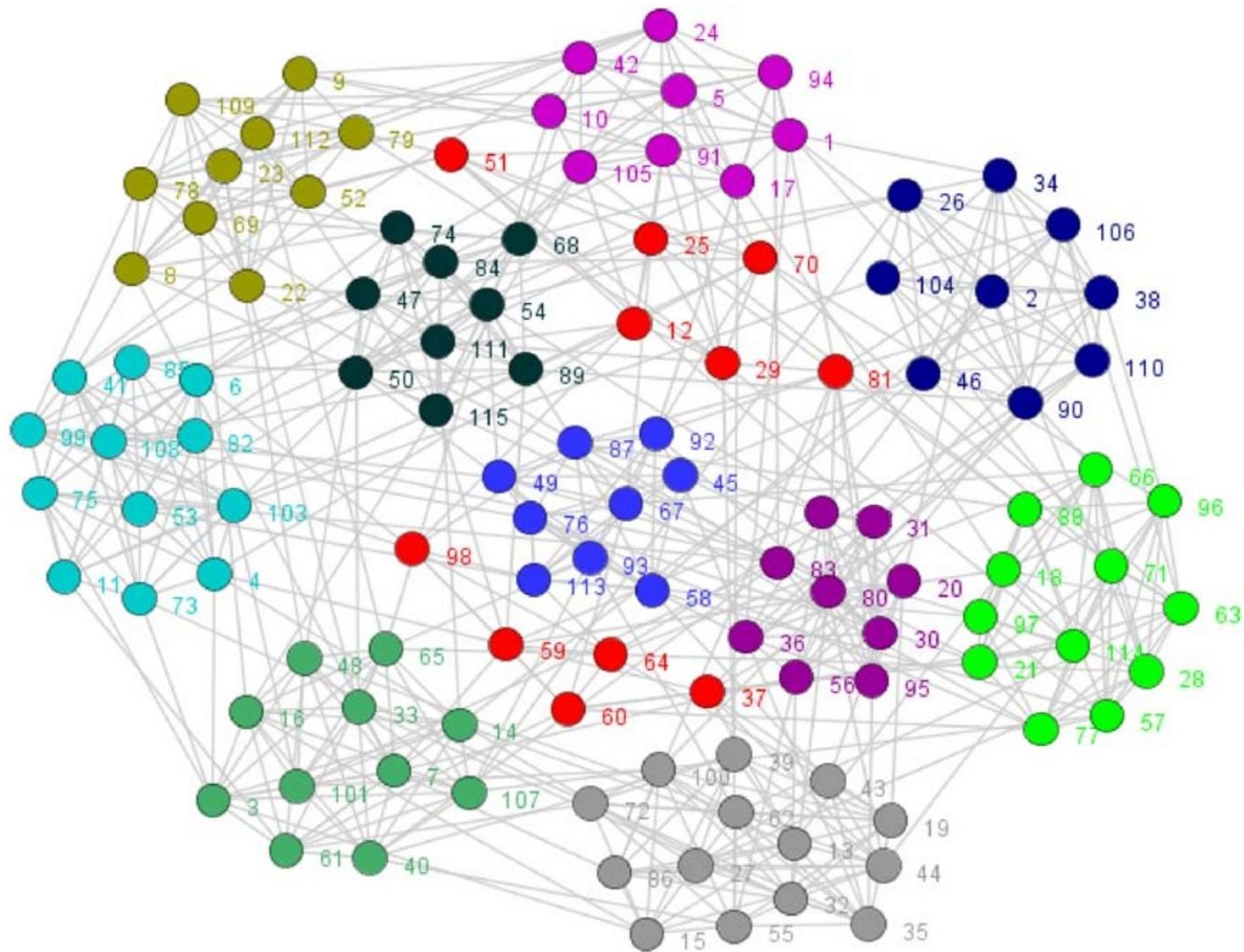


Networks

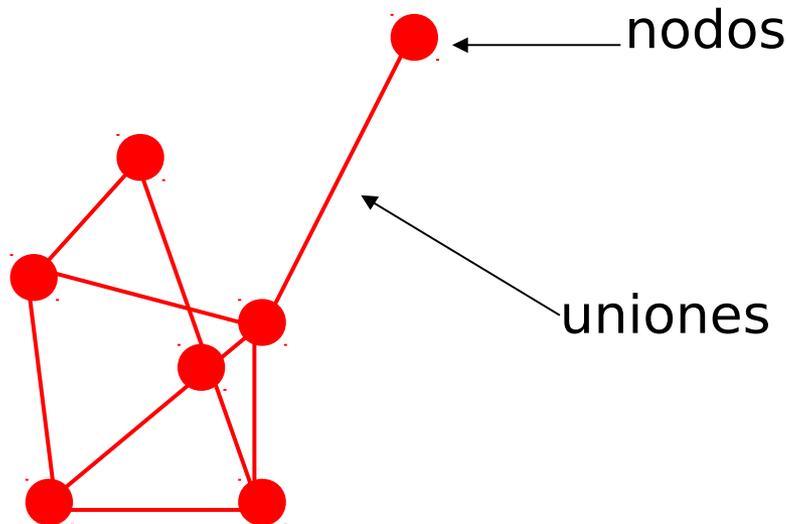
侍



~

Que es una red?

Una red esta compuesta por



Entonces, una red es un conjunto de n nodos unidos por M uniones

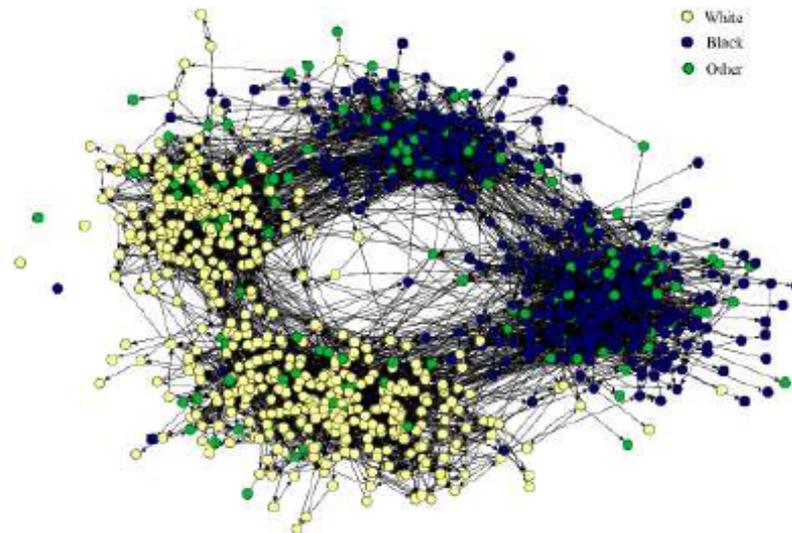
$$M \leq \frac{n(n-1)}{2} = \begin{array}{|c|} \hline n \\ \hline 2 \\ \hline \end{array}$$

Comunidades

Estudiaremos esto en profundidad en el futuro pero definimos (no estricto) que :

la comunidades de un grapho son los subgraphos cuyos nodos estan mas "unidos" entre si que con los nodos pertenecientes a otras comunidades"

La presencia de comunidades da lugar a la *modularidad*



Caminos y conectividad

Una de las características interesantes de una red es la accesibilidad de los nodos.

Sea un camino $P_{i_0 i_n}$ definido por

Consideramos redes conexas.

$$G = (\mathcal{V}, \mathcal{E})$$

$n + 1$ nodos

$$\mathcal{V}_p = \{i_0, i_1, \dots, i_n\}$$

n links

$$\mathcal{E}_p = [(i_0, i_1), (i_1, i_2), \dots, (i_{n-1}, i_n)]$$

El camino $P_{i_0 i_n}$ conecta entonces i_0 con i_n por n pasos y esa es la longitud del camino.

