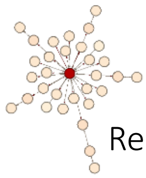


Redes Sociales

Jorge Fábrega

Formación de redes

1

¿Por qué estudiar modelos de formación de redes?

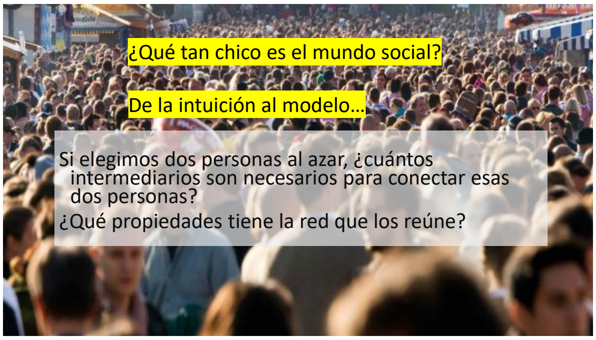
- Los modelos no buscan replicar cada red, sino ofrecer **modelos nulos de contraste**.
- Permiten identificar qué aspectos de las redes reales son **emergentes** y no triviales.
- Un modelo nulo es un punto de comparación.
  - Nos permite preguntar: ¿lo observado podría deberse solo al azar o al mecanismo X, Y o Z?
  - Diferencia entre estructura trivial y estructura significativa.

2

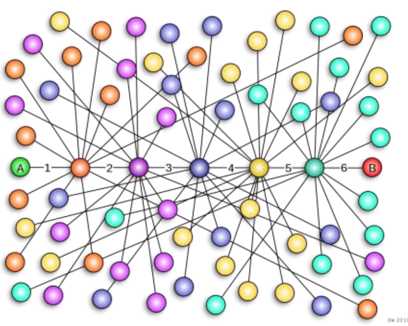
¿Qué tan chico es el mundo social?

De la intuición al modelo…

3



4



5

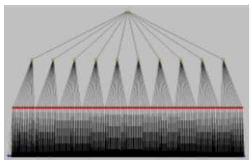
El concepto de Mundo Pequeño describe el hecho que a pesar del gran tamaño de un sistema social, en la mayoría de las redes las distancias entre dos nodos son cortas

6

¿Es razonable que sea así?

En el mundo hay ~8mil millones de personas

¿Es razonable que cada uno de nosotros esté a menos de 6 grados de distancia de las demás personas en el mundo (en promedio)?



7

Si la red mundial fuese aleatoria: ¿Cuántas personas debería uno conocer para que d<6 a nivel mundial?

$$k^d = N$$

Cada nodo tiene *k* links. Por ende, en una red aleatoria probabilísticamente existen k^d nodos a una distancia *d*.

8

Si la red mundial fuese aleatoria: ¿Cuántas personas debería uno conocer para que d<6 a nivel mundial?

$$k^d = N$$
$$d = \log N / \log k$$

9

Si la red mundial fuese aleatoria: ¿Cuántas personas debería uno conocer para que d<6 a nivel mundial?

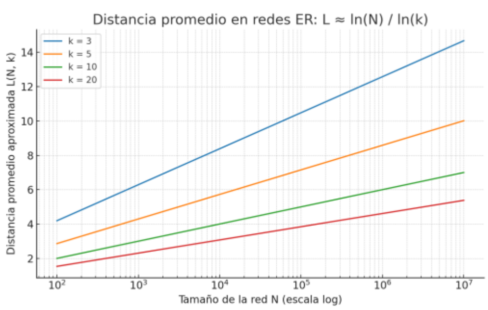
$$k^d = N$$

$$d = \log N / \log k$$

$$6 = \log 8\,000000000 / \log k$$

$$k \approx 45\ pers.$$

10



11

Ejemplo 1

¿Cuántos intermediarios son necesarios para enviar una carta desde la persona A a la persona B a través de conocidos? Milgram (1967)

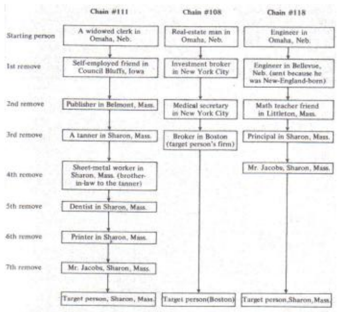
Experimento:

