

Network Science

dataScience UDD



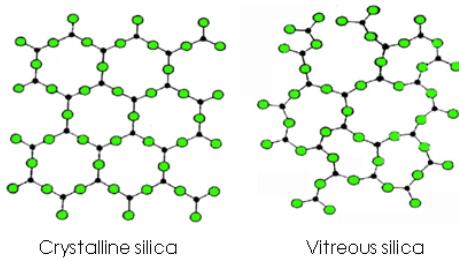
Cristian Candia-Castro Vallejos, Ph.D.

cristiancandia@udd.cl

Director Magister en Data Science UDD
Profesor Investigador, Facultad de Ingeniería, UDD
External Faculty Northwestern Institute on Complex Systems, Kellogg School of
Management, Northwestern University

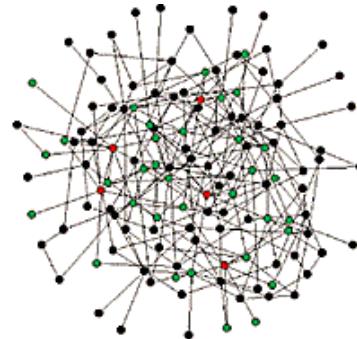
Introducción

Desde sistemas físicos a sociales



Crystalline silica

Vitreous silica

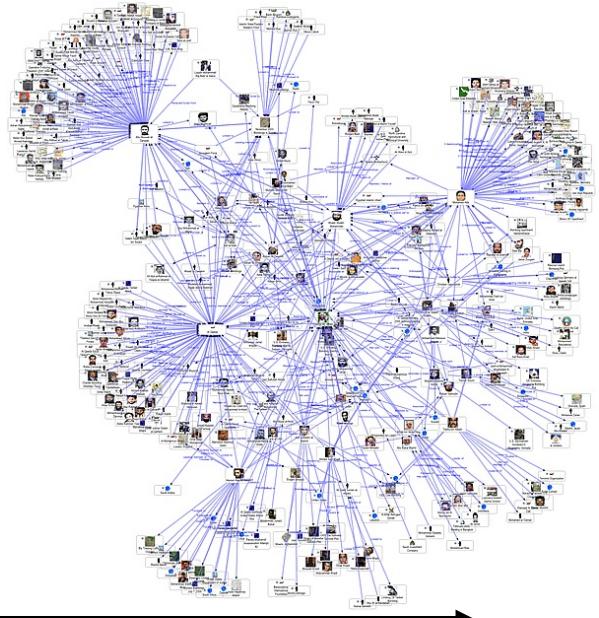


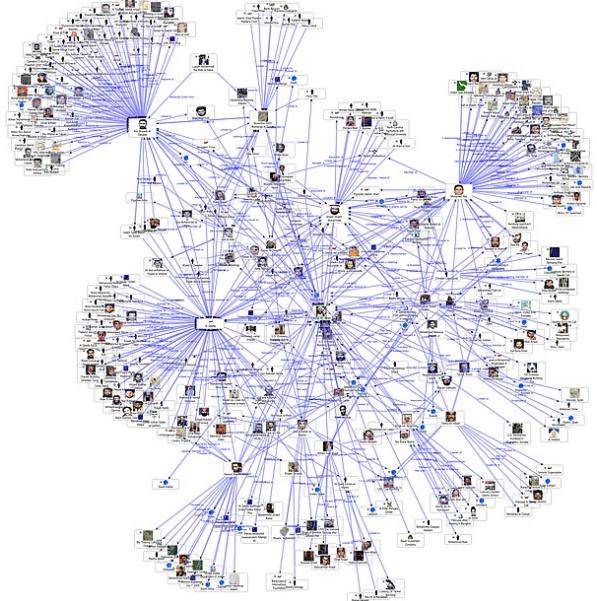
increasing network heterogeneity/complexity

regular, lattices
physical systems

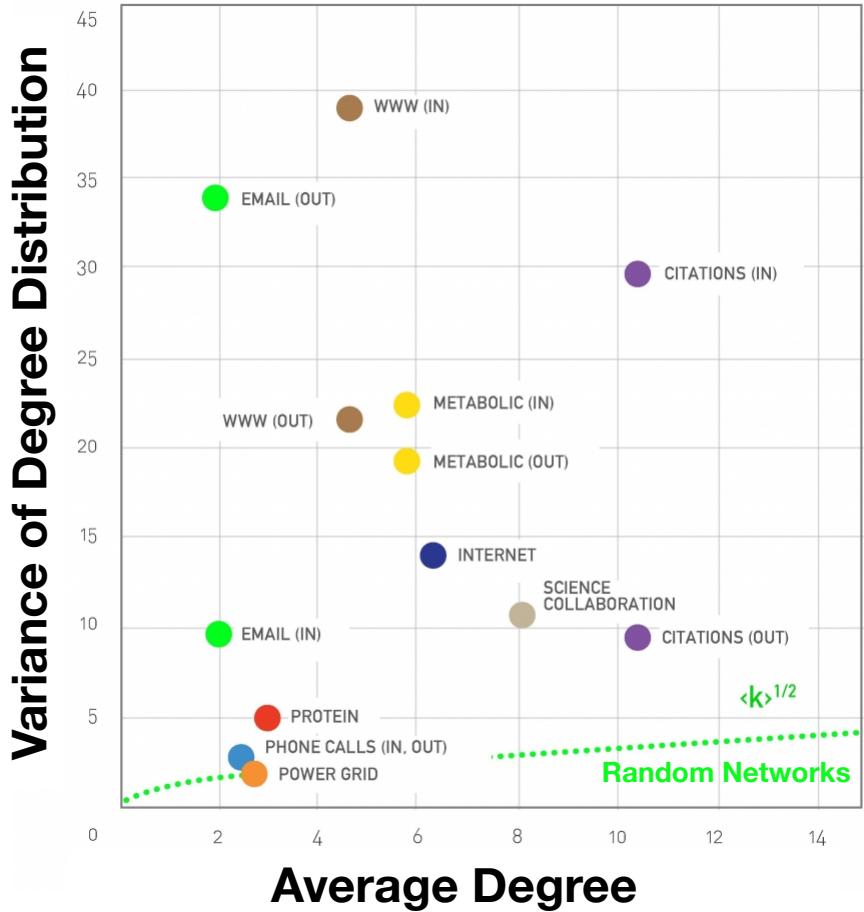
random structures
mathematics

complex structures
social systems





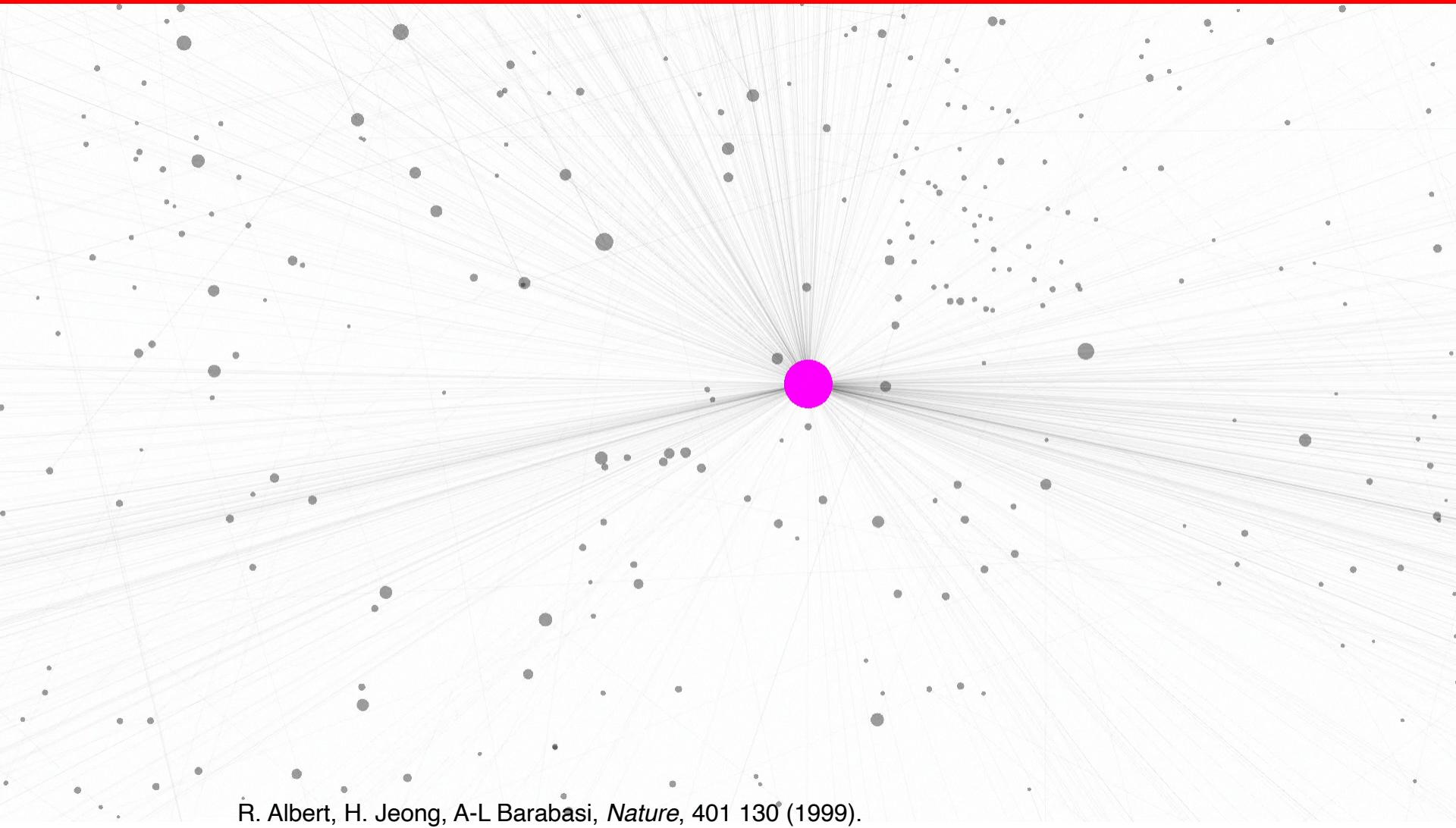
Las redes reales no son random son libre de escala!



Leyes de potencia y redes libre de escala

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

WORLD WIDE WEB



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

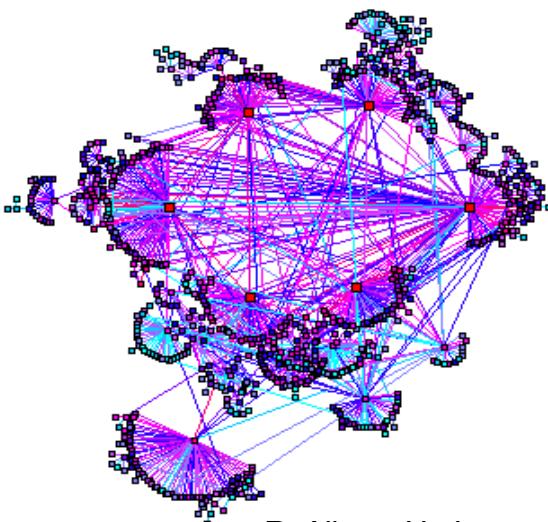
WORLD WIDE WEB

Nodos: **documentos WWW**

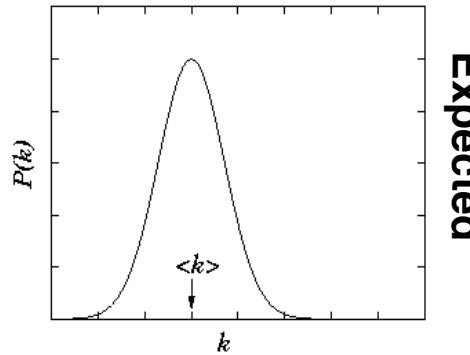
Enlaces: **links URL**

Más de 3 mil millones de documentos.

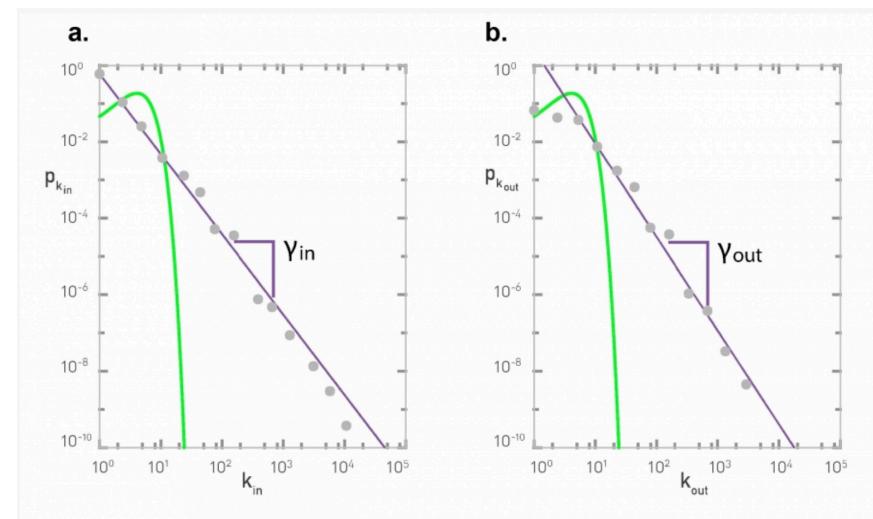
ROBOT: recopila todas las URL encontradas en un documento y las sigue de manera recursiva



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



Expected



Formalismo discreto vs. continuo.

Formalismo discreto

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

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

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

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

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

INTERPRETATION:

$$p_k$$

Formalismo continuo

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

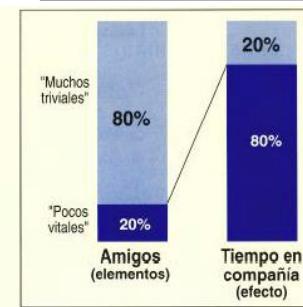
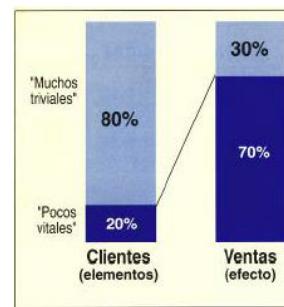
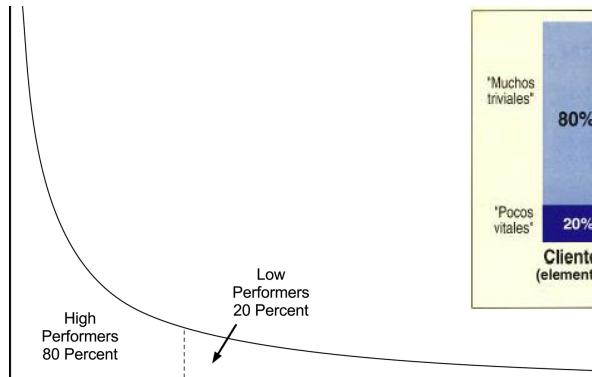
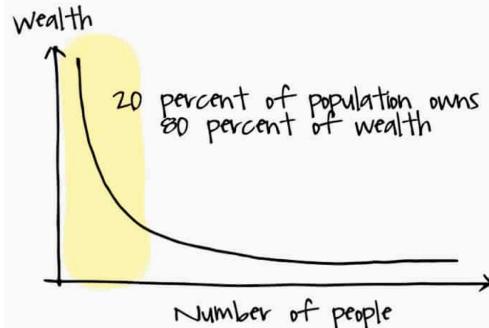
$$p(k) = Ck^{-\gamma}.$$

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

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

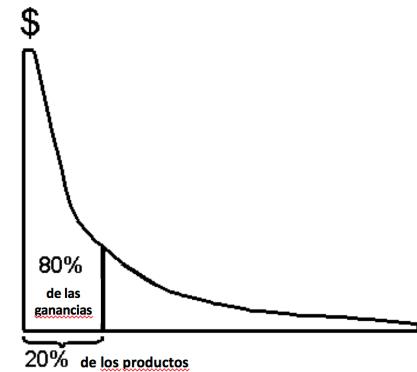
$$p(k) = (\gamma - 1)k_{\min}^{\gamma-1}k^{-\gamma}.$$