El número de caminos de longitud n entre dos nodos dados, i y j es la potencia n -ésima de la matriz de adyacencias.

$$N_{ij} = (\mathbf{X}^n)_{ij}$$

Dos componentes $G_1 = (V_1, E_1)$; $G_2 = (V_2, E_2)$ de una red están **desconectadas** si es imposible construir un camino $P_{i_1 i_2}$ con $i_1 \in G_1$ y $i_2 \in G_2$

Distancia entre dos nodos : longitud del camino mínimo l_{ij} , que los unen.

Entonces un grafo puede ser representado por su matriz de adyacencias $\|A\|$, donde

$$A_{ij} = 1 \quad \text{si est\u00e1n conectados}$$

$$A_{ij} = 0 \quad \text{en otro caso}$$

$$A_{ij} = \alpha_{ij} \quad \text{conectados pesados}$$

Para no dirigidos $A_{ij} = A_{ji}$

Propiedades fundamentales de los complex networks

Mundo pequeño

(el mundo es un pañuelo)

Experimento de Milgram (1967):

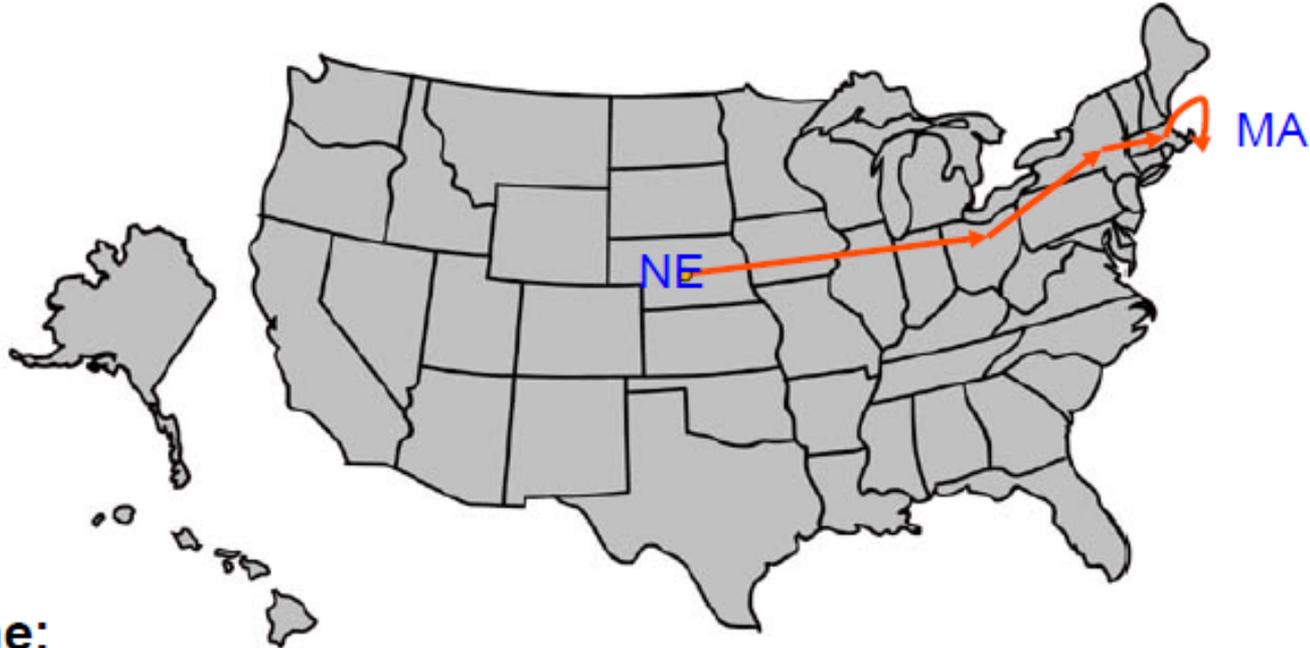
Envió paquetes a personas elegidas al azar en el medio oeste De USA y les pidió que se los hiciesen llegar a personas en Boston Pero debían hacerlo a través de personas que conociesen por el Nombre (debían llegar por amigos de amigos de amigos...).

Como resultado de este experimento es que en valor medio Eran suficientes “6 amigos”, para alcanzar el éxito.

Esto fue el inicio de una enorme serie de trabajos en el tema.

Instructions:

Given a target individual (stockbroker in Boston), pass the message to a person you correspond with who is “closest” to the target.



Outcome:

20% of initiated chains reached target

average chain length = 6.5

← “Six degrees of separation”

Herramientas para estudiar el efecto de mundo pequeño

Distancia mínima promedio en una red :

Suma sobre todos
Los pares de nodos

$$l(G) = \frac{1}{[n(n-1)]/2} \sum_{i \leq j \in G} d_{ij}$$

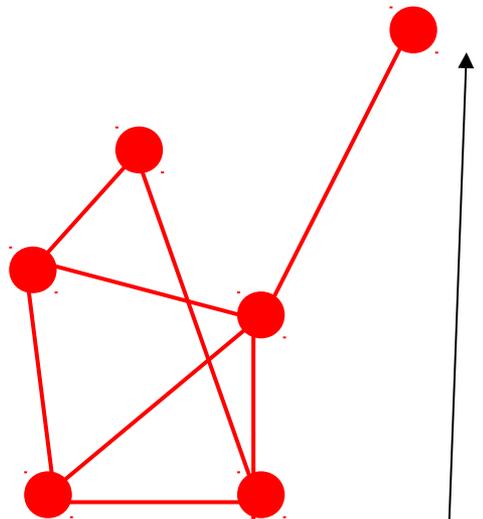
d_{ij} es el camino mínimo entre los nodos i y j

$[n(n-1)]/2$ es el numero pares de nodos para n nodos

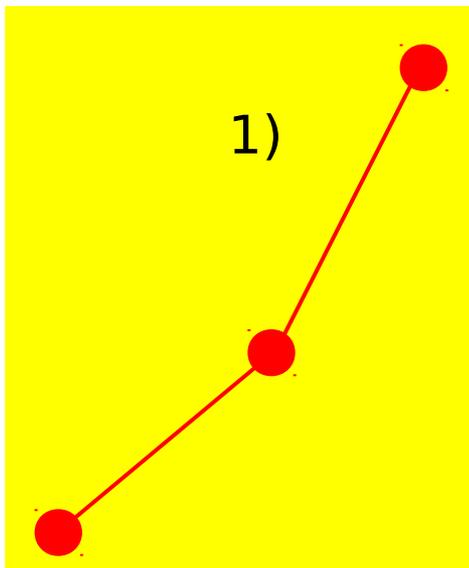
Por ejemplo : d_{ij}

i

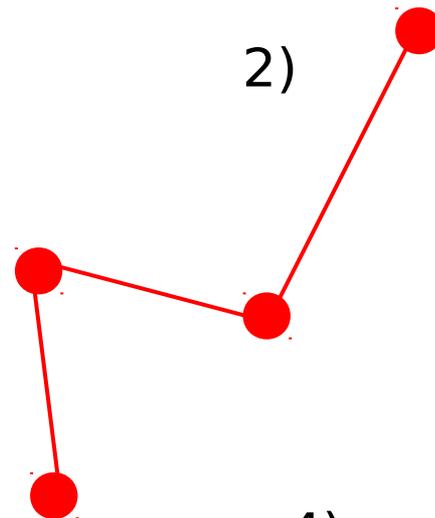
j



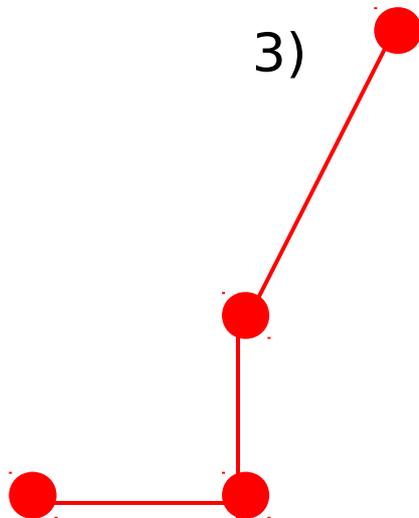
Nodos de
interes



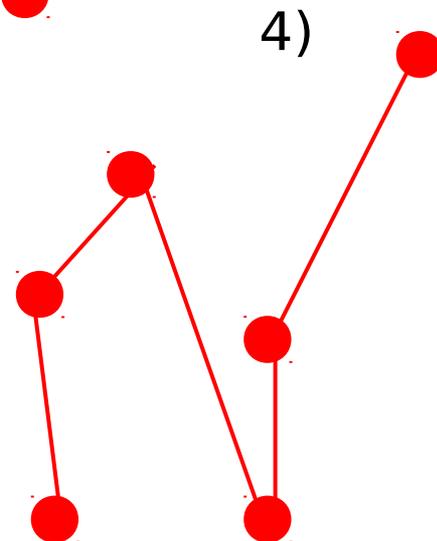
1)