Inicio: Nebraska/Kansas,

Destino: Boston

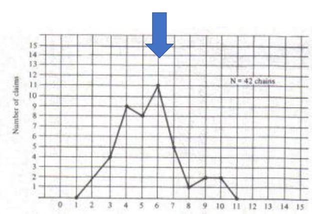
"Si no conoce al destinatario personalmente no trate de contactarlo, sino envíe la carta a una persona que usted conoce y que, a su juicio, tiene más probabilidades de conocerlo"

12



Ejemplos de lo que pasó

13



Stanley Milgram, 1967

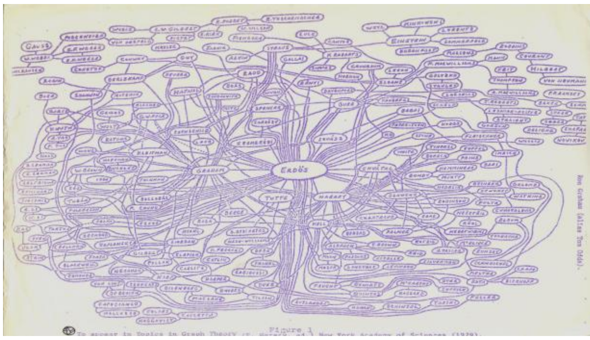
Promedio de intermediarios: 6

14

Ejemplo 2

Paul Erdös, matemático. Publicó 1500 artículos. La gran mayoría coautoreados

15



16

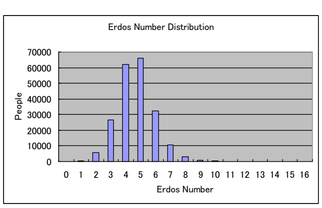
Algunos ejemplos de números de Erdös de grandes físicos

Albert Einstein	1921	2
Niels Bohr	1922	5
Erwin Schrödinger	1933	8
Enrico Fermi	1938	3
Richard P. Feynman	1965	4
Murray Gell-Mann	1969	3
Norman F. Ramsey	1989	3

17

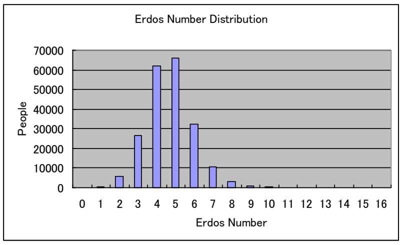
Número de Erdös

Número de Erdös 0 --- 1 persona
Número de Erdös 1 --- 504 personas
Número de Erdös 2 --- 6593 personas
Número de Erdös 3 --- 33605 personas
Número de Erdös 4 --- 83642 personas
Número de Erdös 5 --- 87760 personas
Número de Erdös 6 --- 40014 personas
Número de Erdös 7 --- 11591 personas
Número de Erdös 8 --- 3146 personas
Número de Erdös 9 --- 819 personas
Número de Erdös 10 --- 244 personas
Número de Erdös 11 --- 68 personas
Número de Erdös 12 --- 23 personas
Número de Erdös 13 --- 5 personas



Mediana = 5; Promedio = 4.69, Desviación = 1.27

18



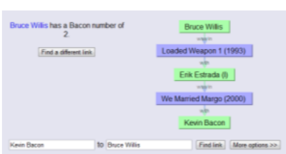
Mediana = 5; Promedio = 4.69, Desviación = 1.27

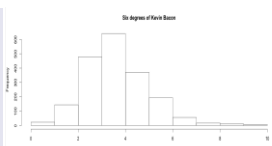
19

Ejemplo 3

El oráculo de Kevin Bacon

http://oracleofbacon.org/



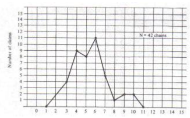


20

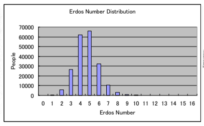
¿Es Kevin Bacon el actor con más conexiones?