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



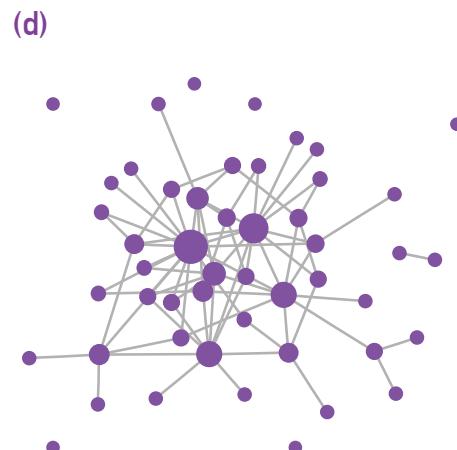
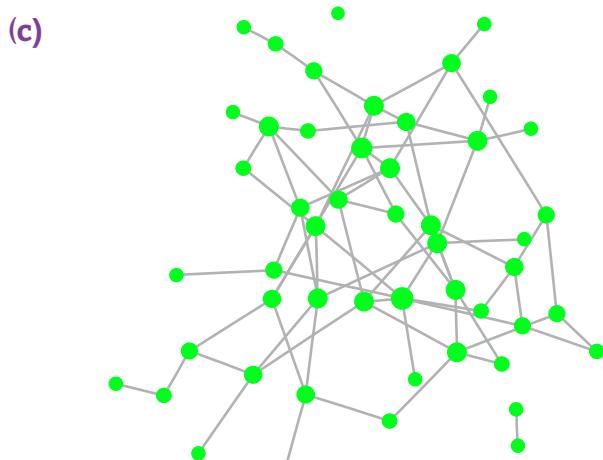
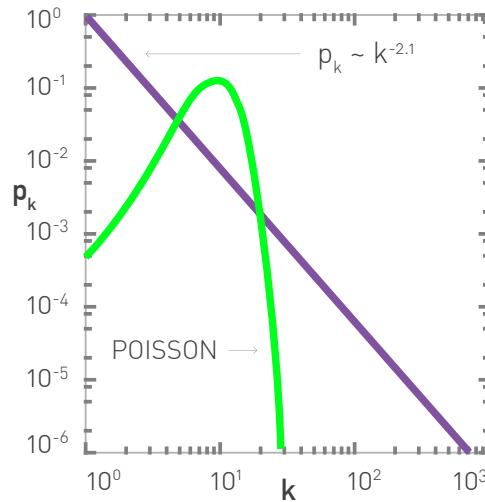
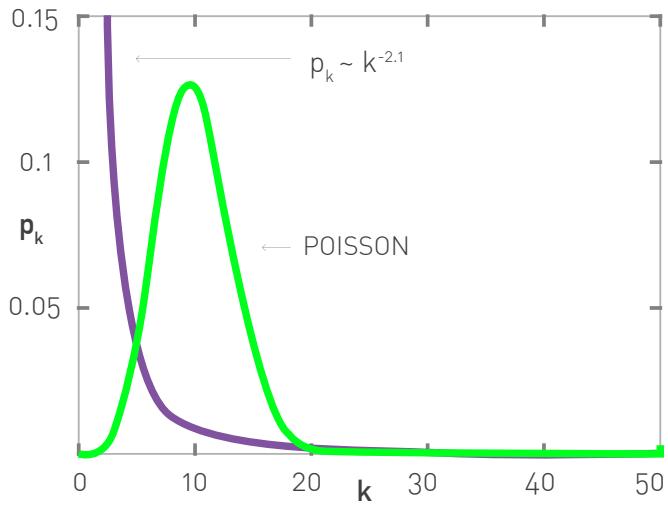
REGLA 80/20

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



Hubs

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

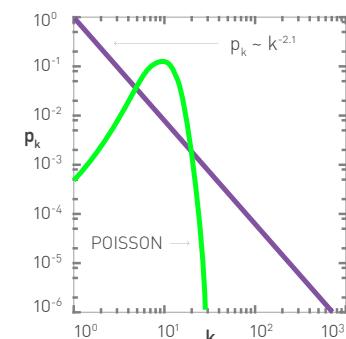


La diferencia entre una ley de potencia y una distribución exponencial.

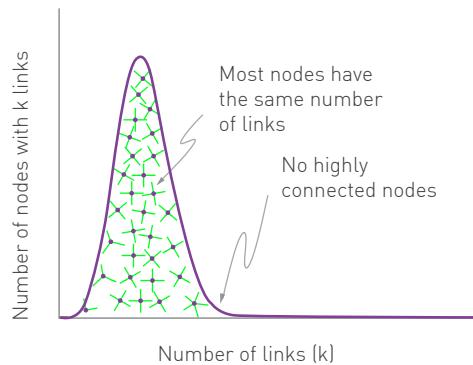
Usemos la WWW para ilustrar las propiedades del régimen de alto-k.

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

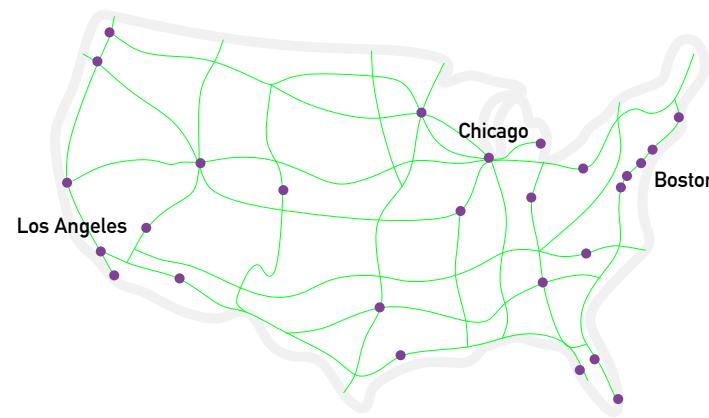
- Cercana a $p_{100} \simeq 10^{-30}$ en una distribución de Poisson
- Cercana a $p_{100} \simeq 10^{-4}$ si p_k sigue una ley de potencia
- En consecuencia, si la WWW fuera una red aleatoria, de acuerdo con la predicción de Poisson, esperaríamos 10^{-18} nodos de grado $k > 100$, o ninguno.
- Para una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



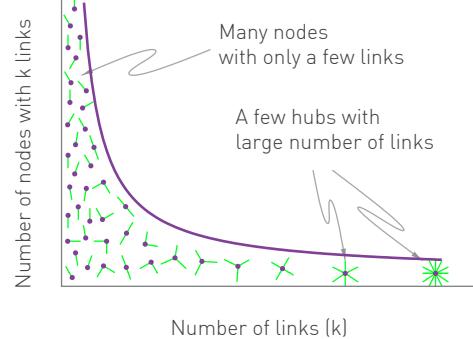
(a) POISSON



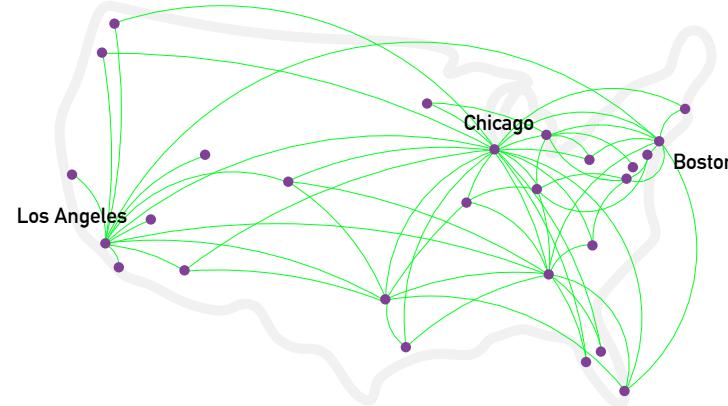
(b)



(c) POWER LAW



(d)



El tamaño del hub más grande.

Todas las redes reales son finitas → exploremos sus consecuencias.

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

Estimando k_{\max}

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

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

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

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

El tamaño del hub más grande.

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

Red libre de escala:

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

Red libre aleatoria (hubs prohibidos):

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

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

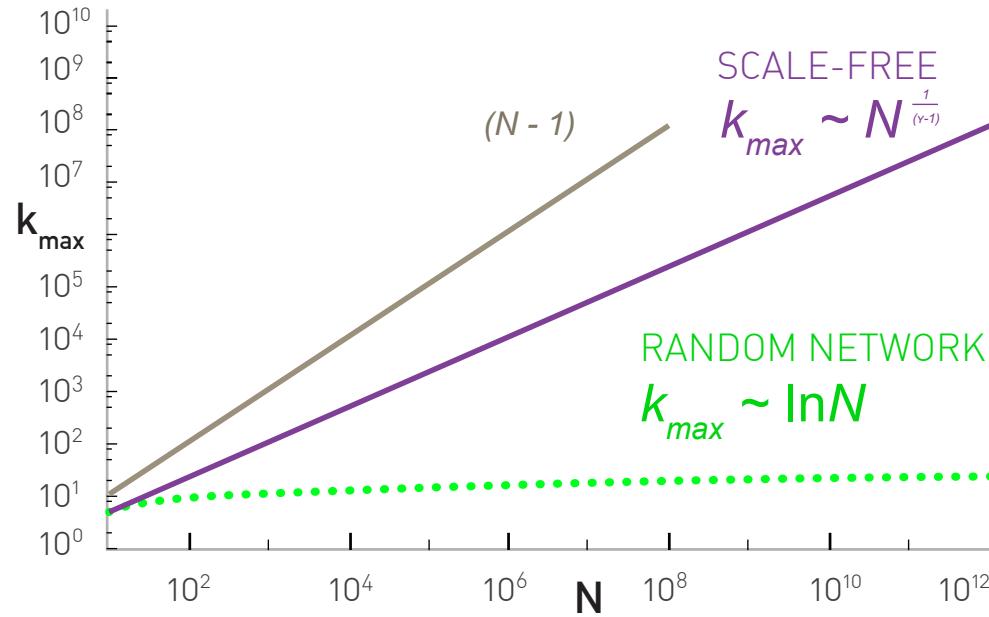
Redes libre de escala finitas

Grado máximo esperado, k_{\max}

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

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

El tamaño del hub más grande.



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

El significado de la propiedad libre de escala

Redes libre de escala: Definición

Definición :

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

¿De dónde viene el nombre?

Fenómenos críticos e invariancia de escala.
(un desvío)

Comportamiento libre de escalas en el espacio.

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

a $T = T_c$:

la longitud de
correlación diverge

Las fluctuaciones emergen
en todas las escalas:

*comportamiento libre de
escala*



Scale invariance at the critical point

by Douglas Ashton

www.kineticallyconstrained.com

FENOMENOS CRITICOS

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

Divergencias en distribuciones sin escala.

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

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

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

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

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

Exponente debe ser negativo

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

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

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

La integral diverge.

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

DIVERGENCIA DE LOS MOMENTOS SUPERIORES

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

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

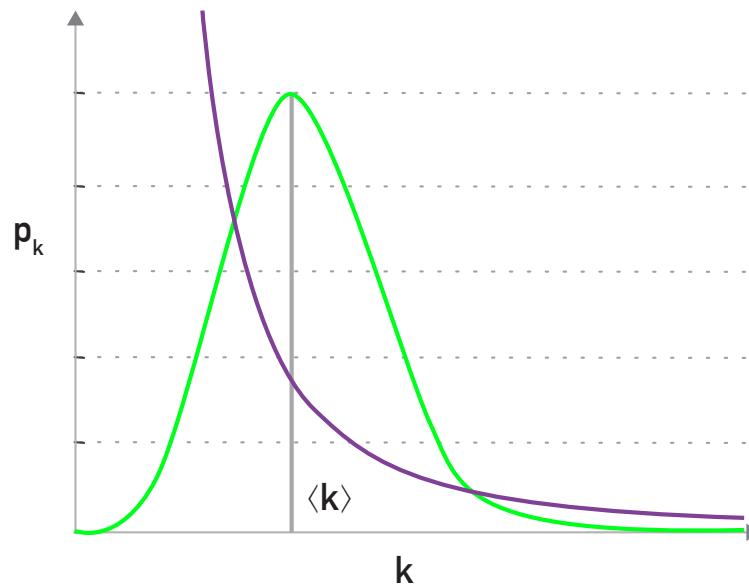
Network	Size	$\langle k \rangle$	κ	γ_{out}	γ_{in}
WWW	325 729	4.51	900	2.45	2.1
WWW	4×10^7	7		2.38	2.1
WWW	2×10^8	7.5	4000	2.72	2.1
WWW, site	260 000				1.94
Internet, domain*	3015–4389	3.42–3.76	30–40	2.1–2.2	2.1–2.2
Internet, router*	3888	2.57	30	2.48	2.48
Internet, router*	150 000	2.66	60	2.4	2.4
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 $N \rightarrow \infty$

El significado de sin escala



Random Network

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

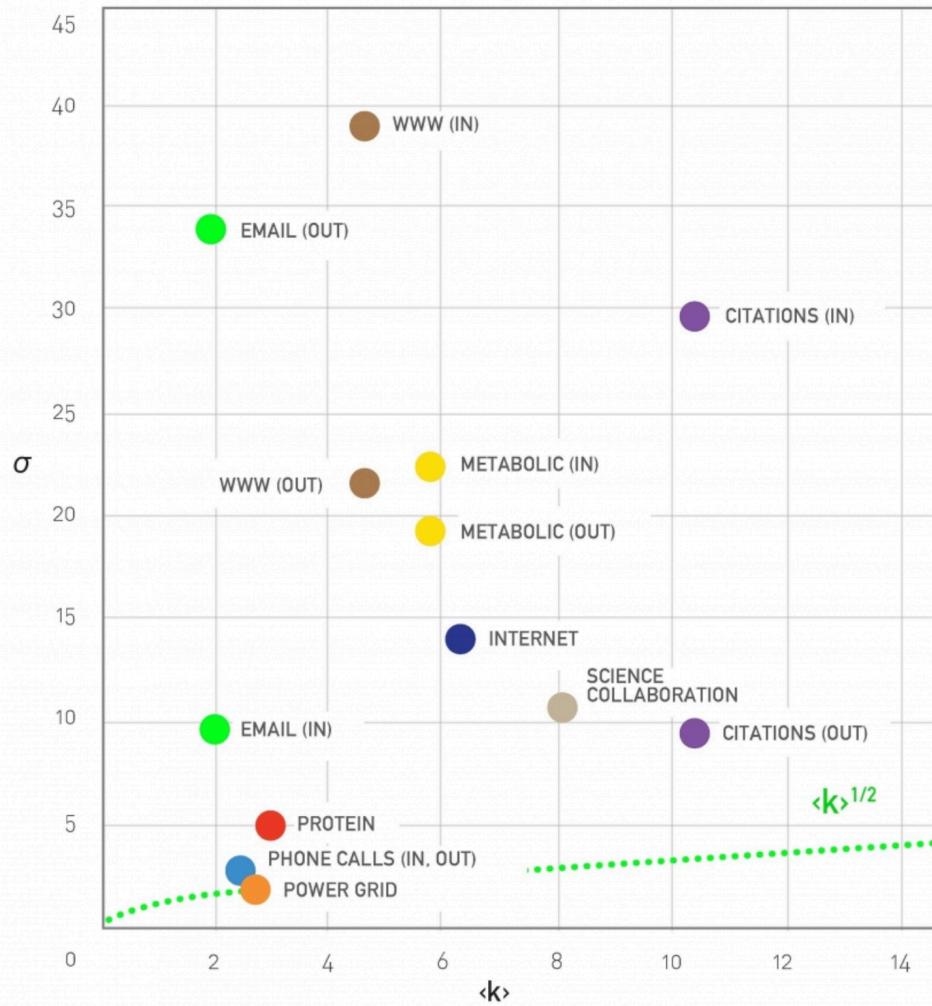
$$k = \langle k \rangle \pm \langle k \rangle^{1/2}$$