2)



3)



4)

etc.

$$l = \frac{1}{\frac{1}{2}n(n-1)} \sum_{i \geq j} d_{ij}$$

Donde d_{ij} es la distancia minima (en terminos del numero de pasos) entre los nodos i y j . O sea la geodesica entre esos dos puntos.(ojo se incluye la distancia a si mismo que es 0)

Esto esta bien definido si trabajamos con networks conexos, pero si no hacemos un analisis preliminar de clusters podrian aparecer infinitos

Quizas entonces es mejor usar

$$l^{-1} = \frac{1}{\frac{1}{2}n(n+1)} \sum_{i \geq j} d_{ij}^{-1}$$

Clusterización

$$C(G) = \frac{1}{n} \sum_i C_i$$

Donde C_i responde la siguiente pregunta

Que fracción de mis amigos son amigos entre si?

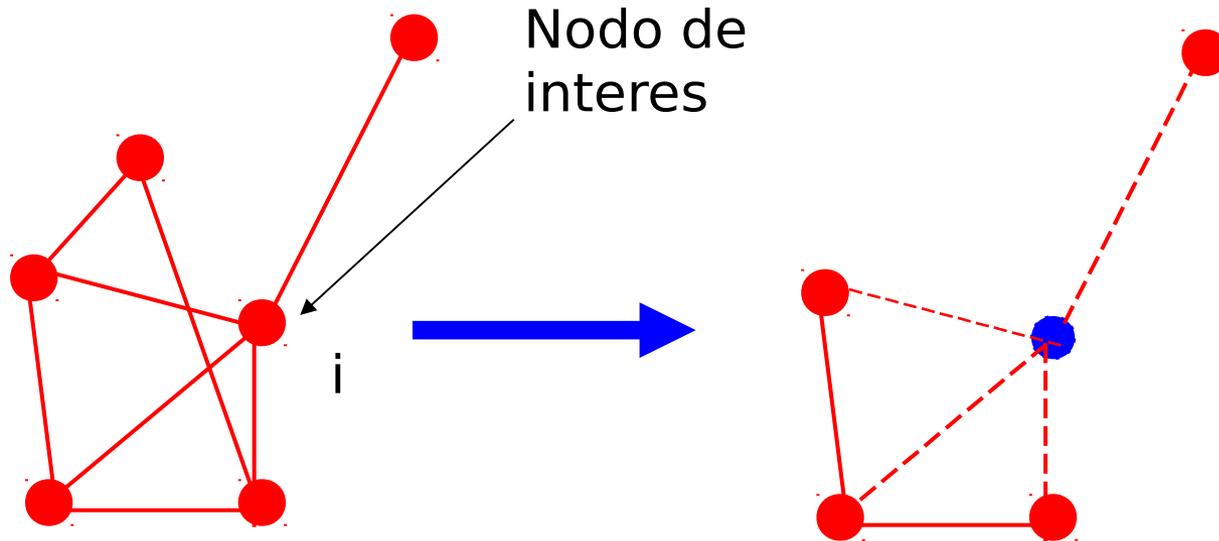
$$C_i = \frac{E_i}{k_i(k_i - 1) / 2}$$

E_i es el número de uniones entre los k_i vecinos de i

Número de pares de nodos con k_i el número de nodos vecinos de i

Si todos mis amigos son amigos entre si entonces $C_i=1$

Por ejemplo, para calcular C_i



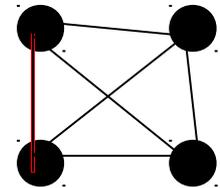
Numero de nodos vecinos al de interes : 4

Numero máximo de uniones de 4 nodos : 6

Uniones presentes

: 2

→ $C_i = 2/6$



Random graph

Un random graph es un grfo que se construye del siguiente modo
(modelo de Renyi)

Dados N nodos rotulados, existen $N_t = N(N - 1)/2$ posibles uniones.

Si el random graph tiene n nodos , estos se eligen al azar de los N_t .

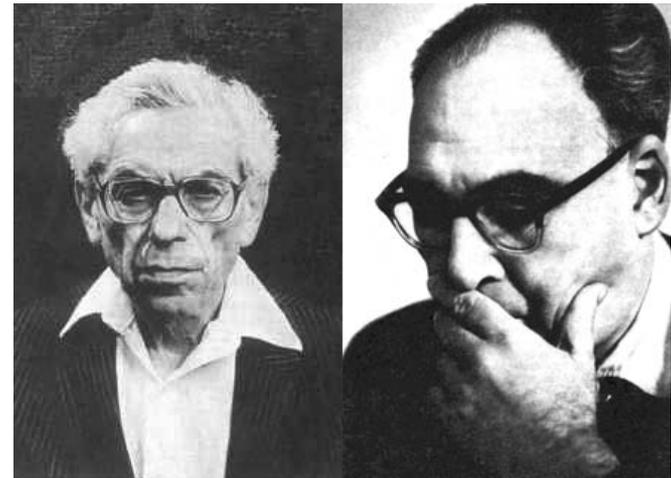
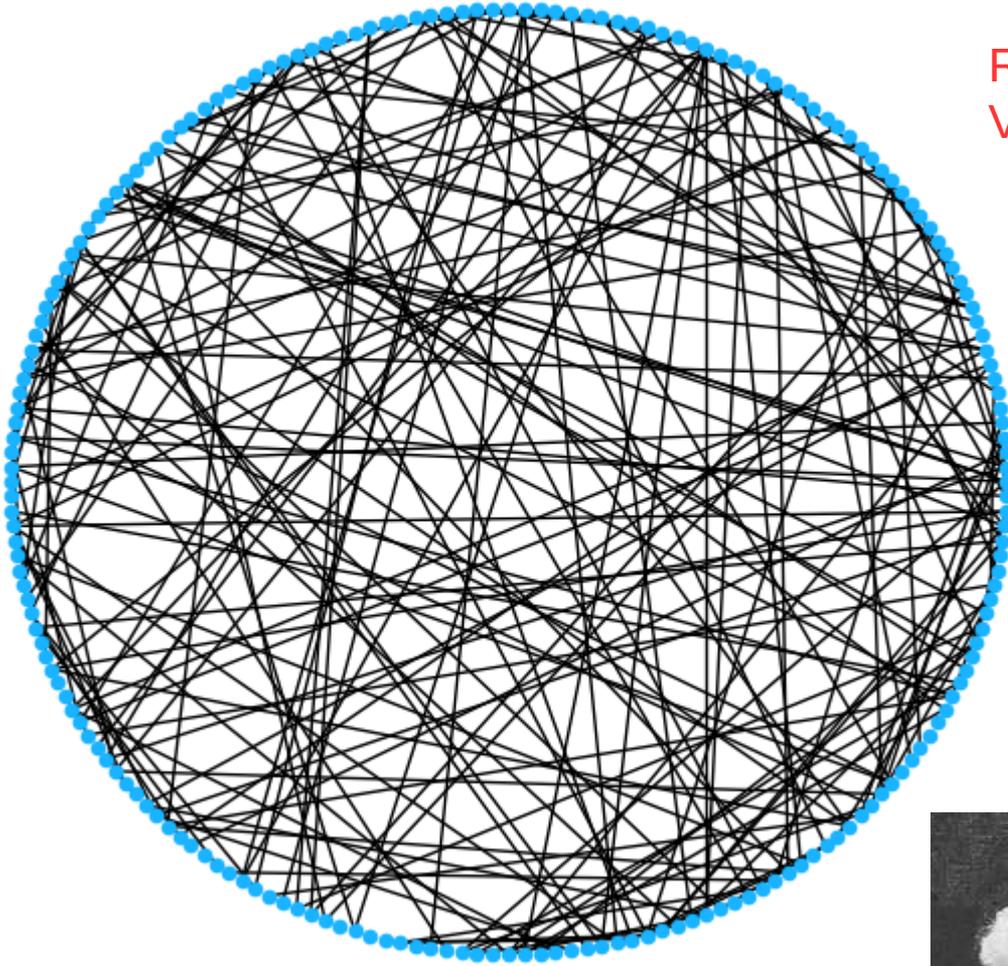
Otra posible definicion es la del grafo binomial