Ranking	Nombre	Distancia promedio	# of películas	# of co-actores
1	Rod Steiger	2.537527	112	2562
2	Donald Pleasence	2.542376	180	2874
3	Martin Sheen	2.551210	136	3501
4	Christopher Lee	2.552497	201	2993
5	Robert Mitchum	2.557181	136	2905
6	Charlton Heston	2.566284	104	2552
7	Eddie Albert	2.567036	112	3333
8	Robert Vaughn	2.570193	126	2761
9	Donald Sutherland	2.577880	107	2865
10	John Gielgud	2.578980	122	2942
11	Anthony Quinn	2.579750	146	2978
12	James Earl Jones	2.584440	112	3787
…				
876	**Kevin Bacon**	**2.786981**	**46**	**1811**
…				

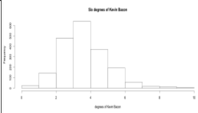
21



Milgram



Erdös



Bacon

En todos los ejemplos aparece el fenómeno de mundos pequeños

22

¿Podría el fenómeno de mundos pequeños ser resultado del puro azar?

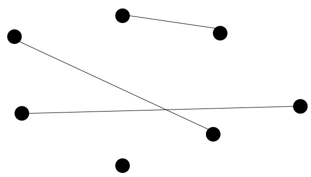
23

Random graph networks

24

Redes aleatorias

- Partir con *N* nodos y para cada par, con probabilidad *p*, añadir un link entre ellos.



25

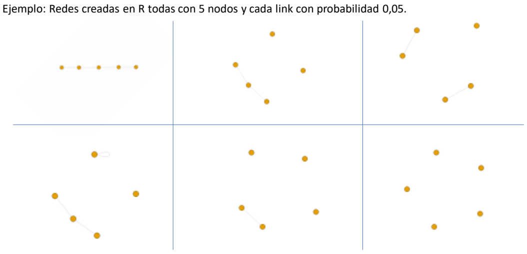
¿Qué propiedades tiene la red que hace posible el fenómeno de mundos pequeños?

Estas son todas redes de 4 nodos y 2 links. Total de posibles grafos: $\binom{\binom{n}{2}}{m}$ = 15

Por lo tanto, cada red *G(n,m)* es seleccionada con probabilidad 1/15

26

Random Graph Model:
Erdös-Renyi-Gilbert proponen *G(n,p)*, donde *p* es la probabilidad de existencia de cada *m*.
Ejemplo: Redes creadas en R todas con 5 nodos y cada link con probabilidad 0,05.



Por lo tanto G(n,p) no es exactamente lo mismo que *G(n,m)*, pero el primero se asemeja al segundo para grandes *n*

27

Asumimos que cada posible link *m* existe con una probabilidad *p* (iid)

La probabilidad *P(G)* de cada grafo de *n* nodos es:

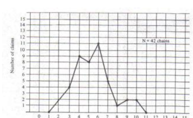
$$P(G) = p^m(1-p)^{\binom{n}{2}-m}$$

Y la probabilidad de un grafo de *m* links sigue una distribución binomial

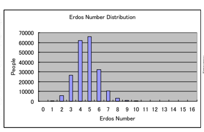
$$P(m) = \binom{\binom{n}{2}}{m} p^m(1-p)^{\binom{n}{2}-m}$$

El grado esperado $\langle m \rangle = (n-1)p$

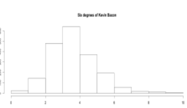
28



Milgram

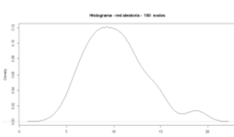


Erdös



Bacon

Red aleatoria de 100 nodos



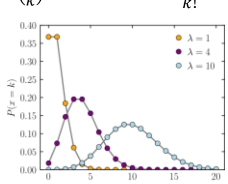
29

Si se mantiene constante el grado medio: $\lambda = \langle m \rangle = (n-1)p$

La probabilidad que un nodo tenga *k* grados es:

$$p_k = \binom{n}{k} p^k(1-p)^{n-k} \cong \frac{\lambda^k e^{-\lambda}}{k!}$$

Es decir, una Poisson



Ejemplos en R

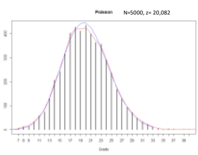
30

→ mundos pequeños

- Por ley de los grandes números a medida que aumenta *n*, una binomial puede aproximarse por una poisson. Tal que *z* es *q*n*

$$Binominal: p(k) = \binom{n-1}{k} q^k (1-q)^{n-k-1}$$

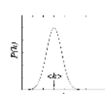
$$Poisson: p(k) = \frac{1}{k!} e^{-z} z^k$$