Scale-Free Network

Nodo elegido al azar:
Scale: none

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

El significado de sin escala



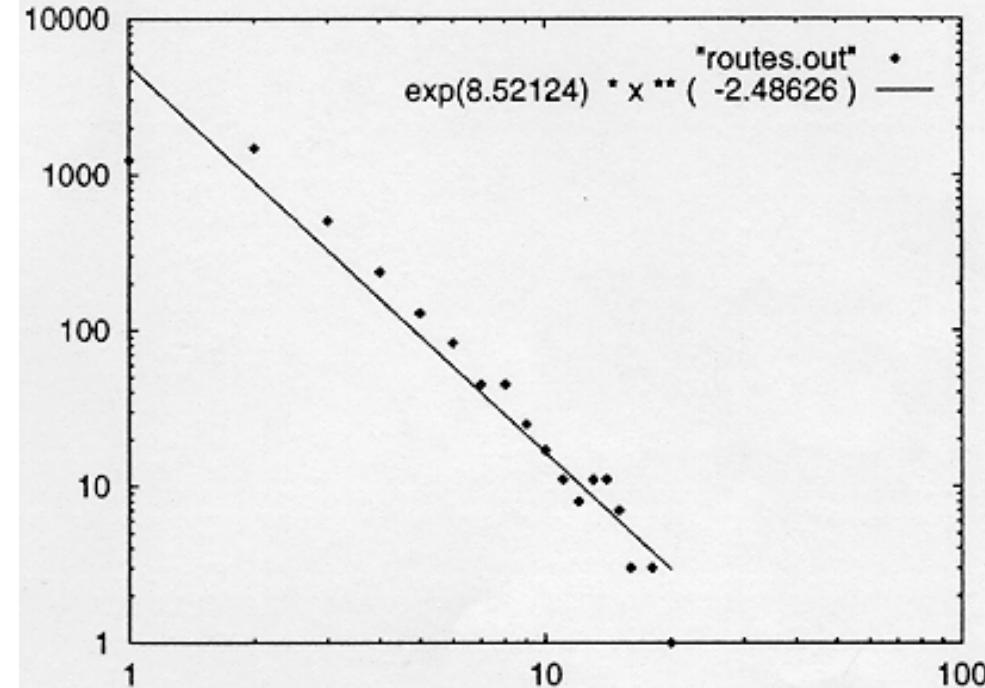
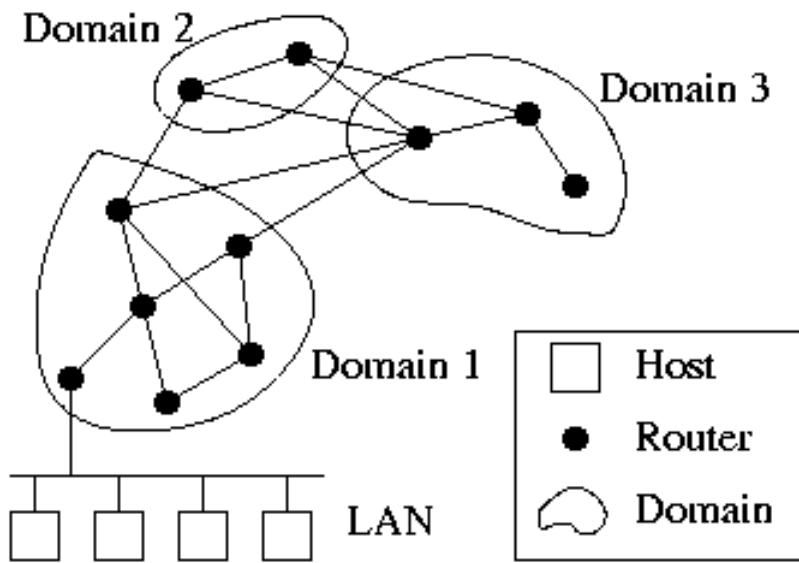
$$k = \langle k \rangle \pm \sigma_k$$

Universalidad

Backbone de internet

Nodos: computadores, routers

Enlaces: líneas físicas



(Faloutsos, Faloutsos and Faloutsos, 1999)

Network Science: Scale-Free Property



Network Science: Scale-Free Property

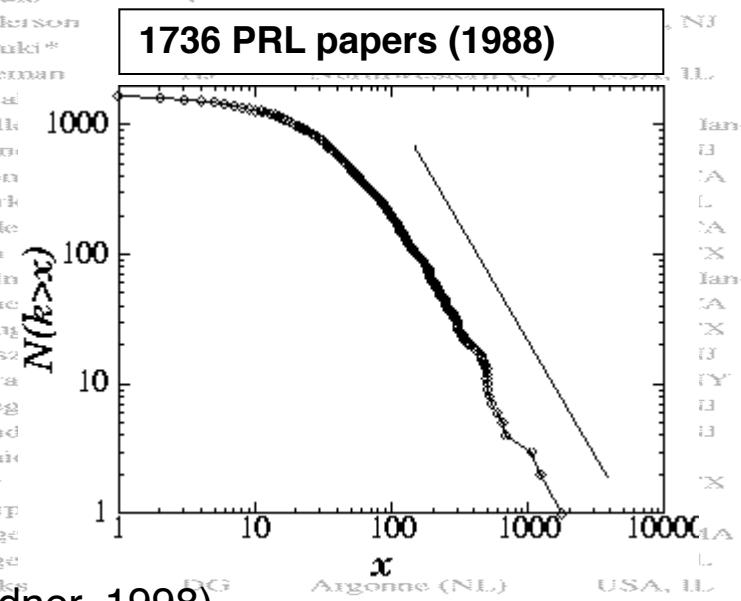
INDICE DE CITACIONES EN CIENCIA

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

Nodos: papers

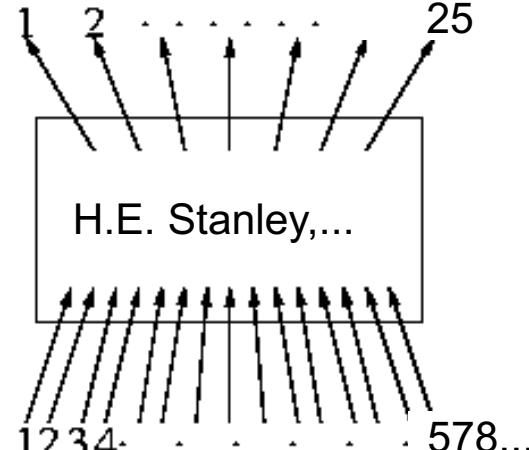
Enlaces: citaciones

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



(S. Redner, 1998)

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

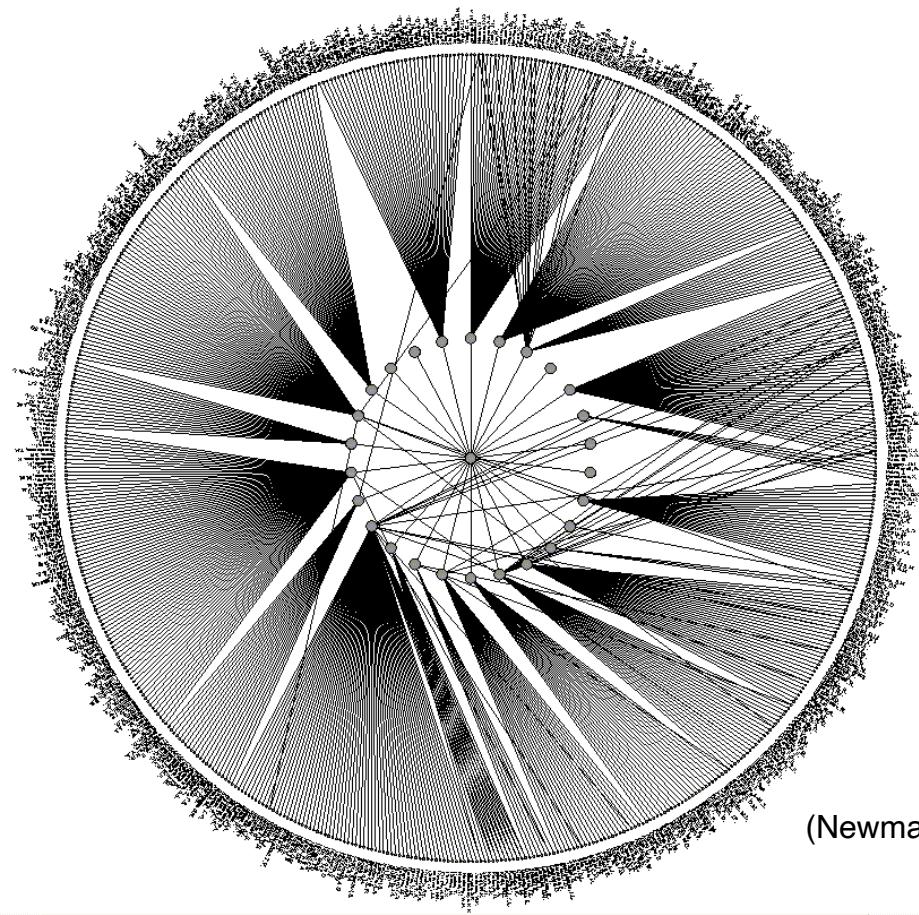


$$P(k) \sim k^{-\gamma} \quad (\gamma = 3)$$

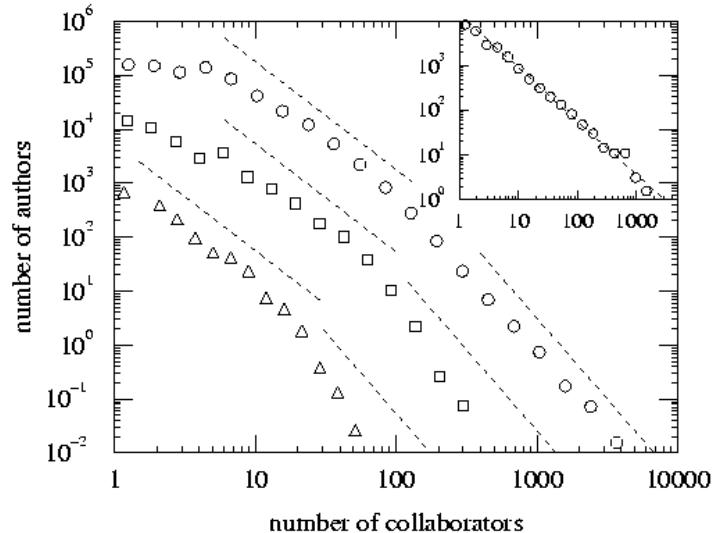
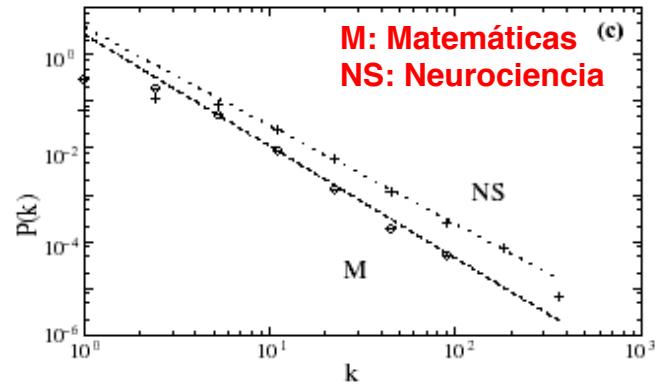
CO-AUTORIA EN CIENCIA

Nodos: Científicos (autores)

Links: Publicaciones conjuntas



(Newman, 2000, Barabasi et al 2001)

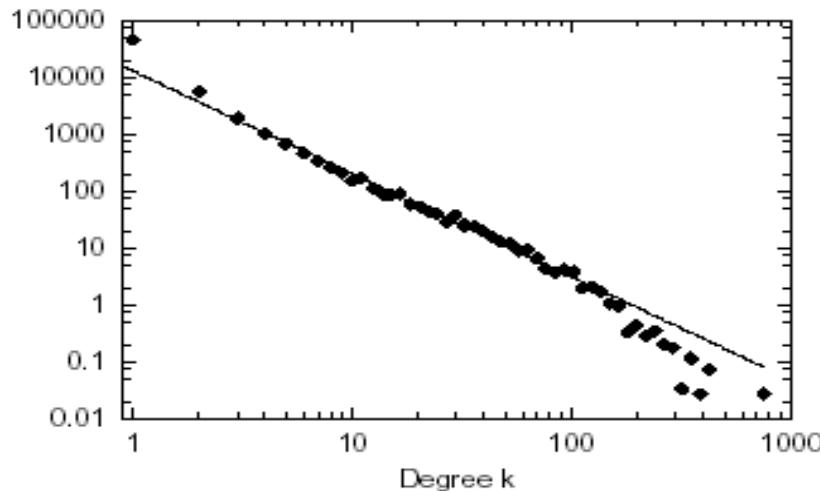


COMUNIDADES ONLINE

Nodos: Usuarios online

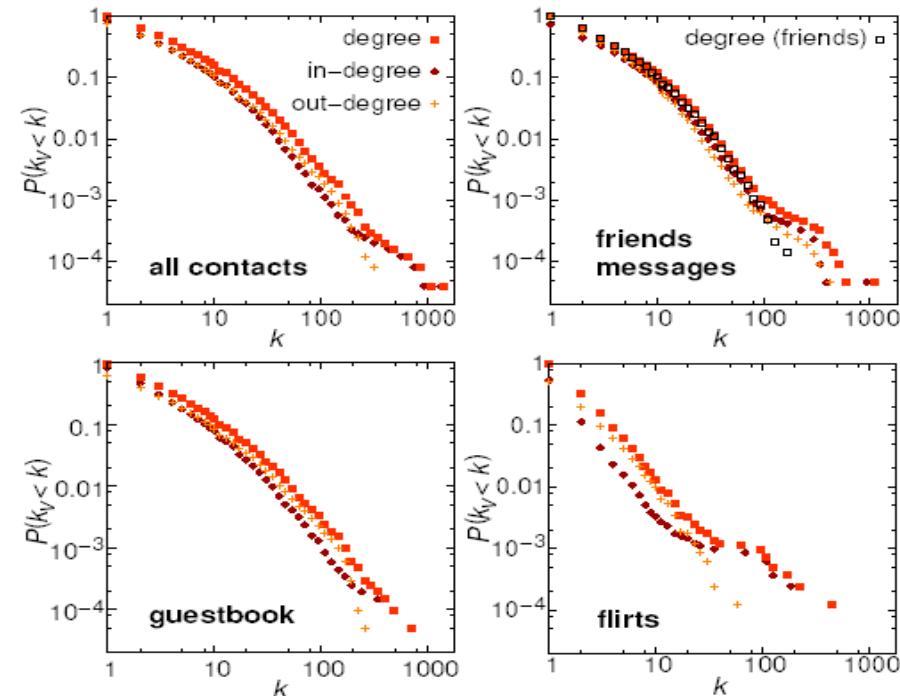
Enlaces: contactos email

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



Ebel, Mielsch, Bornholdtz, PRE 2002.

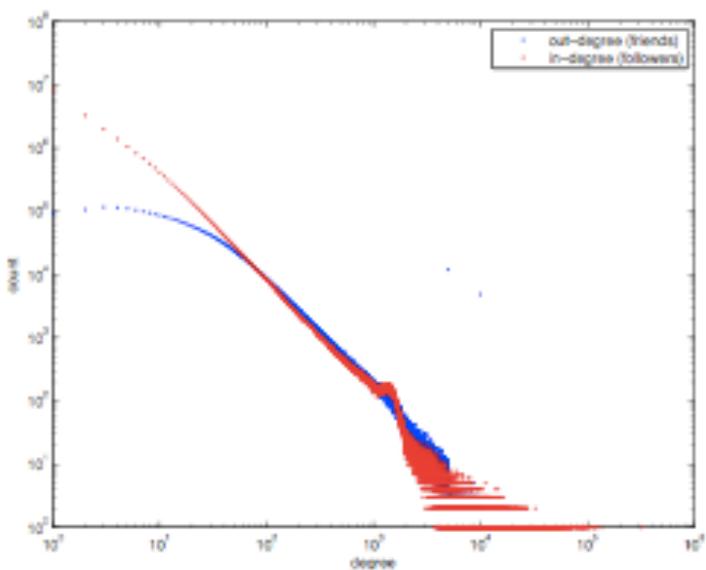
Pussokram.com comunidad online;
512 days, 25,000 users.



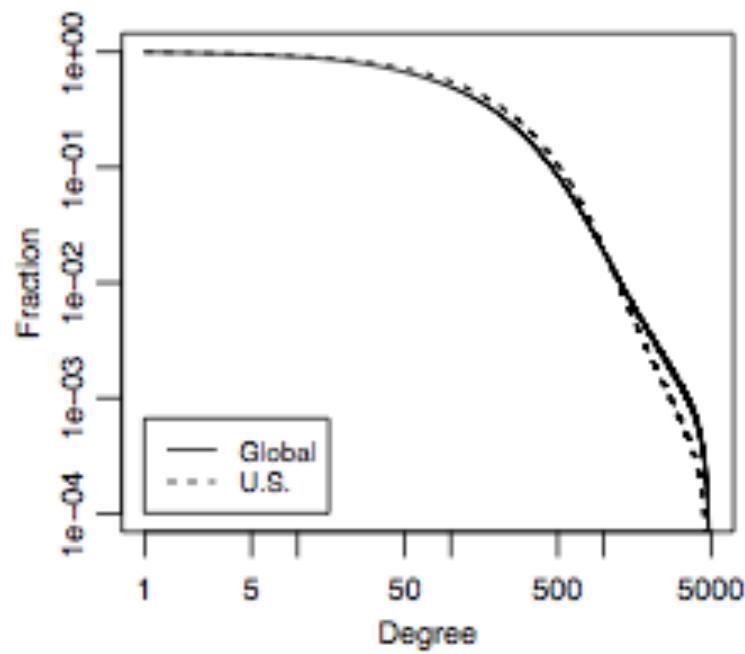
Holme, Edling, Liljeros, 2002.