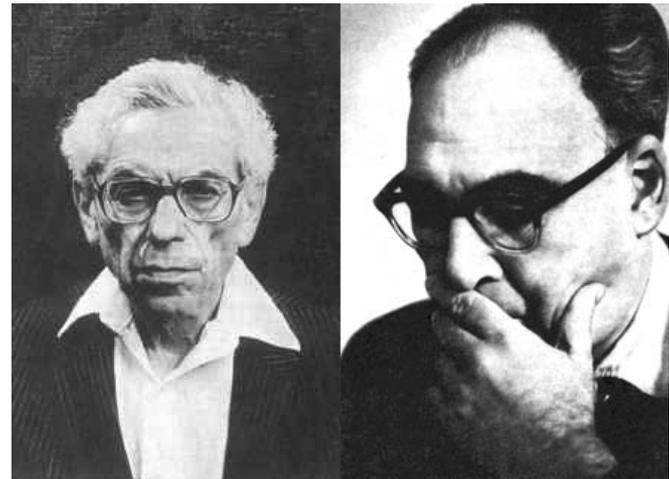
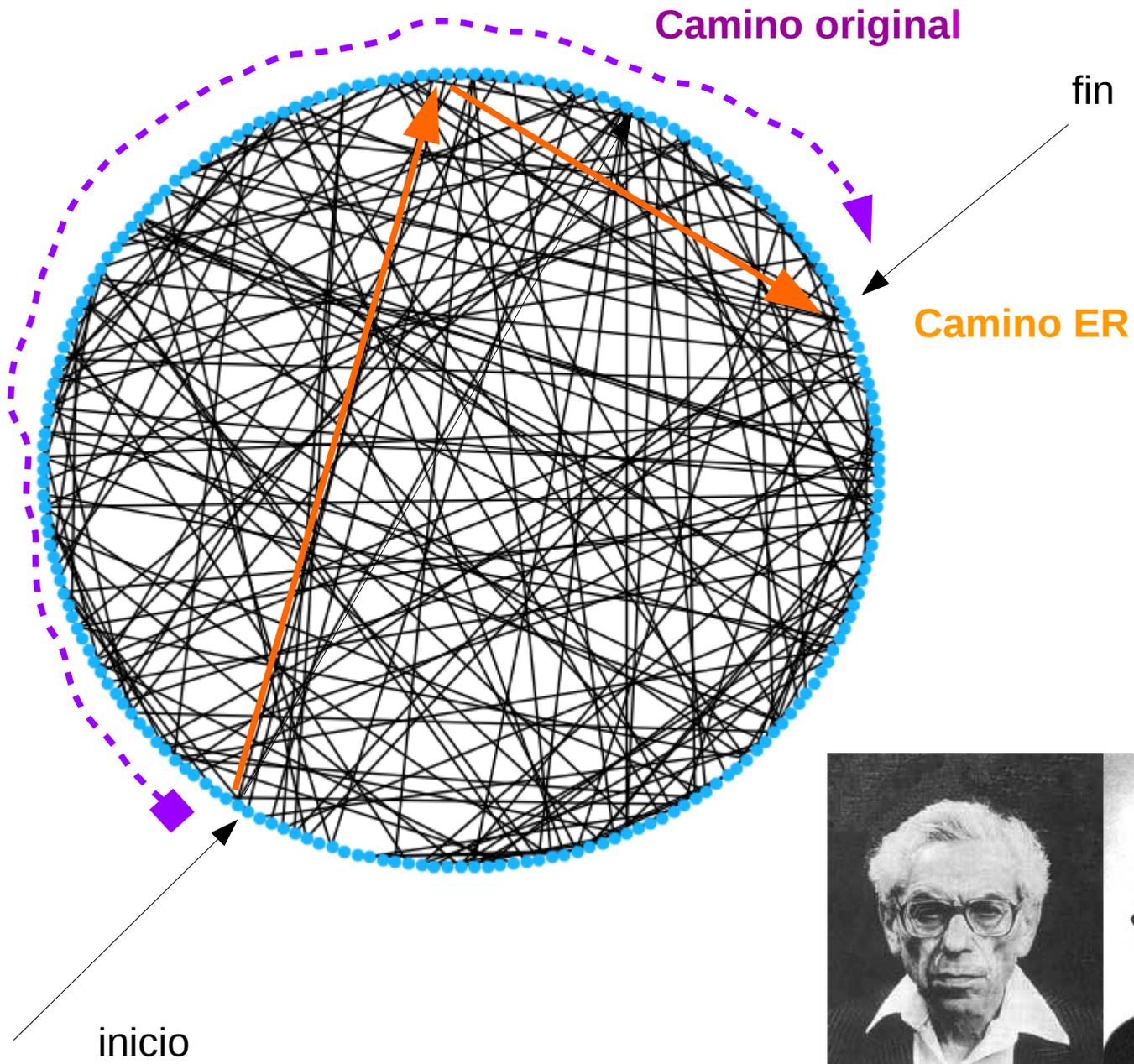
Empezamos con N nodos y luego tomamos los pares y los unimos con probabilidad p

(atención con la conectividad)