31

→ mundos pequeños

- Muchas propiedades de las redes pueden ser explicadas con este simple modelo.
- Por ejemplo, un componente gigante surge cuando el grado promedio es mayor de 1
- Sea λ el grado promedio
  - Si λ < 1 → islas
  - Si λ >1 → componente gigante
- Distancia promedio proporcional a $k \sim \ln N$



32

Percolación → emerge un componente gigante. Sea *S* la fracción del grafo ocupado por el componente gigante

$$S = 1 - e^{-\lambda S}$$

Proporción no conectada al componente gigante

Si λ < 1, S = 0
Si $\lambda \geq$ 1, S > 0

No linearidad
Transición de fase

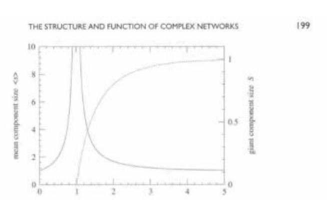


Fig. 4.1 *The mean component size (solid line), excluding the giant component if there is one, and the giant component size (dotted line), for the Poisson random graph, (4.3) and (4.4).*

Ver Newman 2004

33

Modelos aleatorios generalizados

- En Erdös-Renyi la probabiblidad es igual para cada conexión, pero en ocasiones queremos ver redes aleatorias con una distribución de grado específica.
- Son la base para los "modelos nulos"
- En la literatura lo verán como General Random Graph Models

Ejemplos en R

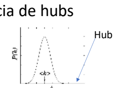
34

→ mundos pequeños… PERO

- Pero los niveles de clusterización de las redes reales son mayores a los que predice una red aleatoria (Erdös-Renyi)

```
> resumen.distancias
                    real aleatorio
distancia empresas  2.865     3.081
clustering empresas 0.401     0.051
```

- Y otras cosas no coinciden. Por ejemplo: existencia de hubs



35

Small world networks

36

Strogatz y Watts (1998): Fenómeno Small World

- Homofilia + Weak links

$$C(v) = \frac{\text{\# de links entre vecinos}}{N(N-1)/2}$$

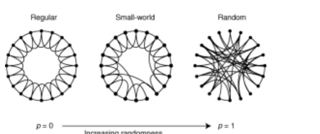
Table 1 Empirical examples of small-world networks

	L_{actual}	L_{random}	C_{actual}	C_{random}
Film actors	3.65	2.99	0.79	0.00027
Power grid	18.7	12.4	0.080	0.005
C. elegans	2.65	2.25	0.28	0.05

37

Strogatz y Watts (1998): Fenómeno Small World

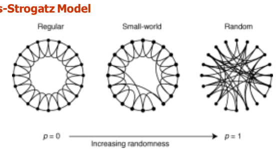
– Alto clustering, diámetro corto

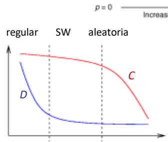


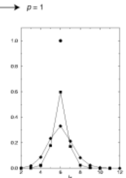
Fuente: Watts y Strogatz (1998)

38

Watts-Strogatz Model





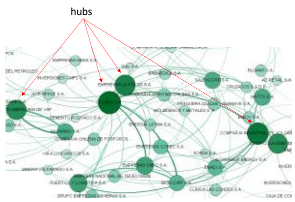


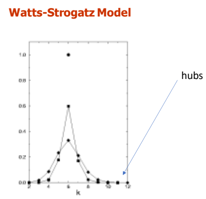
C(p) : clustering coef
D(p) : Distancia promedio

39

Ejemplos en R

Pero, el problema con los hubs persiste.