COMUNIDADES ONLINE

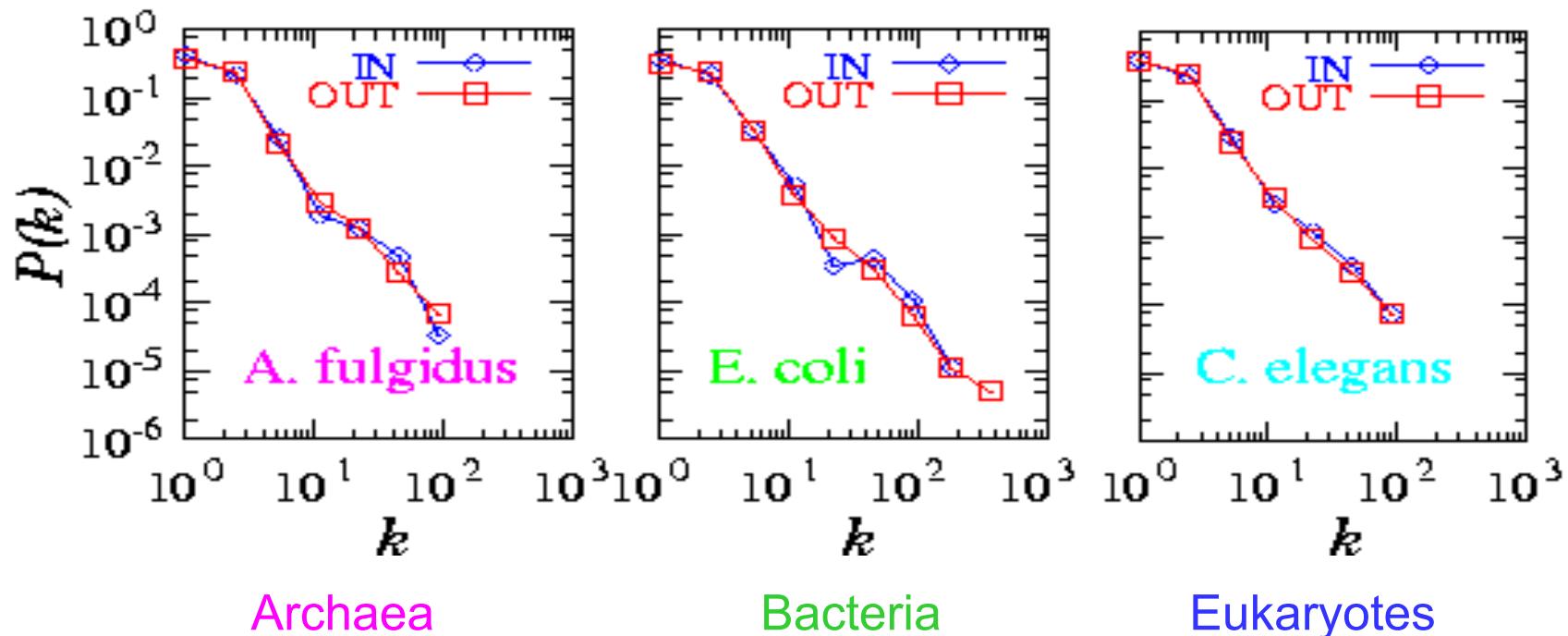
Twitter:



Facebook



RED METABOLICA



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

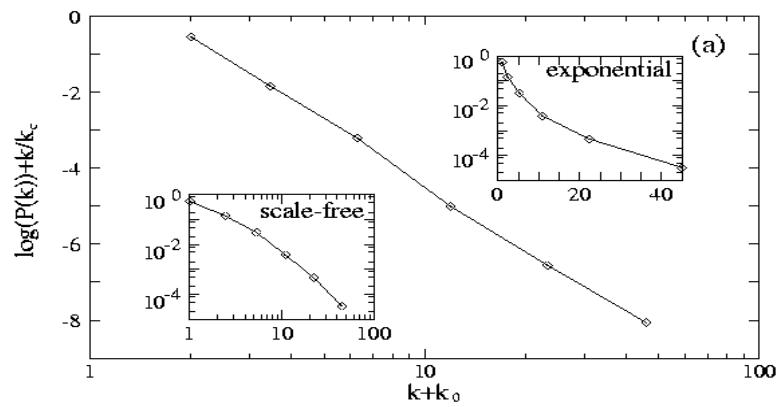
$$P_{in}(k) \approx k^{-2.2}$$

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

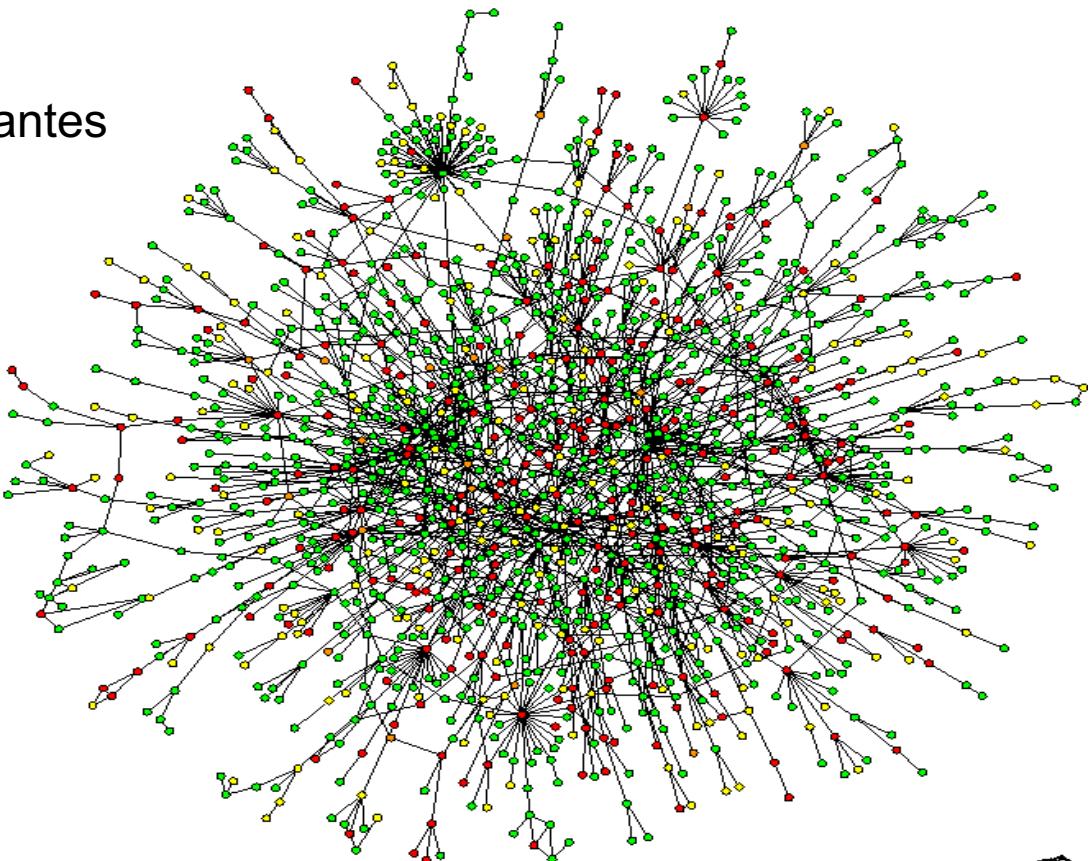
TOPOLOGÍA DE LA RED DE PROTEÍNAS

Nodos: proteínas

Enlaces: interacciones físicas vinculantes



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

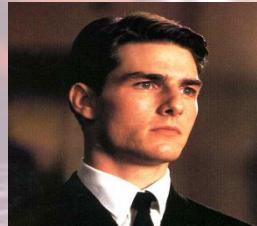


RED DE ACTORES

Nodos: actores

Enlaces: película conjunta

IMDb Internet Movie Database



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

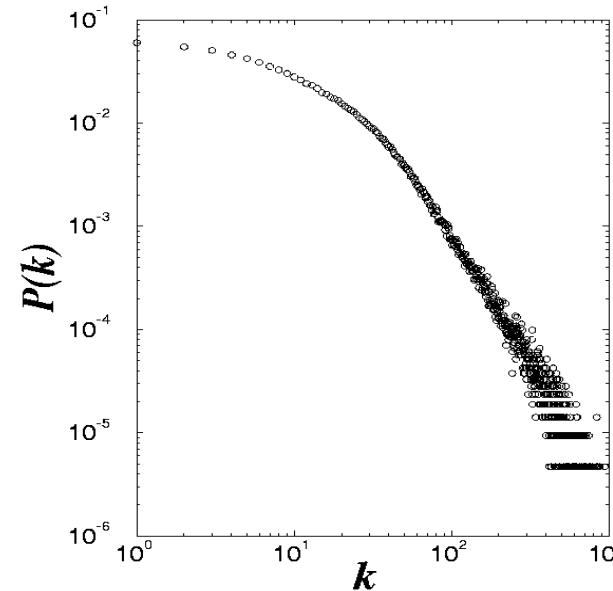


$N = 212,250$ actores

$\langle k \rangle = 28.78$

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

$\gamma=2.3$

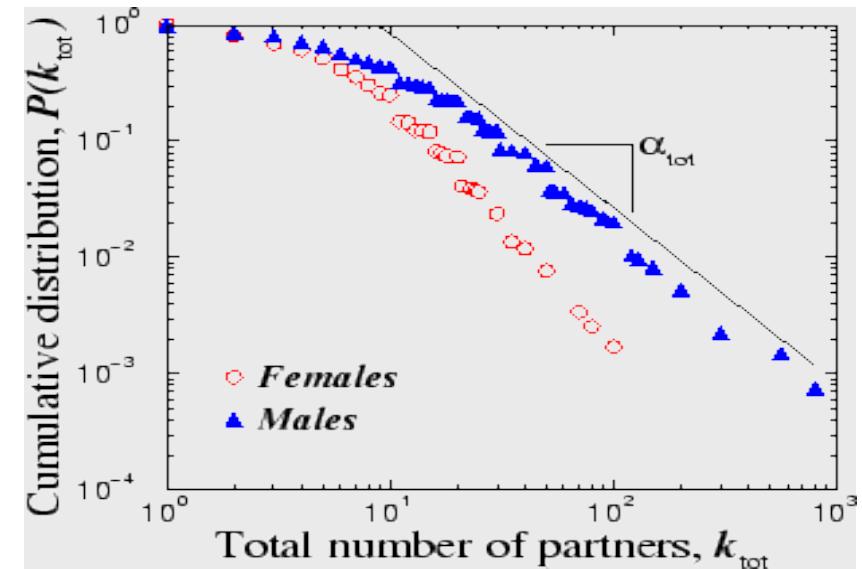


SWEDISH SE-WEB



Nodos: Personas (Mujeres; Hombres)

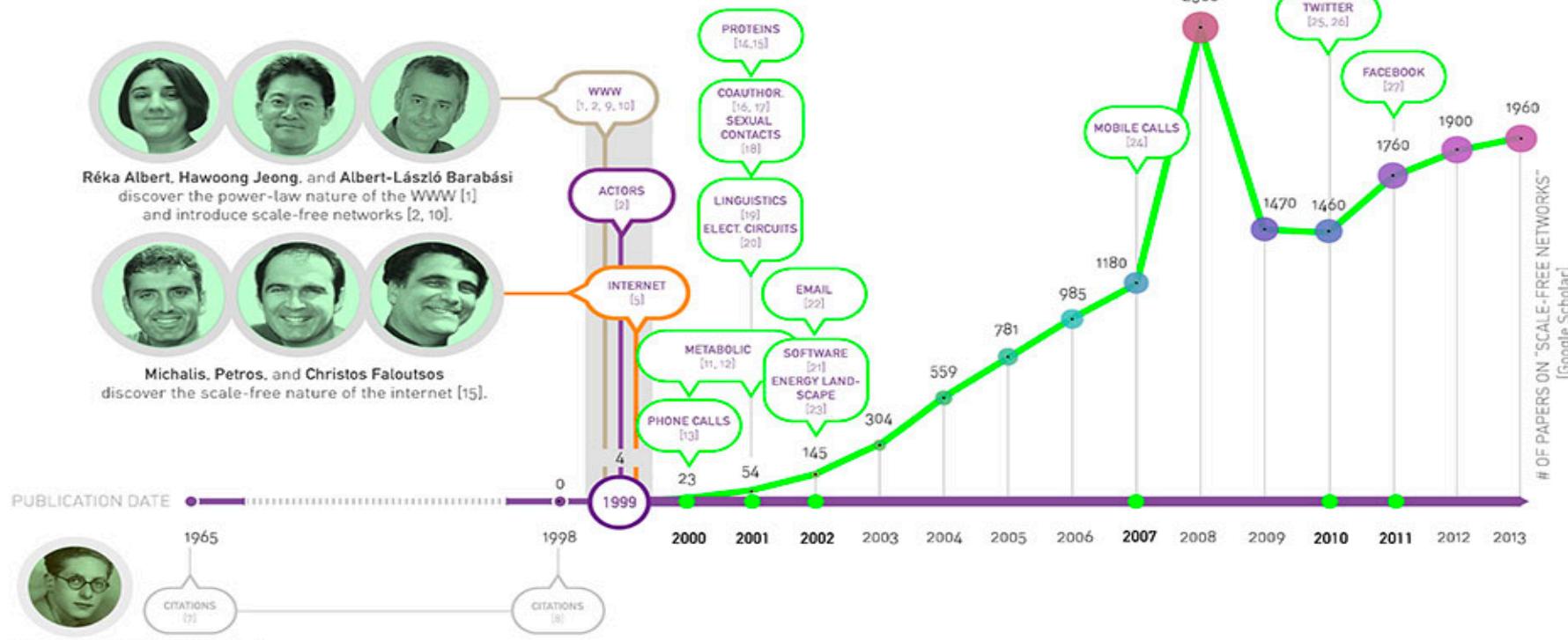
Enlaces: Relaciones sexuales



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

Liljeros et al. Nature 2001

Línea de tiempo de las redes libres de escala

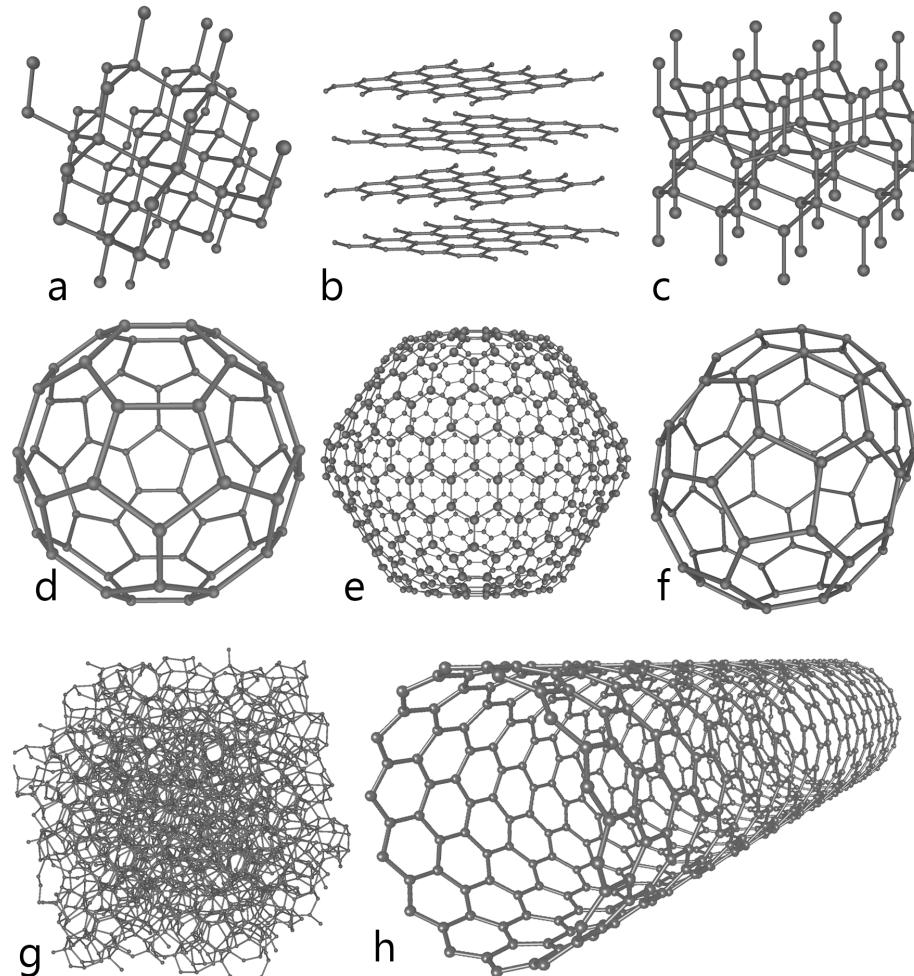


"we expect that the scale-invariant state observed in all systems for which detailed data has been available to us is a generic property of many complex networks, with applicability reaching far beyond the quoted examples."