Tomamos el random graph como referencia
y lo comparamos con los reales

Random de Eordos Renyi + links a Vecinos inmediatos





Aquí se ve facil

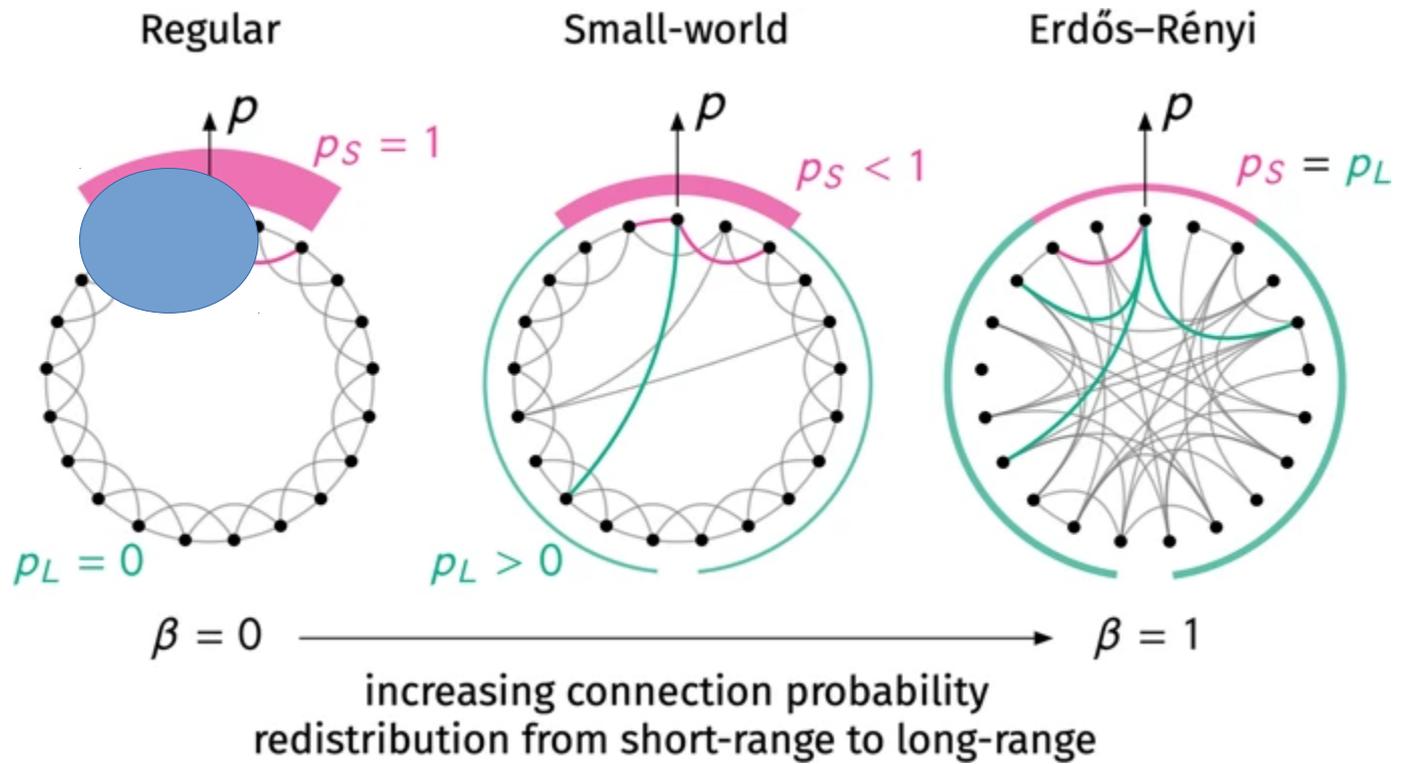


TABLE I. The general characteristics of several real networks. For each network we have indicated the number of nodes, the average degree $\langle k \rangle$, the average path length ℓ , and the clustering coefficient C . For a comparison we have included the average path length ℓ_{rand} and clustering coefficient C_{rand} of a random graph of the same size and average degree. The numbers in the last column are keyed to the symbols in Figs. 8 and 9.

Network	Size	$\langle k \rangle$	ℓ	ℓ_{rand}	C	C_{rand}	Reference	Nr.
WWW, site level, undir.	153 127	35.21	3.1	3.35	0.1078	0.00023	Adamic, 1999	1
Internet, domain level	3015–6209	3.52–4.11	3.7–3.76	6.36–6.18	0.18–0.3	0.001	Yook <i>et al.</i> , 2001a, Pastor-Satorras <i>et al.</i> , 2001	2
Movie actors	225 226	61	3.65	2.99	0.79	0.00027	Watts and Strogatz, 1998	3
LANL co-authorship	52 909	9.7	5.9	4.79	0.43	1.8×10^{-4}	Newman, 2001a, 2001b, 2001c	4
MEDLINE co-authorship	1 520 251	18.1	4.6	4.91	0.066	1.1×10^{-5}	Newman, 2001a, 2001b, 2001c	5
SPIRES co-authorship	56 627	173	4.0	2.12	0.726	0.003	Newman, 2001a, 2001b, 2001c	6
NCSTRL co-authorship	11 994	3.59	9.7	7.34	0.496	3×10^{-4}	Newman, 2001a, 2001b, 2001c	7
Math. co-authorship	70 975	3.9	9.5	8.2	0.59	5.4×10^{-5}	Barabási <i>et al.</i> , 2001	8
Neurosci. co-authorship	209 293	11.5	6	5.01	0.76	5.5×10^{-5}	Barabási <i>et al.</i> , 2001	9
<i>E. coli</i> , substrate graph	282	7.35	2.9	3.04	0.32	0.026	Wagner and Fell, 2000	10
<i>E. coli</i> , reaction graph	315	28.3	2.62	1.98	0.59	0.09	Wagner and Fell, 2000	11
Ythan estuary food web	134	8.7	2.43	2.26	0.22	0.06	Montoya and Solé, 2000	12
Silwood Park food web	154	4.75	3.40	3.23	0.15	0.03	Montoya and Solé, 2000	13
Words, co-occurrence	460.902	70.13	2.67	3.03	0.437	0.0001	Ferrer i Cancho and Solé, 2001	14
Words, synonyms	22 311	13.48	4.5	3.84	0.7	0.0006	Yook <i>et al.</i> , 2001b	15
Power grid	4941	2.67	18.7	12.4	0.08	0.005	Watts and Strogatz, 1998	16
<i>C. Elegans</i>	282	14	2.65	2.25	0.28	0.05	Watts and Strogatz, 1998	17

Observamos que

a) los caminos mínimos medios de los networks reales son del orden de los random

b) la clusterización de los reales es mucho mayor que las de los random.

TABLE I. The general characteristics of several real networks. For each network we have indicated the average degree $\langle k \rangle$, the average path length ℓ , and the clustering coefficient C . For a comparison we also show the average path length ℓ_{rand} and clustering coefficient C_{rand} of a random graph of the same size and average degree. The first two columns are keyed to the symbols in Figs. 8 and 9.

Network	Size	$\langle k \rangle$	ℓ	ℓ_{rand}	C	C_{rand}	
WWW, site level, undir.	153 127	35.21	3.1	3.35	0.1078	0.00023	
Internet, domain level	3015–6209	3.52–4.11	3.7–3.76	6.36–6.18	0.18–0.3	0.001	
Movie actors	225 226	61	3.65	2.99	0.79	0.00027	
LANL co-authorship	52 909	9.7	5.9	4.79	0.43	1.8×10^{-4}	Ne
MEDLINE co-authorship	1 520 251	18.1	4.6	4.91	0.066	1.1×10^{-5}	Ne
SPIRES co-authorship	56 627	173	4.0	2.12	0.726	0.003	Ne
NCSTRL co-authorship	11 994	3.59	9.7	7.34	0.496	3×10^{-4}	Ne
Math. co-authorship	70 975	3.9	9.5	8.2	0.59	5.4×10^{-5}	
Neurosci. co-authorship	209 293	11.5	6	5.01	0.76	5.5×10^{-5}	
<i>E. coli</i> , substrate graph	282	7.35	2.9	3.04	0.32	0.026	
<i>E. coli</i> , reaction graph	315	28.3	2.62	1.98	0.59	0.09	
Ythan estuary food web	134	8.7	2.43	2.26	0.22	0.06	
Silwood Park food web	154	4.75	3.40	3.23	0.15	0.03	
Words, co-occurrence	460.902	70.13	2.67	3.03	0.437	0.0001	Fer
Words, synonyms	22 311	13.48	4.5	3.84	0.7	0.0006	
Power grid	4941	2.67	18.7	12.4	0.08	0.005	
<i>C. Elegans</i>	282	14	2.65	2.25	0.28	0.05	

ario grandinetti tiene un Bacon number de 3.

ario Grandinetti actuó en Amore ritorna, L' (2004) con Fabrizio Bentivoglio

fabrizio Bentivoglio actuó en Apartment Zero (1988) con Colin Firth

colin Firth actuó en Where the Truth Lies (2005) con Kevin Bacon

sana gimenez tiene Bacon number de 3.

sana Gimenez actuó en A los cirujanos se les va la mano (1980) con

ria Andreoli

ria Andreoli actuó en Apartment Zero (1988) con Colin Firth

lin Firth actuó en Where the Truth Lies (2005) con Kevin Bacon

jorge lanata tiene Bacon number de 3.