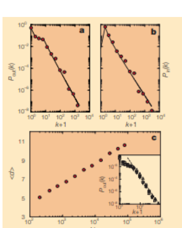


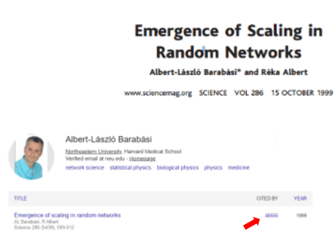
40

Scale free netwoks

41

Preferential Attachment





42

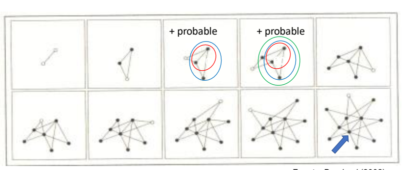
Preferencial Attachment

$P(k) \propto k^{-\alpha} \quad con\ \alpha$>1

43

Preferencial Attachment

Probabilidad de conectarse con un nodo más popular es mayor



Fuente: Barabasi (2002)

44

Comparando los tres modelos

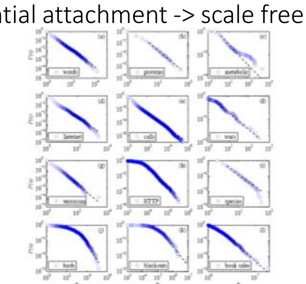
```
> resumen
     distancia clustering
Real     2.865      3.982
ER       0.401      0.295
SW       4.525      3.571
PA       0.010      0.025
```

45

Omnipresencia de las redes scale free?

46

Preferential attachment -> scale free

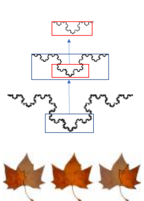


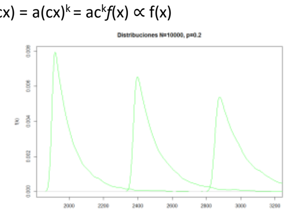
Clauset et al 2009

47

Preferential attachment -> scale free

Sean a y c constantes: $f(cx) = a(cx)^k = ac^k f(x) \propto f(x)$

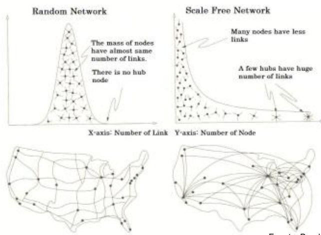




Cambiando la escala por un factor c=1.25 y luego por c=1.5

48

Preferencial Attachment

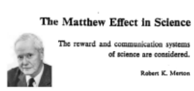


Fuente: Barabasi (2002)

49

Preferencial Attachment

- Implicancias:
  1. Emergen los hubs con mayor probabilidad que bajo un esquema aleatorio
  2. El “rico” es cada vez más “rico” (efecto “matthews”)


The Matthew Effect in Science

The reward and communication systems of science are considered.

Robert K. Merton

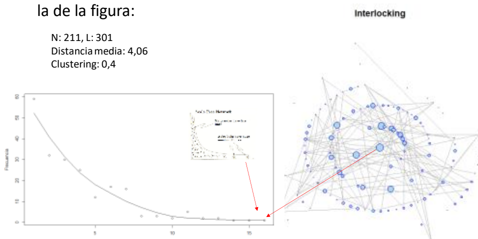
(Science 1968)

50

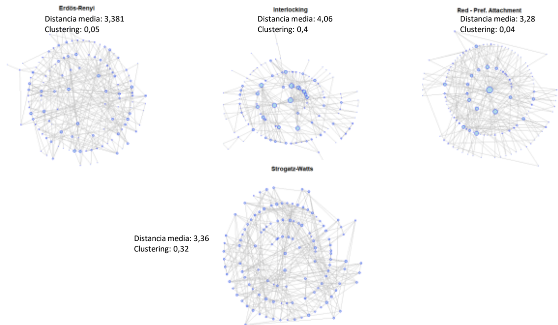
- Consideremos una red real como la de la figura:

N: 211, L: 301
Distancia media: 4,06
Clustering: 0,4

Interlocking



51



52

+ sobre preferential attachment

Esta afirmación se ve una y otra vez en los últimos 20-25 años.

