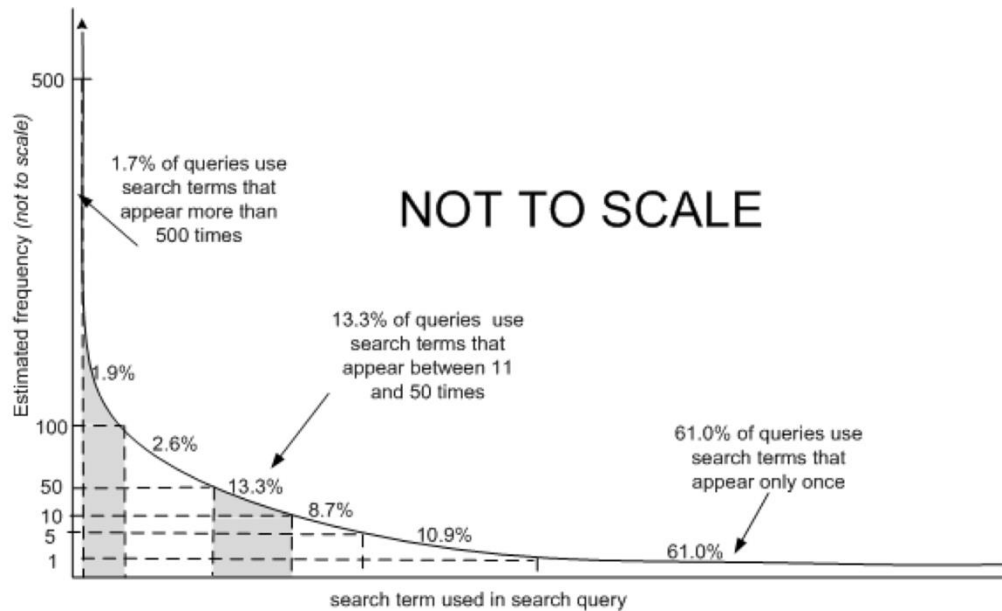
Exponente debe ser negativo

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



¿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.

El tamaño del hub más grande

Todas las redes reales son finitas \rightarrow exploremos sus consecuencias.

\rightarrow 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

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

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

El tamaño del hub más grande.

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

Red libre de escala:

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

Red libre aleatoria (hubs prohibidos):

$$k_{\max} = 13$$

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
à 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!