Jorge Lanata actuó en Lado oscuro del corazon, El (1992) con

Sandra Ballesteros

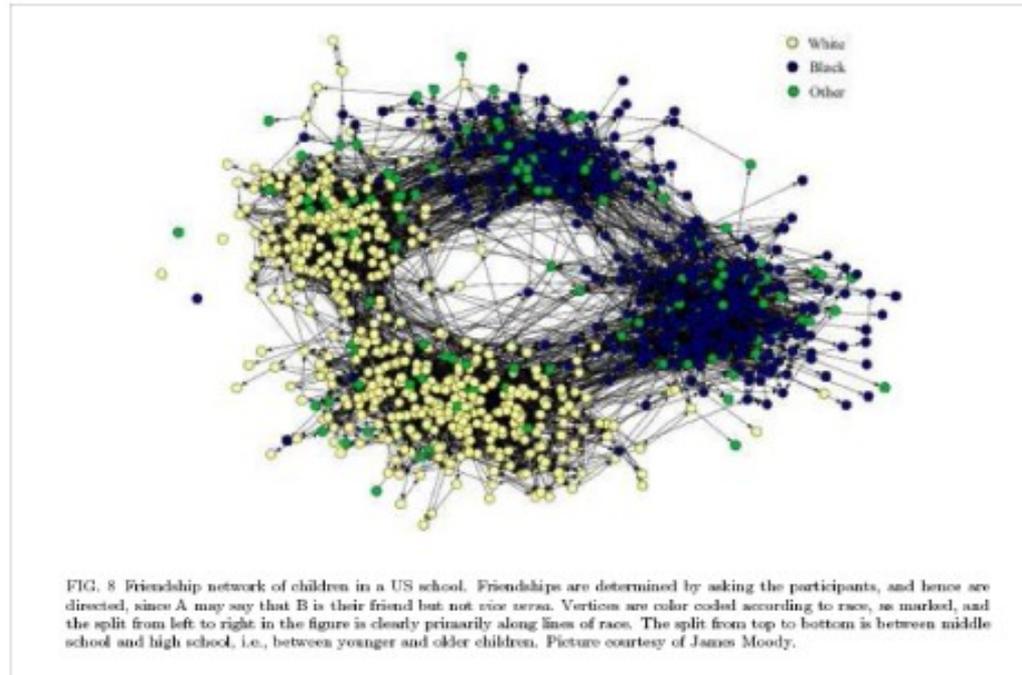
Sandra Ballesteros actuó en Sekuestro, El (1997) con Jose Ramon

Rosario

Jose Ramon Rosario actuó en Mystic River (2003) con Kevin Bacon

e) Estructura de Comunidades

definición intuitiva de comunidad → un subconjunto de nodos (subgrafo) tal que los nodos en el subgrafo están más unidos entre sí que con el resto

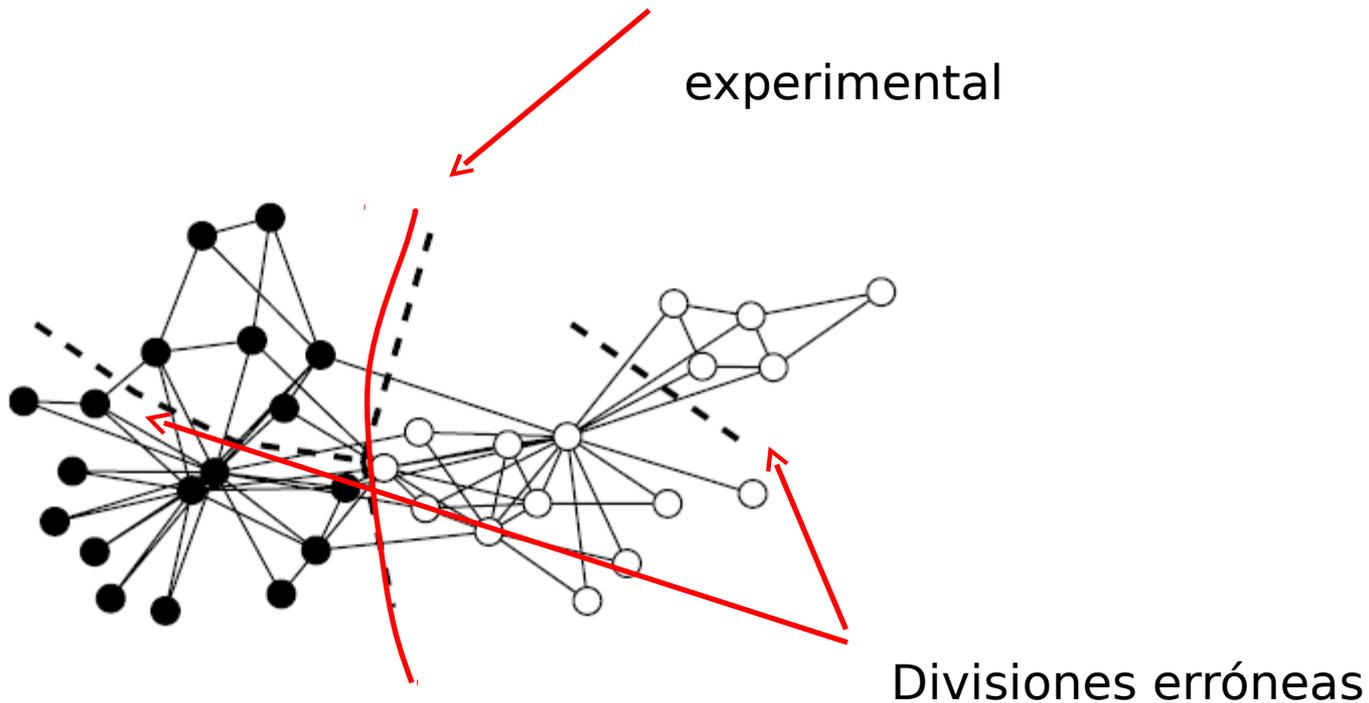


Comunidad

Definición de Comunidad (intuitiva)

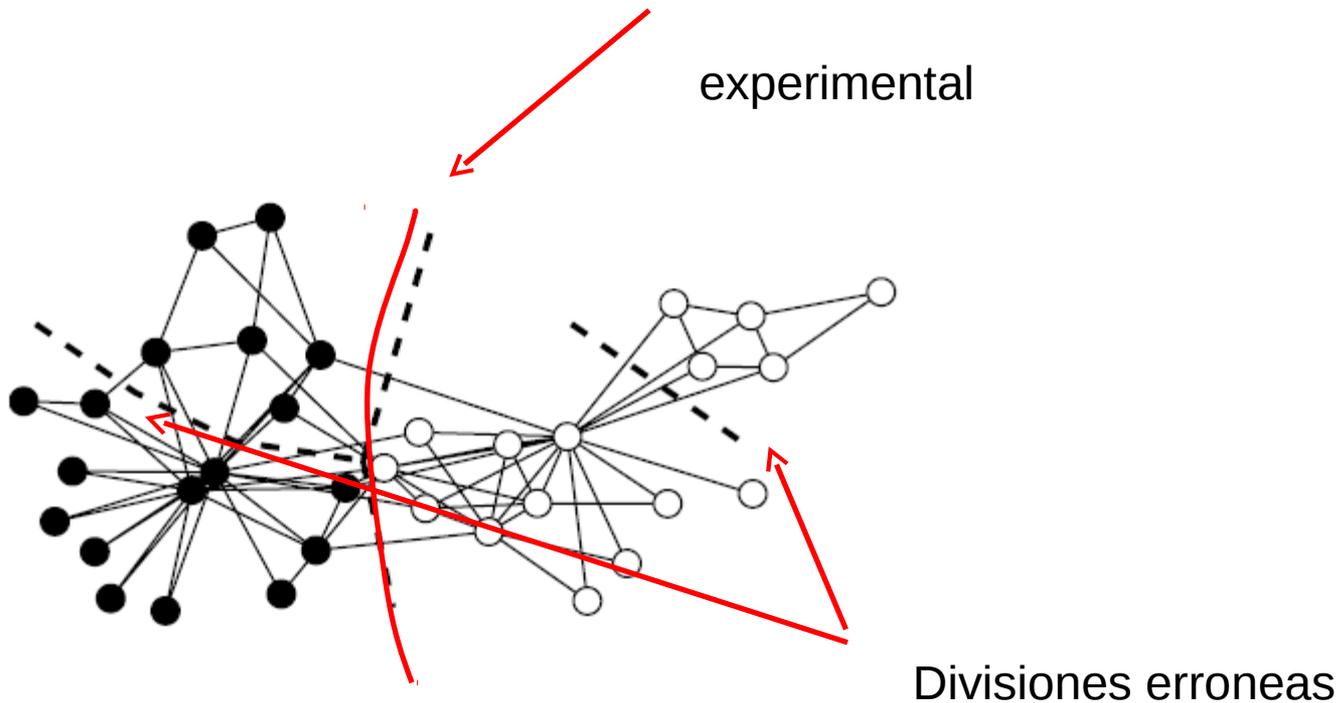
Dado un conjunto de nodos y links que constituyen un grafo conexo