"scale-free networks represent
a foundational idea
in network science"

53

Power-Law Distributions in Empirical Data

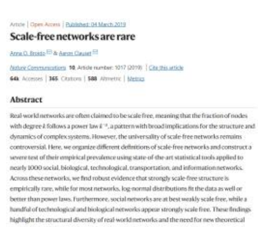
Authors: Aaron Clauset, Cosma Rohilla Shalizi, and M. E. J. Newman | AUTHORS INFO & AFFILIATIONS

https://doi.org/10.1137/070710111

GET ACCESS | BibTeX | Tools

Abstract

Power-law distributions occur in many situations of scientific interest and have significant consequences for our understanding of natural and man-made phenomena. Unfortunately, the detection and characterization of power laws is complicated by the large fluctuations that occur in the tail of the distribution—the part of the distribution representing large but rare events—and by the difficulty of identifying the range over which power-law behavior holds. Commonly used methods for analyzing power-law data, such as least-squares fitting, can produce substantially inaccurate estimates of parameters for power-law distributions, and even in cases where such methods return accurate answers they are still unsatisfactory because they give no indication of whether the data obey a power law at all. Here we present a principled statistical framework for discerning and quantifying power-law behavior in empirical data. Our approach combines maximum-likelihood fitting methods with goodness-of-fit tests based on the Kolmogorov–Smirnov (KS) statistic and likelihood ratios. We evaluate the effectiveness of the approach with tests on synthetic data and give critical comparisons to previous approaches. We also apply the proposed methods to twenty-four real-world data sets from a range of different disciplines, each of which has been conjectured to follow a power-law distribution. In some cases we find these conjectures to be consistent with the data, while in others the power law is ruled out.



Article | Open Access | Published: 04 March 2019

Scale-free networks are rare

Anna D. Broido & Aaron Clauset

Nature Communications 10, Article number: 1017 (2019) | Cite this article

64k Accesses | 365 Citations | 588 Altmetric | Metrics

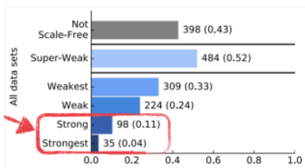
Abstract

Real-world networks are often claimed to be scale free, meaning that the fraction of nodes with degree k follows a power law $k^{-\alpha}$, a pattern with broad implications for the structure and dynamics of complex systems. However, the universality of scale-free networks remains controversial. Here, we organize different definitions of scale-free networks and construct a severe test of their empirical prevalence using state-of-the-art statistical tools applied to nearly 1000 social, biological, technological, transportation, and information networks. Across these networks, we find robust evidence that strongly scale-free structure is empirically rare, while for most networks, log-normal distributions fit the data as well or better than power laws. Furthermore, social networks are at best weakly scale free, while a handful of technological and biological networks appear strongly scale free. These findings highlight the structural diversity of real-world networks and the need for new theoretical explanations of these non-scale-free patterns.

54

+ sobre preferential attachment

- Controversia reciente que pone en duda la universalidad del fenómeno



Fuente: Clauset 2018 at SFI

55

+ sobre preferential attachment

- Y en las redes sociales ninguna pasaría el test (!!)

Fuente: Clauset 2018 at SFI

56

+ sobre preferential attachment

- ¿Y qué dice Barabasi?

→https://www.barabasilab.com/post/love-is-all-you-need

“***they fail to see the scale-free property because they invent a new criterion of weak and strong scale-free networks***”

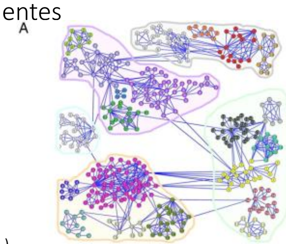
Barabasi, 6/3/2018

Ejemplos en R

57

Exploraciones más recientes

- Las redes reales a menudo no pueden describirse de manera adecuada mediante un único modelo estructural (sea E-R, SW, PA u otra)
- Auge de modelos multicapa o hypergraphs
  - Escalas diferentes en la misma red
  - Estos modelos generan distribuciones alternativas más flexibles que los modelos scale-free (log-normal,exponencial truncada,…)



Delvenne, Yaliraki y Barahona (2010)

58