Barabási and Albert, 1999

No todas las redes son libres de escala

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



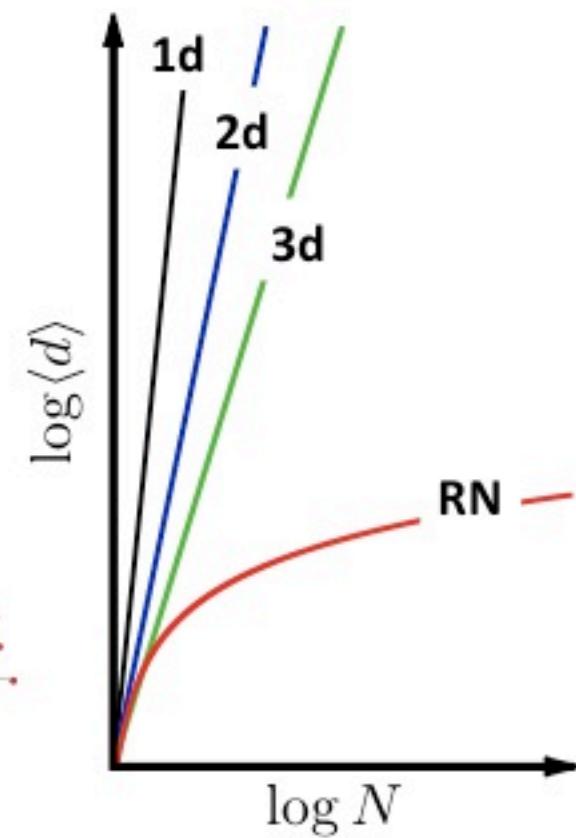
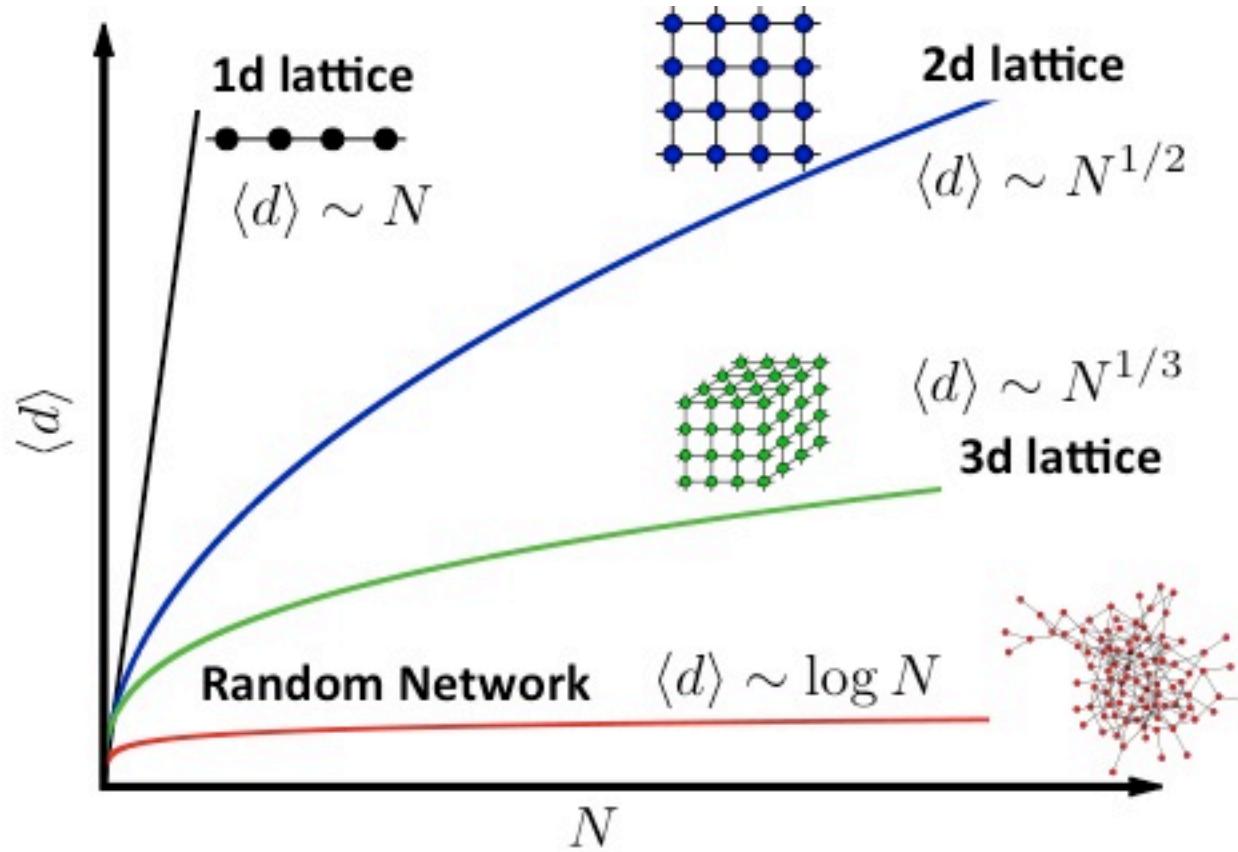
Propiedad ultra-pequeña

COMPORTAMIENTO DE MUNDO PEQUEÑO EN REDES SIN ESCALA

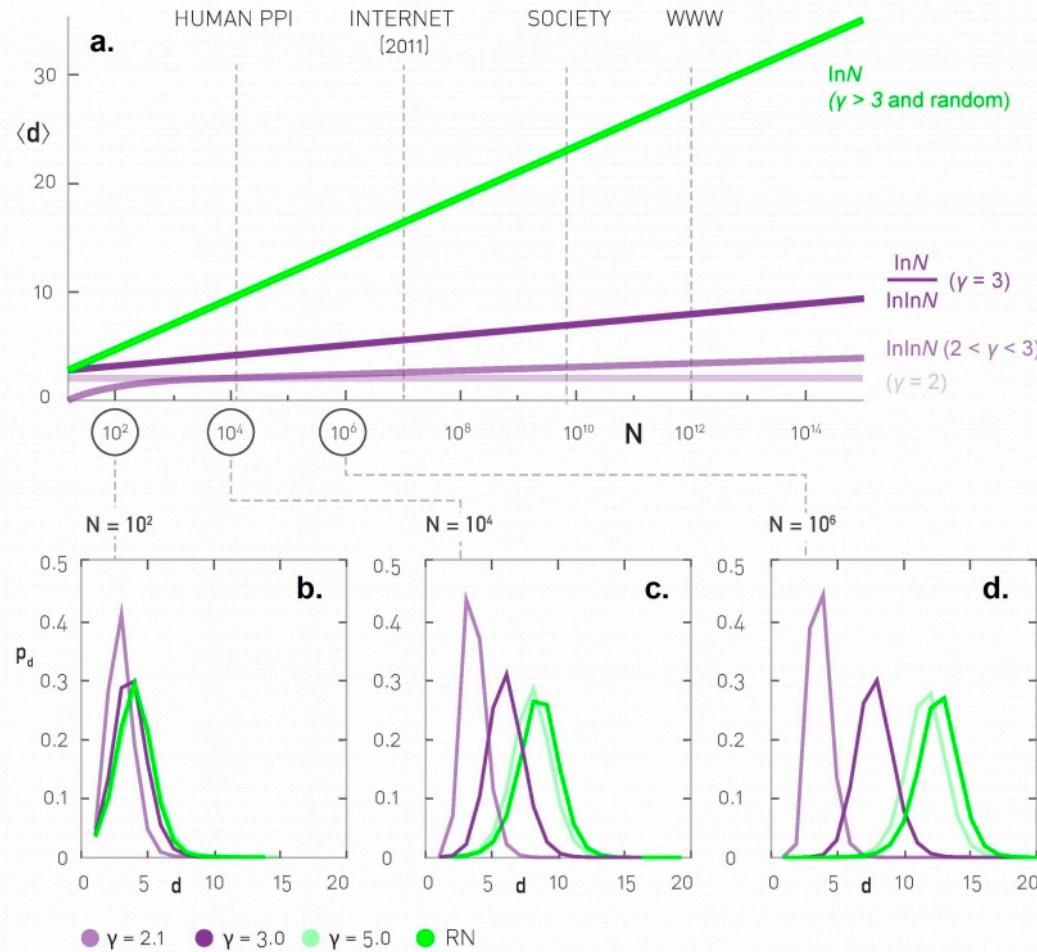
**Ultra
Small
World**

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

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



COMPORTAMIENTO DE MUNDO PEQUEÑO EN REDES SIN ESCALA

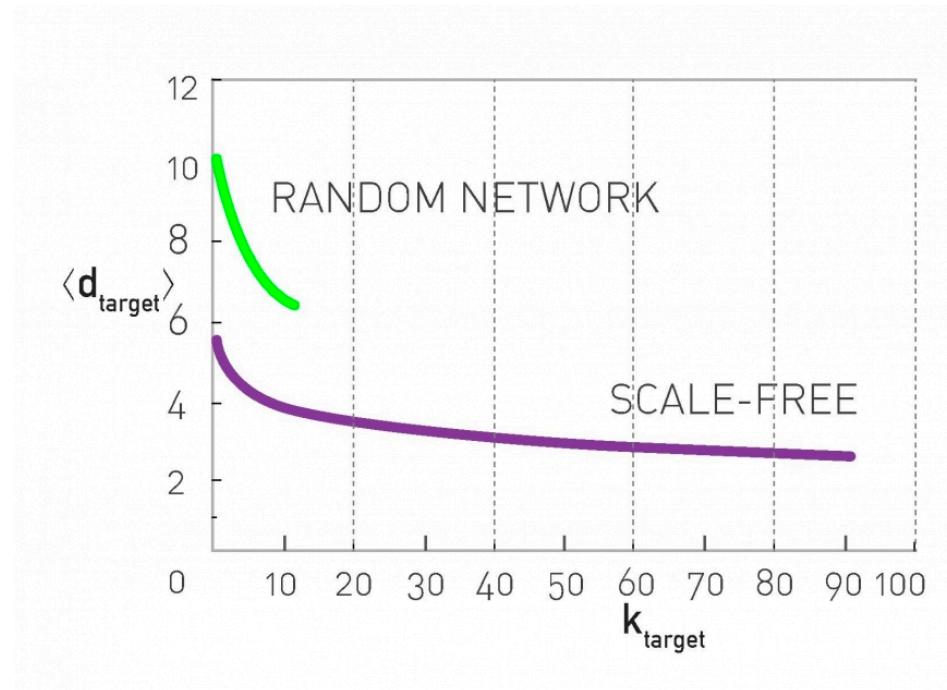


$$\langle d \rangle \sim \begin{cases} \text{const.} & \gamma = 2, \\ \frac{\ln \ln N}{\ln(\gamma - 1)} & 2 < \gamma < 3, \\ \frac{\ln N}{\ln \ln N} & \gamma = 3, \\ \ln N & \gamma > 3. \end{cases}$$

Siempre estamos cerca de los hubs.

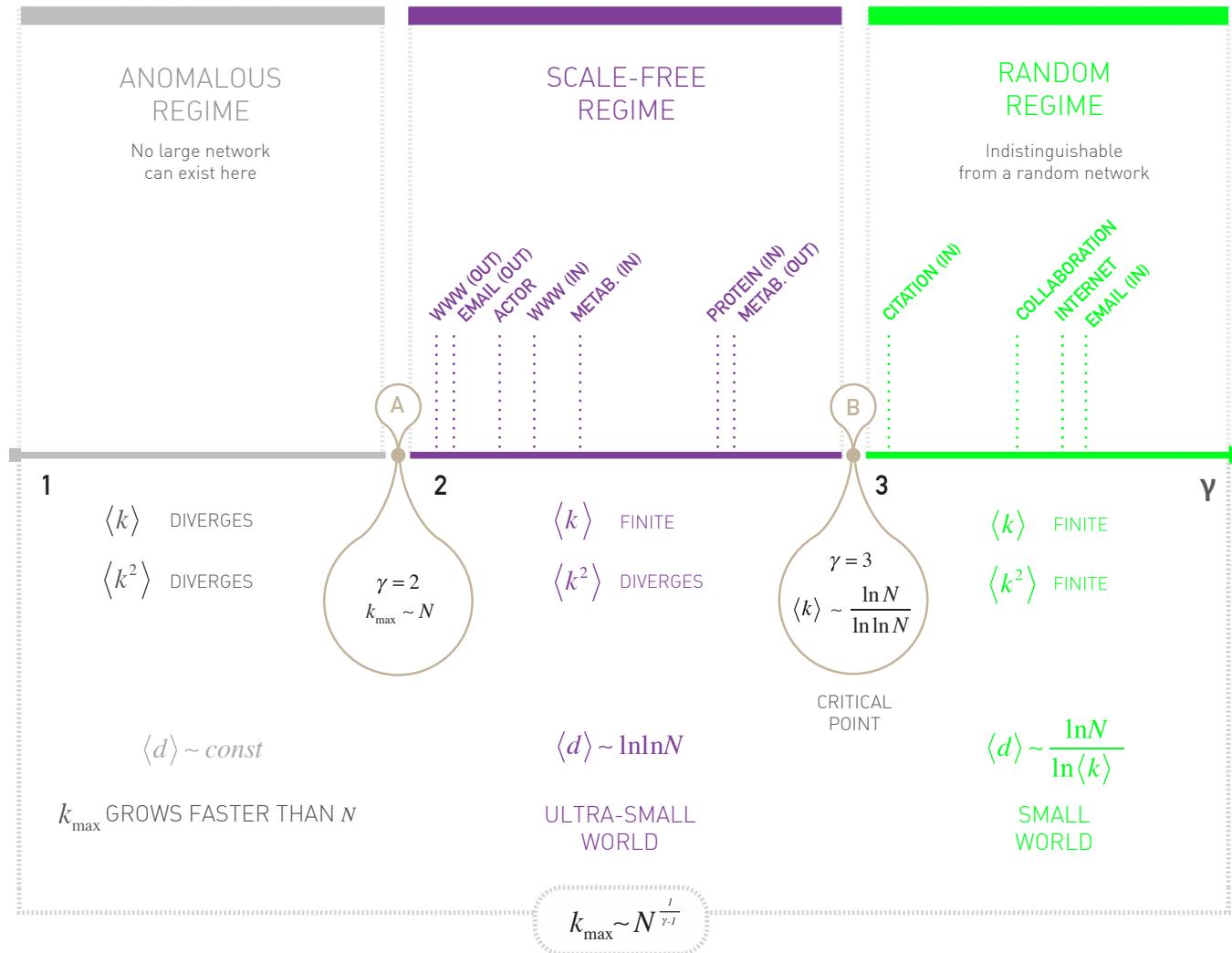
"Siempre es más fácil encontrar a alguien que conozca a una figura famosa o popular que a una persona insignificante."

(Frigyes Karinthy, 1929)



El rol del exponente del grado.