Descomponerlo en comunidades es encontrar aquellos subconjuntos de nodos mas unidos entre si que con respecto al resto del grafo)



Comunidad

Resolver un problema de comunas es :
dar definición de comuna
dar algoritmo para resolverlo



Respecto de las comunidades

El análisis de Newman

Sea

$$A_{v\omega} = \begin{cases} 1 & \text{si los vertices } v \text{ y } \omega \text{ estan conectados} \\ 0 & \text{en otro caso} \end{cases}$$

links

La fracción de lados que pertenecen a las mismas comunidades es

$$\frac{\sum_{v\omega} A_{v\omega} \delta(c_v, c_\omega)}{\sum_{v\omega} A_{v\omega}} = \frac{1}{2m} \sum_{v\omega} A_{v\omega} \delta(c_v, c_\omega)$$

Por otro lado

$$k_v = \sum_{\omega} A_{v\omega}$$

$$k_v = \sum_{\omega} A_{v\omega}$$

son los links que salen de el nodo v (tomar en cuenta que $A_{vv} = 0$)

$$k_v = \sum_{\omega} A_{v\omega}$$

Define ahora Q como

$$Q = \frac{1}{2m} \sum_{v\omega} \left[A_{v\omega} - \frac{k_v k_{\omega}}{2m} \right] \delta(c_v, c_{\omega})$$

o tambien

$$Q = \sum_{v\omega} \left[\frac{A_{v\omega}}{2m} - \frac{k_v \cdot k_{\omega}}{2m \cdot 2m} \right] \delta(c_v, c_{\omega})$$

donde el termino $\frac{k_v}{2m}$ es la fraccion de nodos que salen del nodo v o sea

$$\frac{k_v}{2m} = \frac{\sum_{\omega} A_{v\omega}}{\sum_{v\omega} A_{v\omega}}$$

Esto deberia ser la probabilidad de que un link al azar salga del nodo v entonces la proba de que un nodo salga de v y llegue a ω es

$$P_{v\omega} = \frac{k_v}{m} \cdot \frac{k_{\omega}}{m}$$

Escribe ahora $\delta(c_v, c_\omega) = \sum_i \delta(c_v, i) \delta(c_\omega, i)$

$$Q = \frac{1}{2m} \sum_{v\omega} \left[A_{v\omega} - \frac{k_v k_\omega}{2m} \right] \sum_i \delta(c_v, i) \delta(c_\omega, i)$$

operando con esto es

$$Q = \sum_i \left\{ \frac{1}{2m} \sum_{v\omega} A_{v\omega} \delta(c_v, i) \delta(c_\omega, i) - \sum_{v\omega} \frac{k_v}{2m} \frac{k_\omega}{2m} \delta(c_v, i) \delta(c_\omega, i) \right\}$$

$$Q = \sum_i \left\{ \frac{1}{2m} \sum_{v\omega} A_{v\omega} \delta(c_v, i) \delta(c_\omega, i) - \sum_v \frac{k_v}{2m} \delta(c_v, i) \sum_\omega \frac{k_\omega}{2m} \delta(c_\omega, i) \right\}$$

$$Q = \sum_i \left\{ \frac{1}{2m} \sum_{v\omega} A_{v\omega} \delta(c_v, i) \delta(c_\omega, i) - \sum_v \frac{k_v}{2m} \delta(c_v, i) \sum_\omega \frac{k_\omega}{2m} \delta(c_\omega, i) \right\}$$

$\sum_v \frac{k_v}{2m} \delta(c_v, i) = \frac{1}{2m} \sum_v k_v \delta(c_v, i)$, se suman los numeros de links que salen de cada nodo que pertenecen a la comunidad i

$$Q = \sum_i \left\{ \frac{1}{2m} \sum_{v\omega} A_{v\omega} \delta(c_v, i) \delta(c_\omega, i) - \left[\sum_v \frac{k_v}{2m} \delta(c_v, i) \right]^2 \right\}$$

El termino $\sum_{v\omega} \frac{1}{2m} \frac{k_v k_\omega}{2m} \delta(c_v, c_\omega)$ representa la probabilidad de que si yo "coloco" un link al azar preservando el grado de cada nodo este caiga entre nodos de la comunidad. Es como elegir un link "saliente" con proba $\frac{k_v}{2m}$ y uno "entrante" con proba $\frac{k_\omega}{2m}$

Entonces

$$Q = \sum_{v\omega} \left[\frac{A_{v\omega}}{2m} - \frac{k_v \cdot k_\omega}{2m \cdot 2m} \right] \delta(c_v, c_\omega)$$