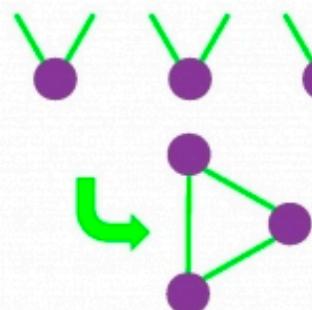
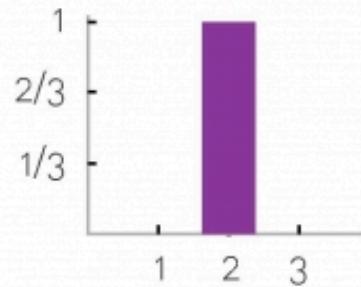
RESUMEN DEL COMPORTAMIENTO DE REDES SIN ESCALA



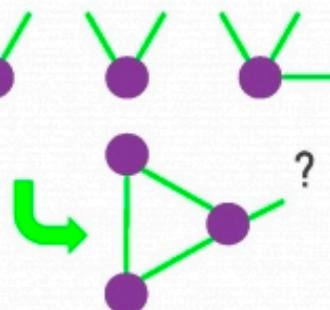
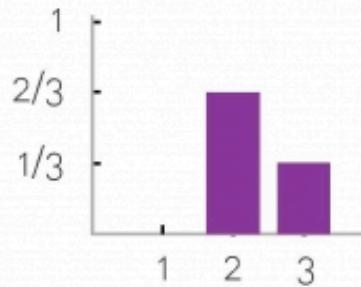
Gáficamente: No hay redes grandes para $\gamma < 2$

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

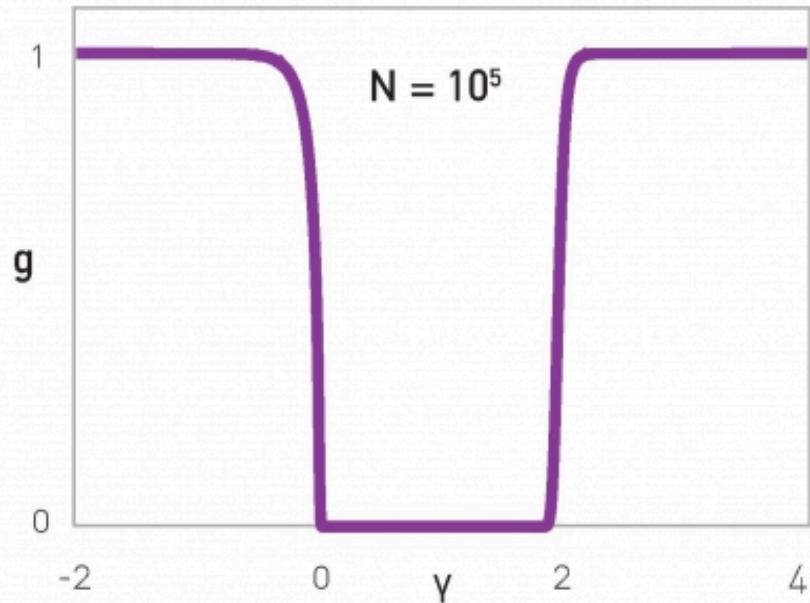
a. Graphical



b. Not Graphical



c.



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

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

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

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

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

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

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

Mobile Call Network

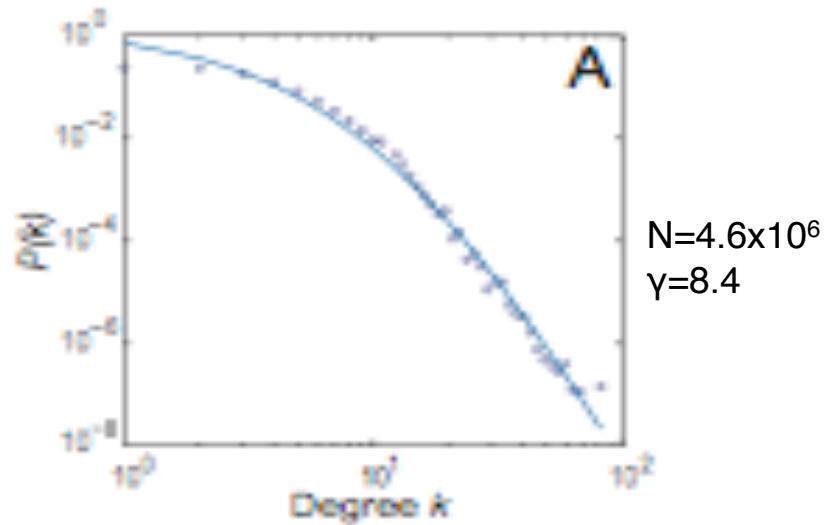
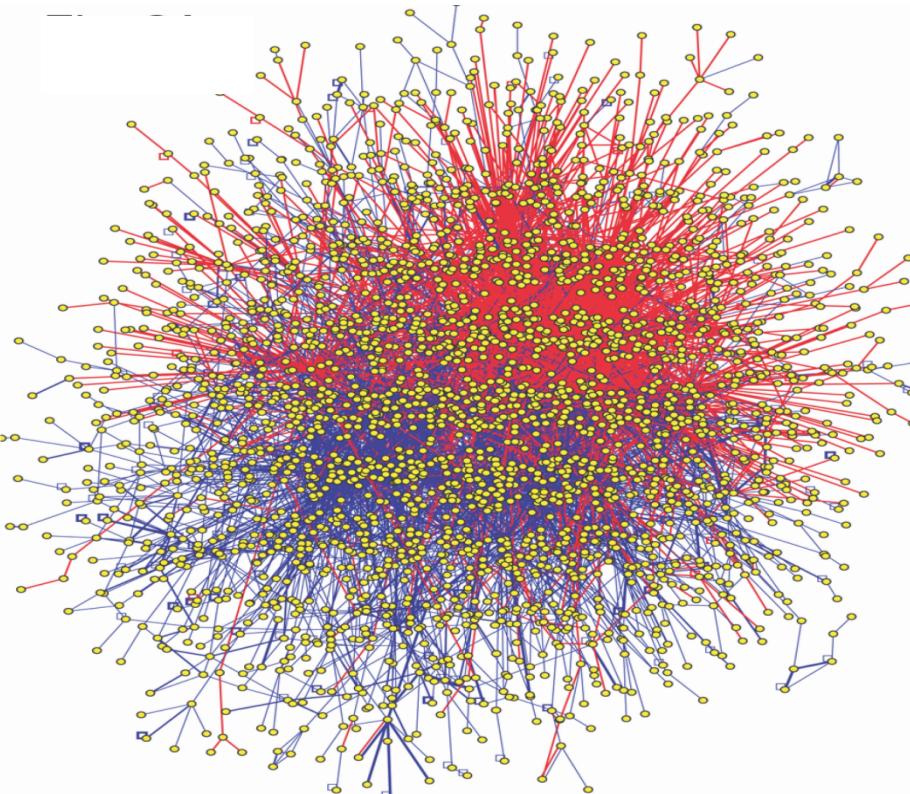


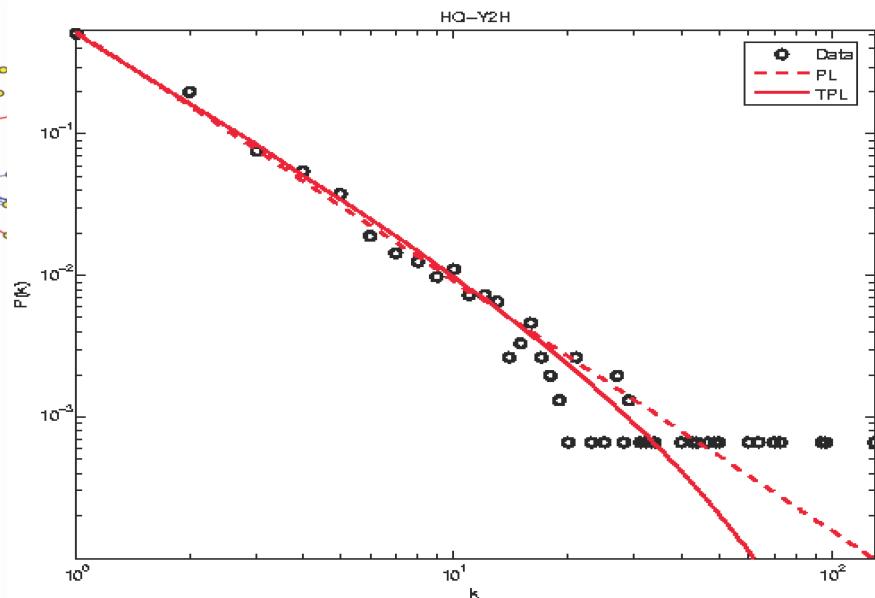
Fig. 1. Characterizing the large-scale structure and the tie strengths of the mobile call graph. (A) and (B) Vertex degree (A) and tie strength distribution (B). Each distribution was fitted with $P(x) = a(x + x_0)^{-\gamma} \exp(-x/x_0)$, shown as a blue curve, where x corresponds to either k or w .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

GRAFICANDO LEYES DE POTENCIA

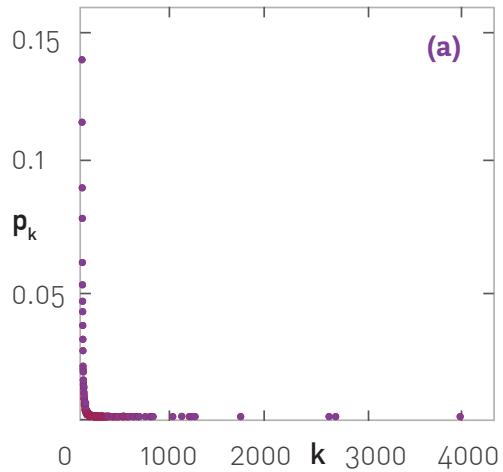
RED DE INTERACCION HUMANA



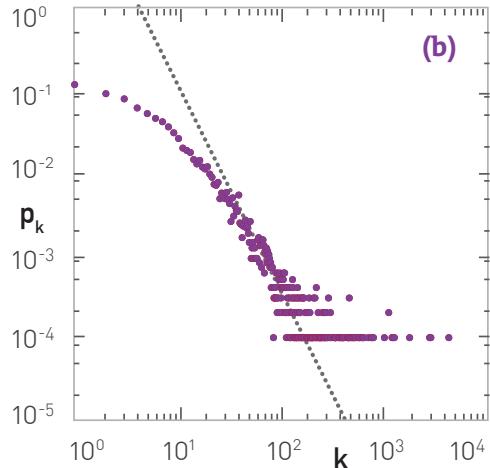
2,800 Y2H interactions
4,100 binary LC interactions
(HPRD, MINT, BIND, DIP, MIPS)



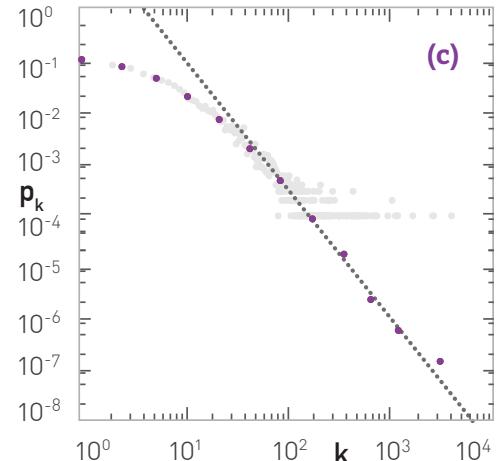
LINEAR SCALE



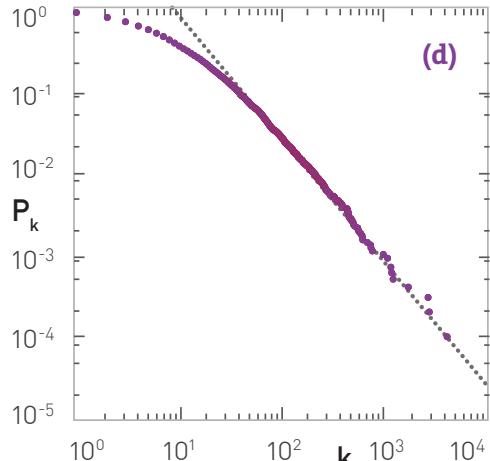
LINEAR BINNING



LOG-BINNING



CUMULATIVE



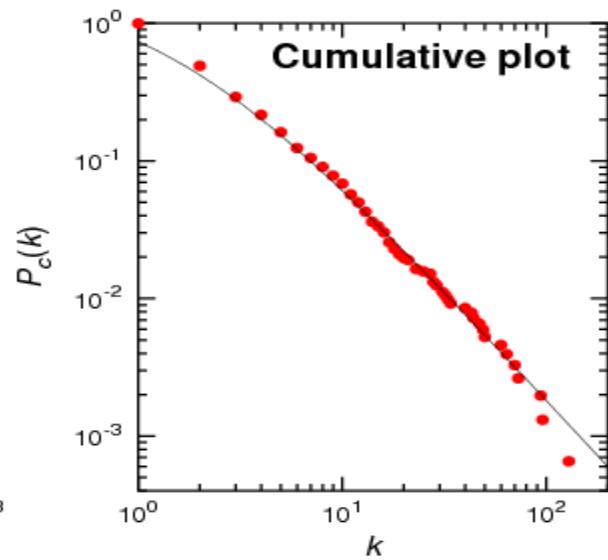
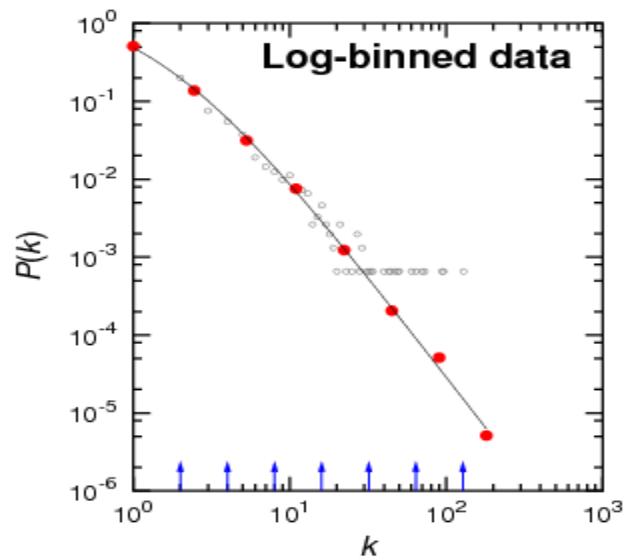
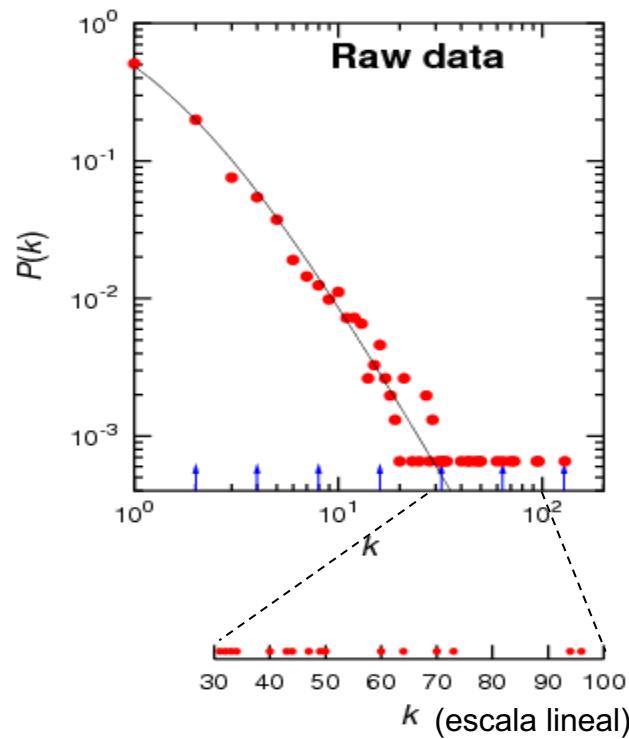
Usa gráficos log-log

Evita el bining lineal

Usa bining logarítmico

Usa la distribución acumulada

DATA DE INTERACCION HUMANA POR RUAL ET AL.



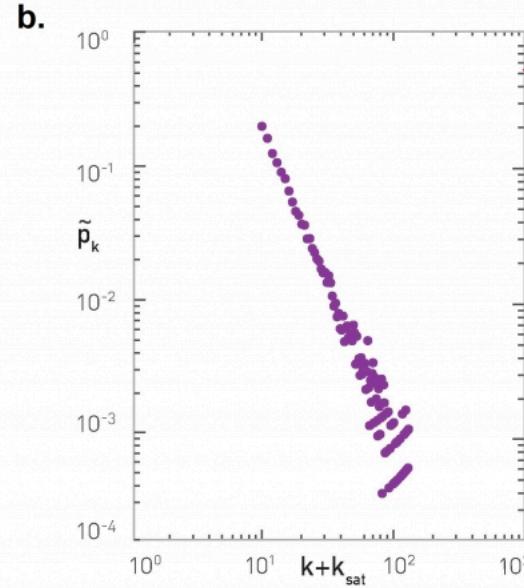
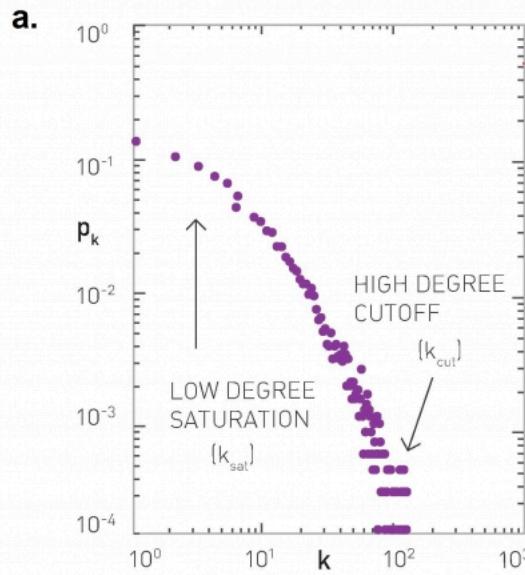
$$P(k) \sim (k+k_0)^{-\gamma}$$
$$k_0 = 1.4, \gamma = 2.6.$$

Fallas comunes

Reescalando la distribución del grado

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

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



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

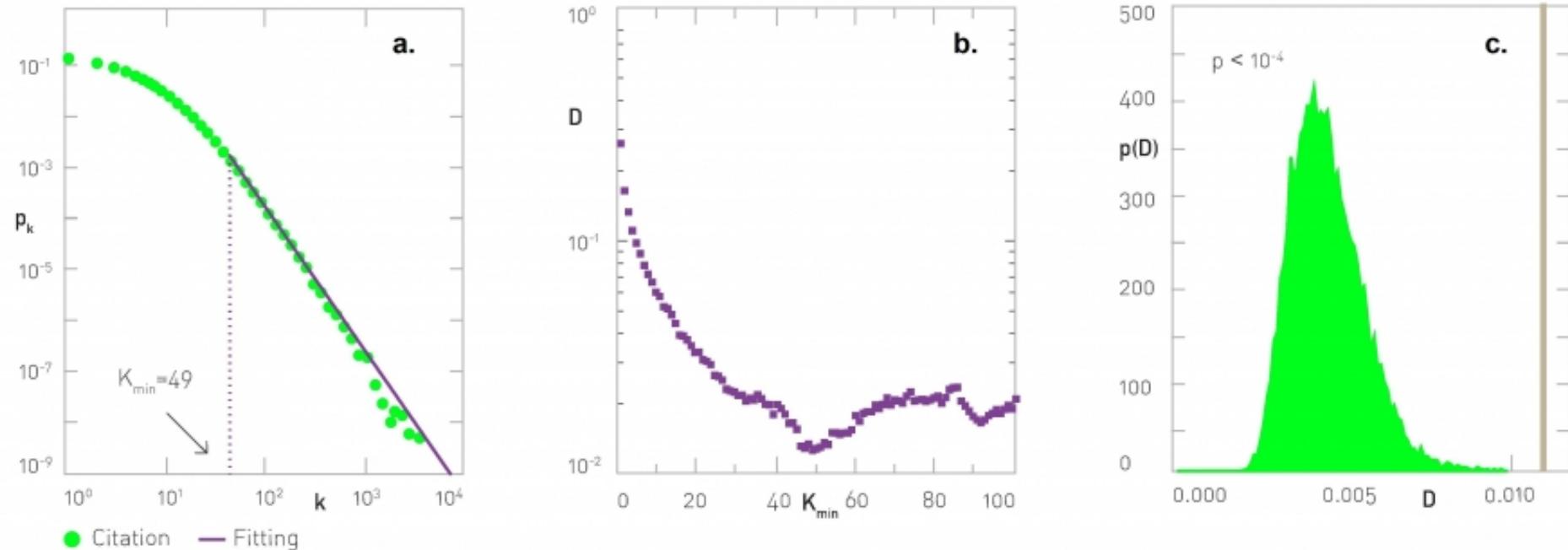
$$p_x = a(k + k_{sat})^{-\gamma} \exp\left(-\frac{k}{k_{cut}}\right) \quad (4.39)$$

Para extraer la extensión completa de la escala se grafica:

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

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

Procedimiento de ajuste



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

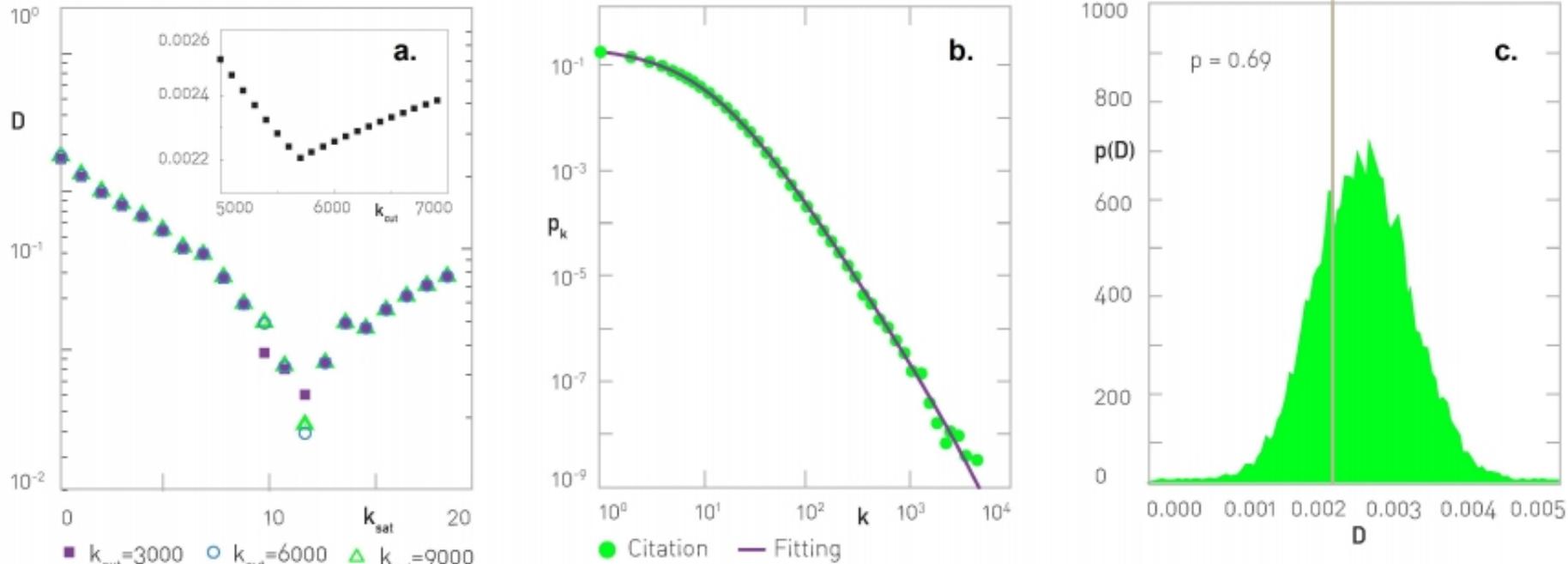
$$P_k = 1 - \frac{\zeta(\gamma, k)}{\zeta(\gamma, K_{\min})}$$

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

<https://aaronclauset.github.io/powerlaws/>

Network Science: Scale-Free Property

Procedimiento de ajuste



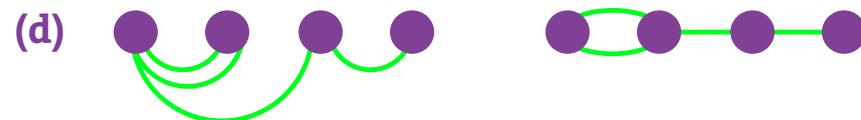
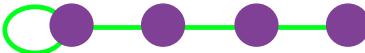
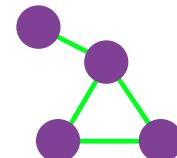
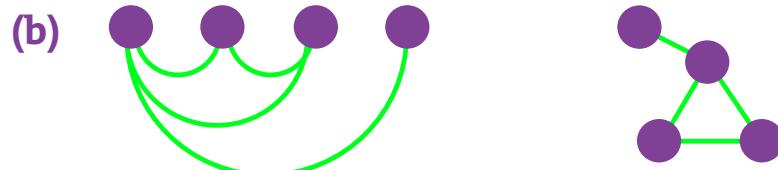
$$p_k = \frac{1}{\sum_{k'=1}^N (k'+k_{sat})^{-\gamma} e^{-k'/k_{cut}}} (k + k_{sat})^{-\gamma} e^{-k/k_{cut}} \quad (4.47)$$

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

Generando redes con un p_k predefinido

Modelo de configuración

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



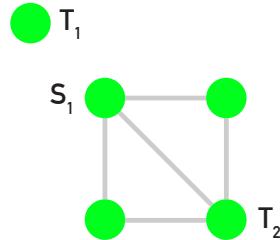
(1) **Secuencia de grados:** Asigne un grado a cada nodo, representado como cabos o medios enlaces. La secuencia de grados se genera analíticamente a partir de una distribución preseleccionada, o se extrae de la matriz de adyacencia de una red real. Debemos comenzar desde un número par de cabos, de lo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

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

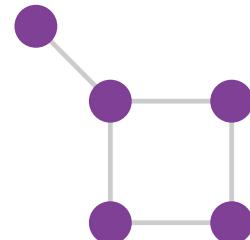
$$p_{ij} = \frac{k_i k_j}{2L - 1}$$

Preservación de grado aleatoria

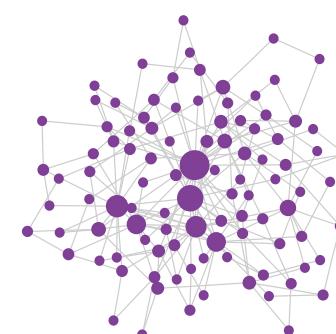
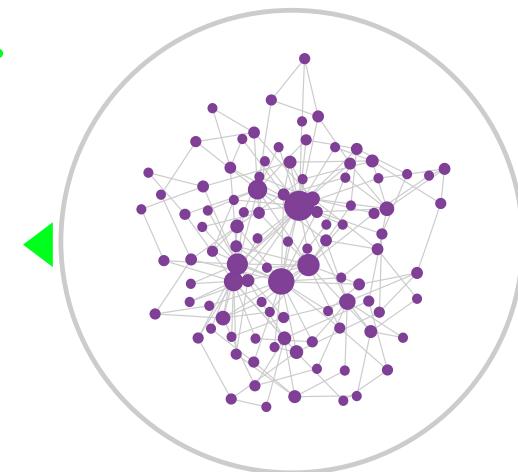
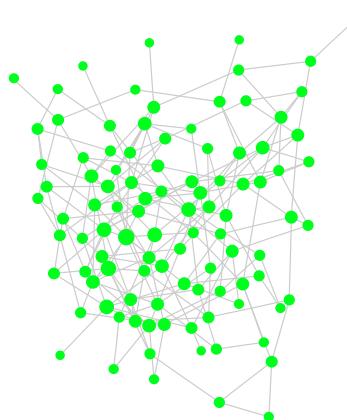
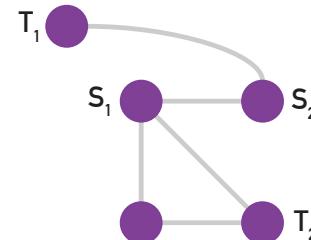
FULL
RANDOMIZATION



ORIGINAL NETWORK



DEGREE-PRESERVING
RANDOMIZATION

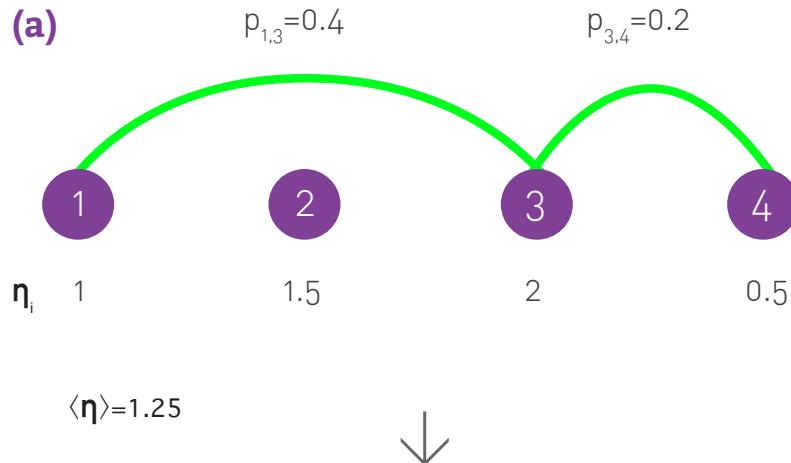


Ejercicio Fuertemente Recomendado

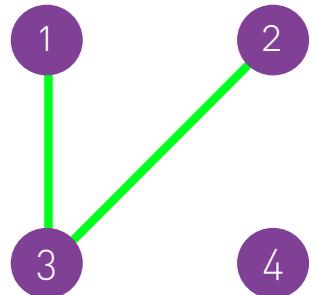
- Estudiar del paquete networkx los modelos de randomización
- `generators.degree_seq.configuration_model`
- https://networkx.org/documentation/latest/reference/generated/networkx.generators.degree_seq.configuration_model.html

Modelo de parámetro oculto

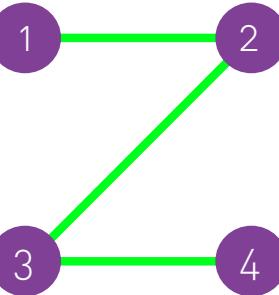
(a)



(b)



(c)



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

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

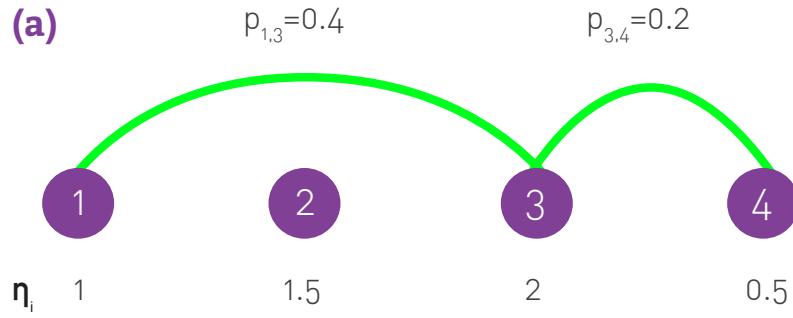
$$\{\eta_1, \eta_2, \dots, \eta_N\}$$

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

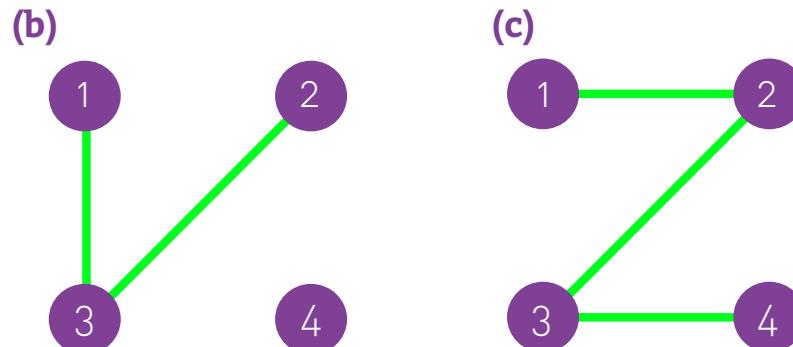
$$\eta_j = \frac{c}{i^\alpha}, i = 1, \dots, N$$

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

Modelo de parámetro oculto



$$\langle \eta \rangle = 1.25$$



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

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

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

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

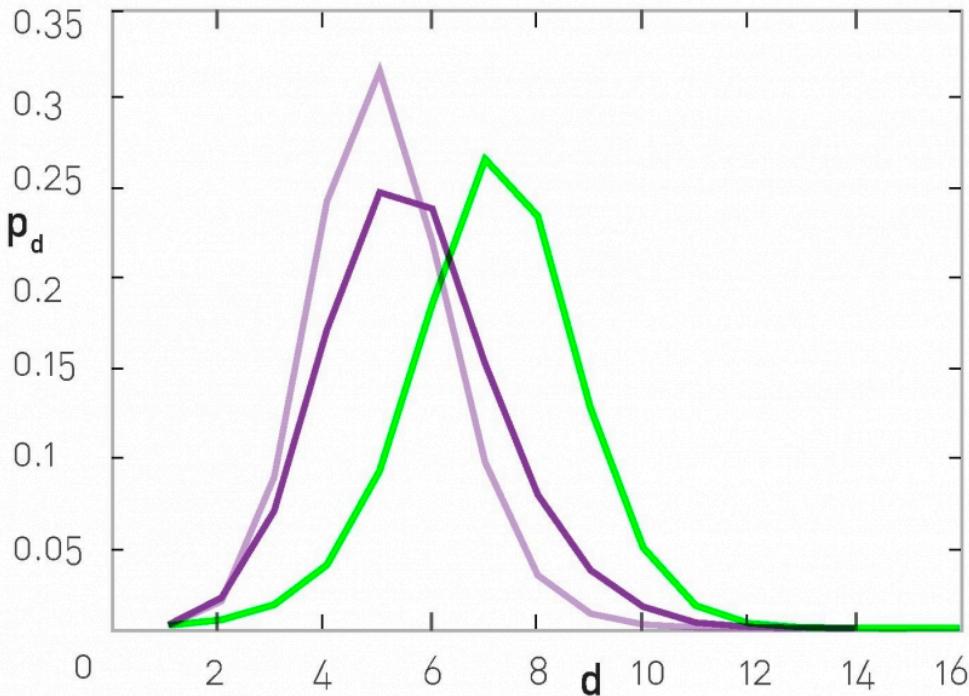
Por ejemplo, la figura muestra la probabilidad de conectar los nodos (1,3) y (3,4).

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

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

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

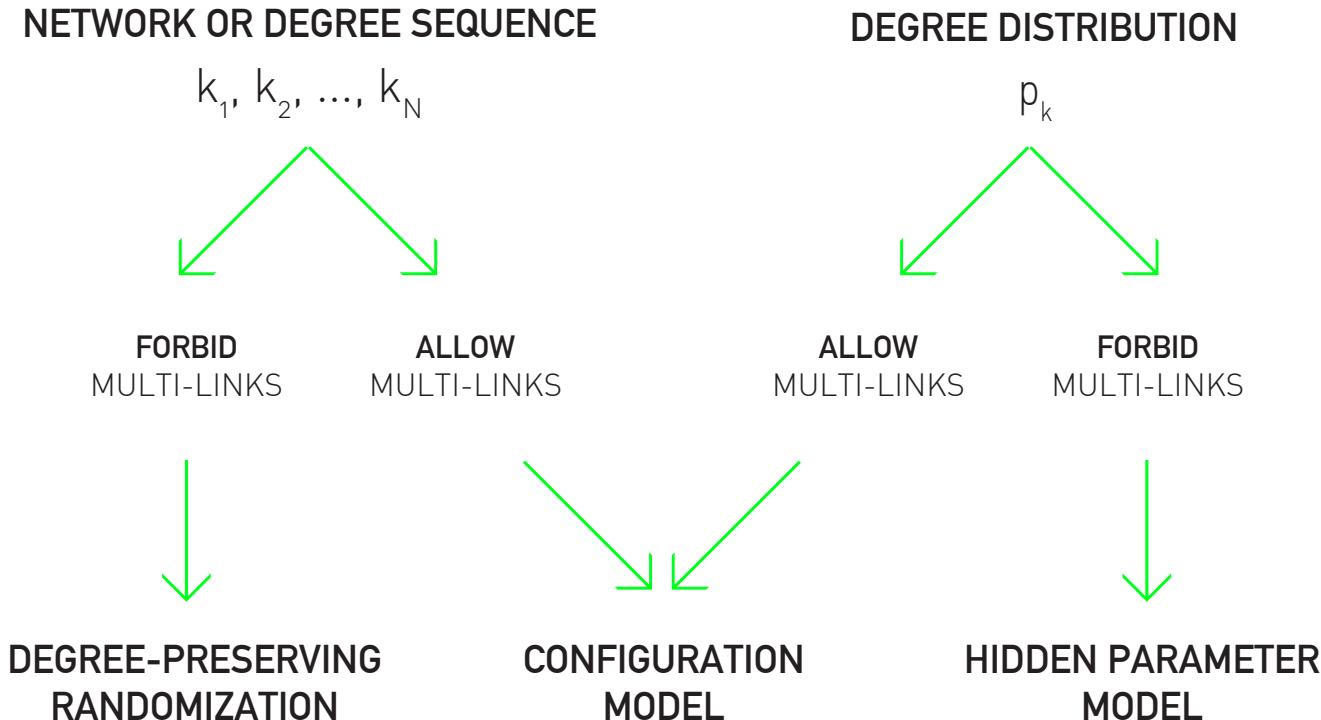
Caso de estudio: Red PPI



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

- Original network
- Degree preserving randomization
- Full randomization

Arbol de decisión



Resumen

Section 9

DEGREE DISTRIBUTION

Discrete form:

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

Continuous form:

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

SIZE OF THE LARGEST HUB

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

MOMENTS OF p_k for $N \rightarrow \infty$

$2 < \gamma < 3$: $\langle k \rangle$ finite, $\langle k^2 \rangle$ diverges.

$\gamma > 3$: $\langle k \rangle$ and $\langle k^2 \rangle$ finite.

DISTANCES

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

Bounded Networks

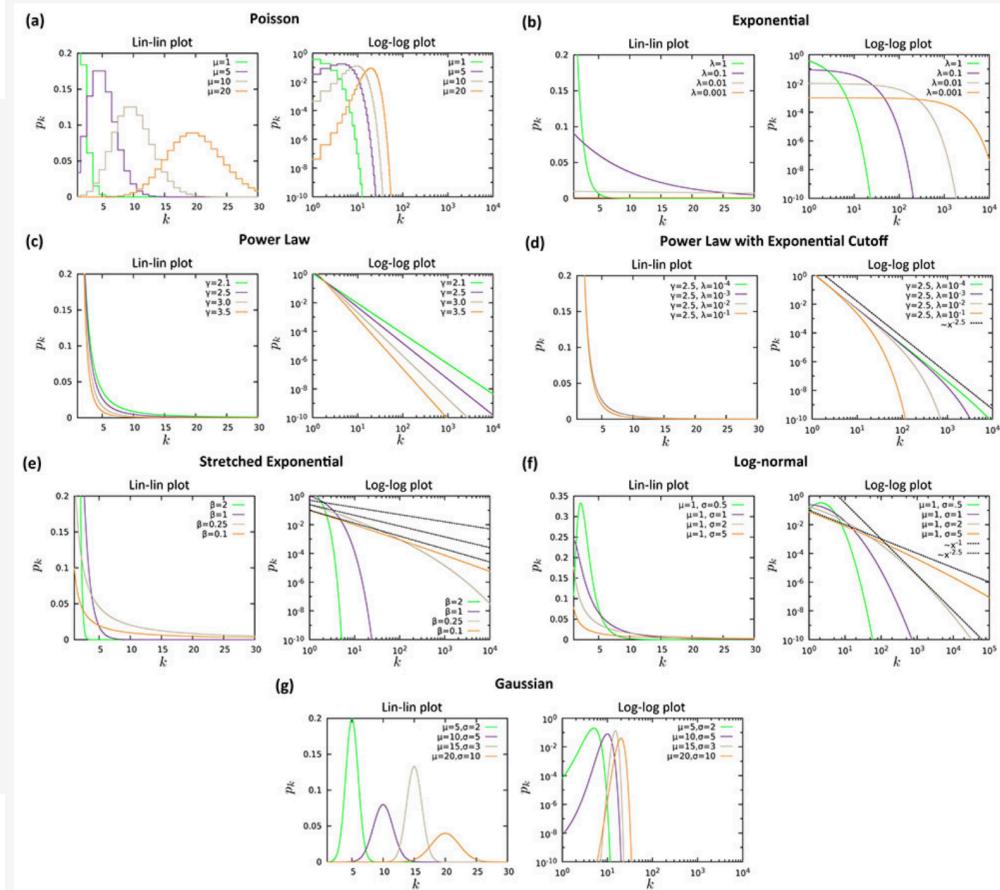
We call a network *bounded* if its degree distribution decrease exponentially or faster for high k . As a consequence $\langle k^2 \rangle$ is smaller than $\langle k \rangle$, implying that we lack significant degree variations. Examples of p_k in this class include the Poisson, Gaussian, or the simple exponential distribution (Table 4.2). The Erdős-Rényi and the Watts-Strogatz networks are the best known network models belonging to this class. Bounded networks lack outliers, consequently most nodes have comparable degrees. Real networks in this class include highway networks and the power grid.

Unbounded Networks

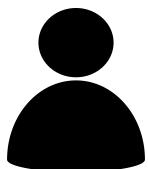
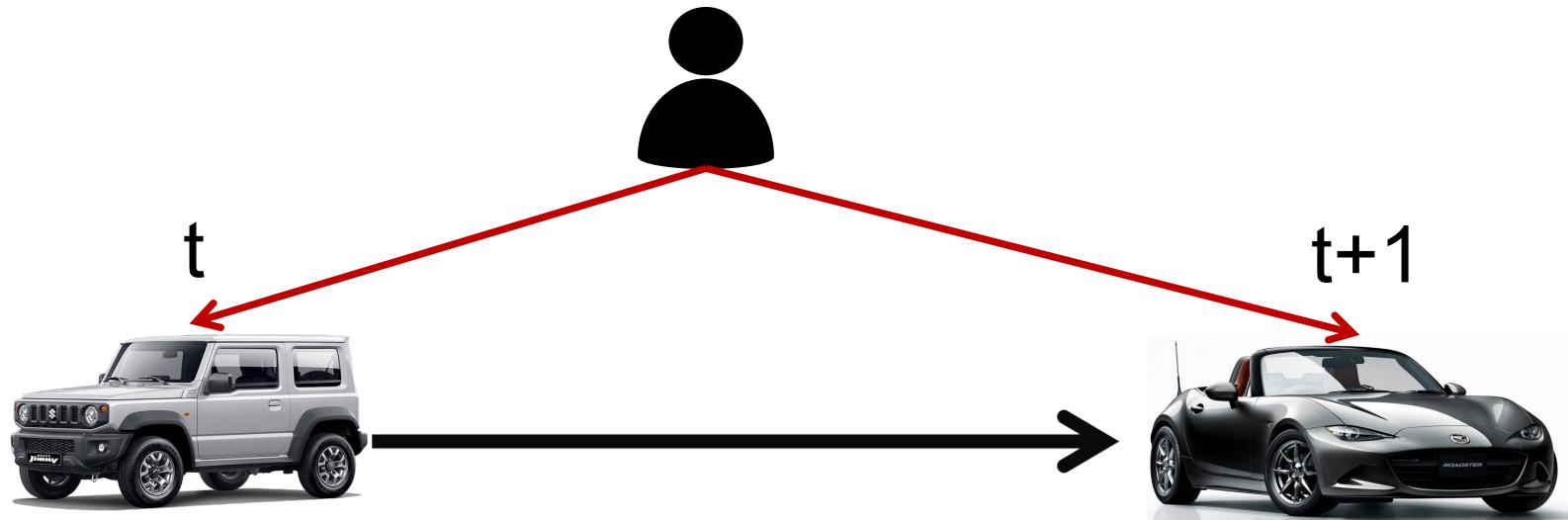
We call a network *unbounded* if its degree distribution has a fat tail in the high- k region. As a consequence $\langle k^2 \rangle$ is much larger than $\langle k \rangle$, resulting in considerable degree variations. Scale-free networks with a power-law degree distribution (4.1) offer the best known example of networks belonging to this class. Outliers, or exceptionally high-degree nodes, are not only allowed but are expected in these networks. Networks in this class include the WWW, the Internet, the protein interaction networks, and most social and online networks.

Tipos de distribuciones en redes

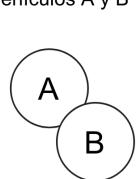
NAME	$p_k/p(x)$	$\langle x \rangle$	$\langle x^2 \rangle$
Poisson (discrete)	$e^{-\mu}\mu^x/x!$	μ	$\mu(1+\mu)$
Exponential (discrete)	$(1-e^{-\lambda})e^{-\lambda x}$	$1/(e^\lambda - 1)$	$(e^\lambda + 1)/(e^\lambda - 1)^2$
Exponential (continuous)	$\lambda e^{-\lambda x}$	$1/\lambda$	$2/\lambda^2$
Power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



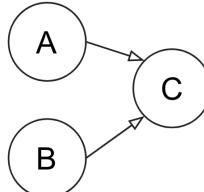
Conectamos vehículos



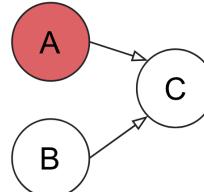
2016
Inicialmente tengo los vehículos A y B



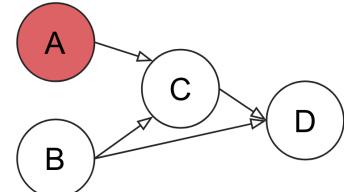
2017
Compro el vehículo C



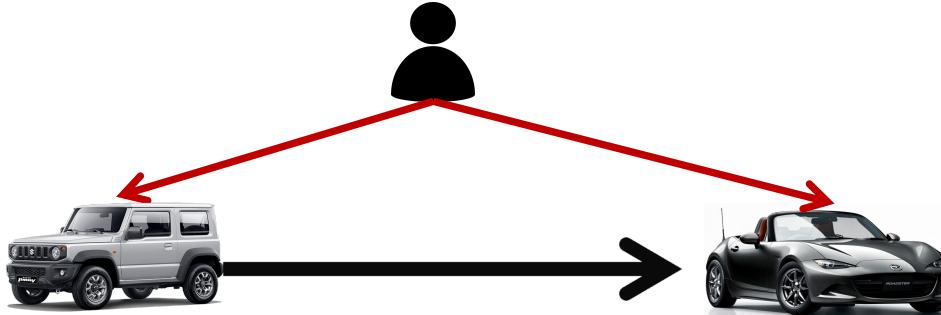
2018
Vendo el vehículo A



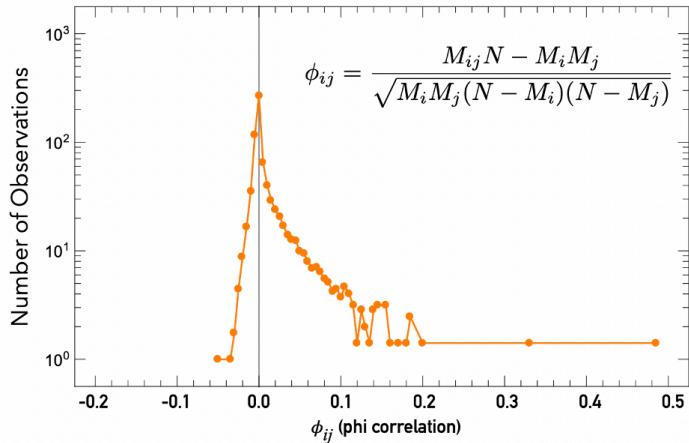
2020
Compro el vehículo D



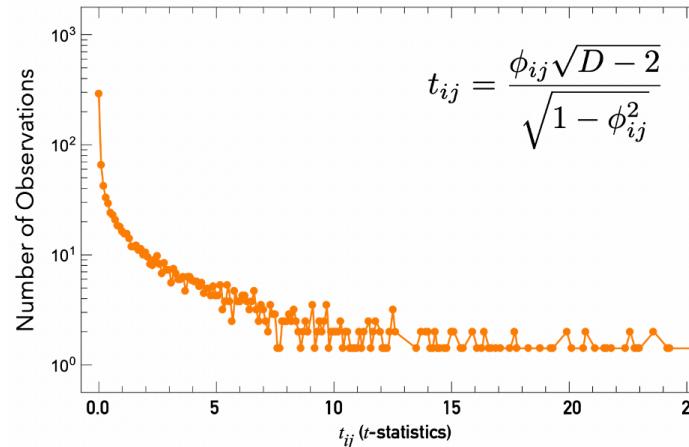
Conectamos vehículos



Encontramos la correlación entre vehículos



Calculamos la significancia de las correlaciones



Descartamos correlaciones negativas

Descartamos conexiones no significativas

Viz: 782 vehículos

55,301 enlaces