correlaciones reales en el sistema

la muy conocida $\rho^{(2)}$ vs $\rho\rho$

Si no hay correlaciones i.e. no hay estructura de comunas $\rightarrow Q=0$

Test sobre Q

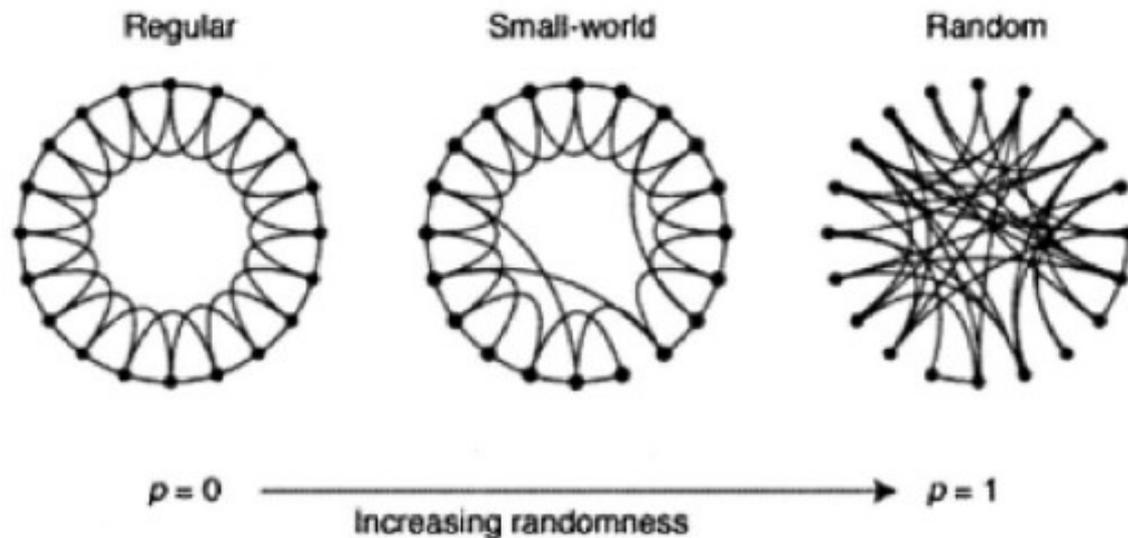
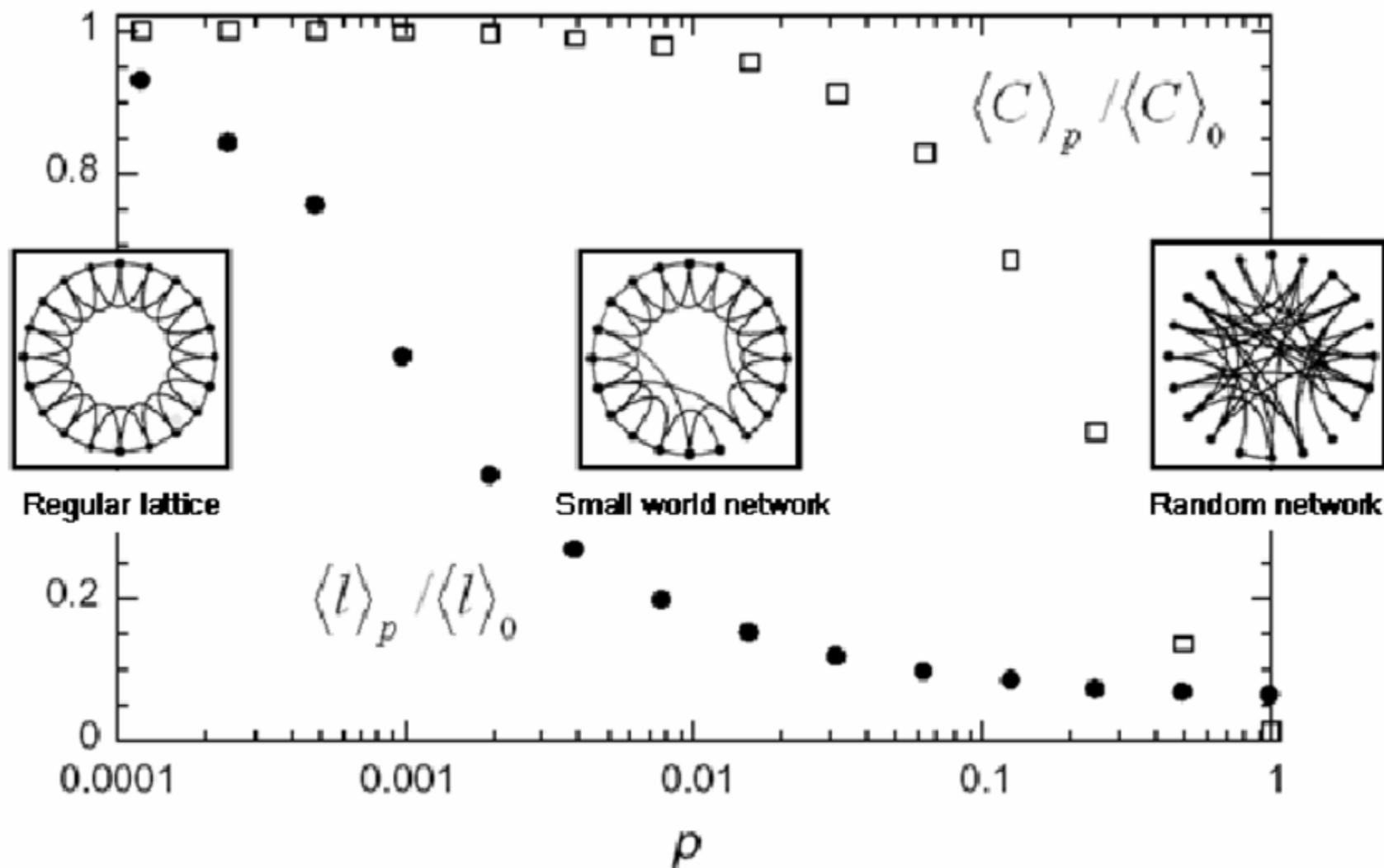


FIG. 15. The random rewiring procedure of the Watts-Strogatz model, which interpolates between a regular ring lattice and a random network without altering the number of nodes or edges. We start with $N=20$ nodes, each connected to its four nearest neighbors. For $p=0$ the original ring is unchanged; as p increases the network becomes increasingly disordered until for $p=1$ all edges are rewired randomly. After Watts and Strogatz, 1998.



Estudiamos el Q de Newman para el siguiente caso

Sea un WS con

N nodos

kN links \Rightarrow de cada nodo salen $2kN$ links

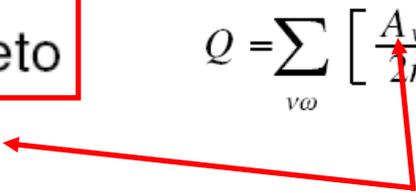
luego el grado de cada nodo es $2k$

Calculamos Q para el grafo completo

l_s es el número de links internos

L es el número total de links

El primer término de Q resulta ser $= \frac{l_s}{L} = \frac{kN}{kN} = 1$

$$Q = \sum_{v\omega} \left[\frac{A_{v\omega}}{2m} - \frac{k_v \cdot k_\omega}{2m \cdot 2m} \right] \delta(c_v, c_\omega)$$


Todos los links son internos

El segundo termino de Q es el grado total dividido por dos veces el numero de links = $\frac{2kN}{2kN} = 1$

$$Q = 1 - 1 = 0$$

Ahora lo partimos en 2 (supongo N par), para cada subgrafo

$$Q_i = \left\{ \frac{\frac{kN}{2} - 2k}{kN} \right\} - \left\{ \frac{2kN/2}{2kN} \right\}^2$$

multiplicando por 2

$$Q = 1 - \frac{1}{2} - \frac{4}{N}$$

Con N muy grande

$$Q \simeq 0.5$$

Luego Q crece mal

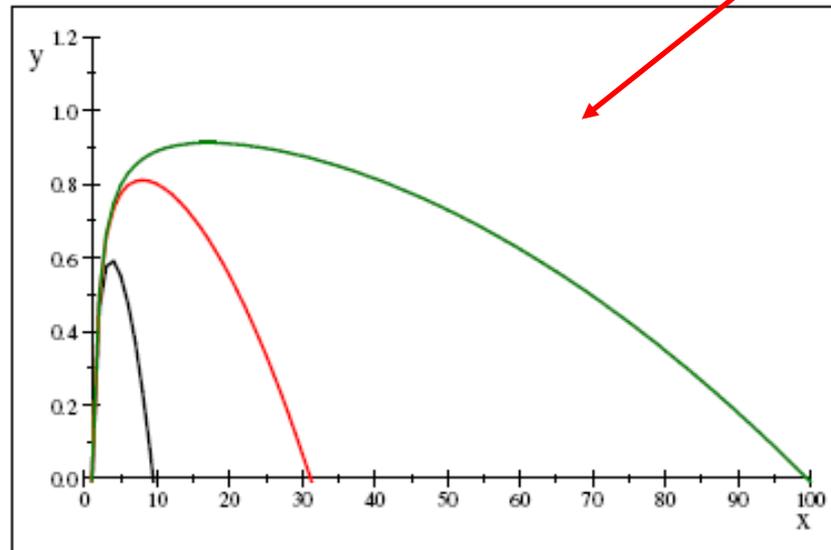
ojo con los links cortados

Entonces si lo parto en n subgrafos

$$Q = 1 - \frac{1}{n} - \frac{n^2}{N}$$

!!!

$$1 - \frac{1}{n} - \frac{n^2}{100}$$



En negro $N = 100$, en rojo $N = 1000$, en verde $N = 10000$

Lo cual invalida los metodos que maximizan Q

Sin embargo al cortar los links los estoy aun contando, asi que pensamos que esos links no existen mas

$$\begin{aligned} Q_i &= \left\{ \frac{\frac{kN}{2} - 2k}{kN} \right\} - \left\{ \frac{2kN/2 - 2k}{2kN} \right\}^2 \\ &= \frac{1}{2} - \frac{2}{N} - \left[\frac{1}{2} - \frac{1}{N} \right]^2 = \frac{1}{2} - \frac{2}{N} - \frac{1}{4} + \frac{1}{N} - \frac{1}{N^2} = \frac{1}{4} - \frac{1}{N} - \frac{1}{N^2} \end{aligned}$$

Entonces Q da

$$Q = \frac{1}{2} - \frac{2}{N} - \frac{2}{N^2}$$

Betweenness

$$b_{ij} = \sum_{paths} \alpha_{no}^{-1} \sum_{l_{km} \in path_{no}} \delta(l_{ij} - l_{km})$$

Se propone entonces

In a recent work, Newman and Girvan [3] have proposed to study the structure of the network by analyzing the effect of the removal of links with highest betweenness. The betweenness b_{ij} of a given link l_{ij} is

$$b_{ij} = \sum_{paths} \alpha_{no}^{-1} \sum_{l_{km} \in path_{no}} \delta(l_{ij} - l_{km}) \quad (2)$$

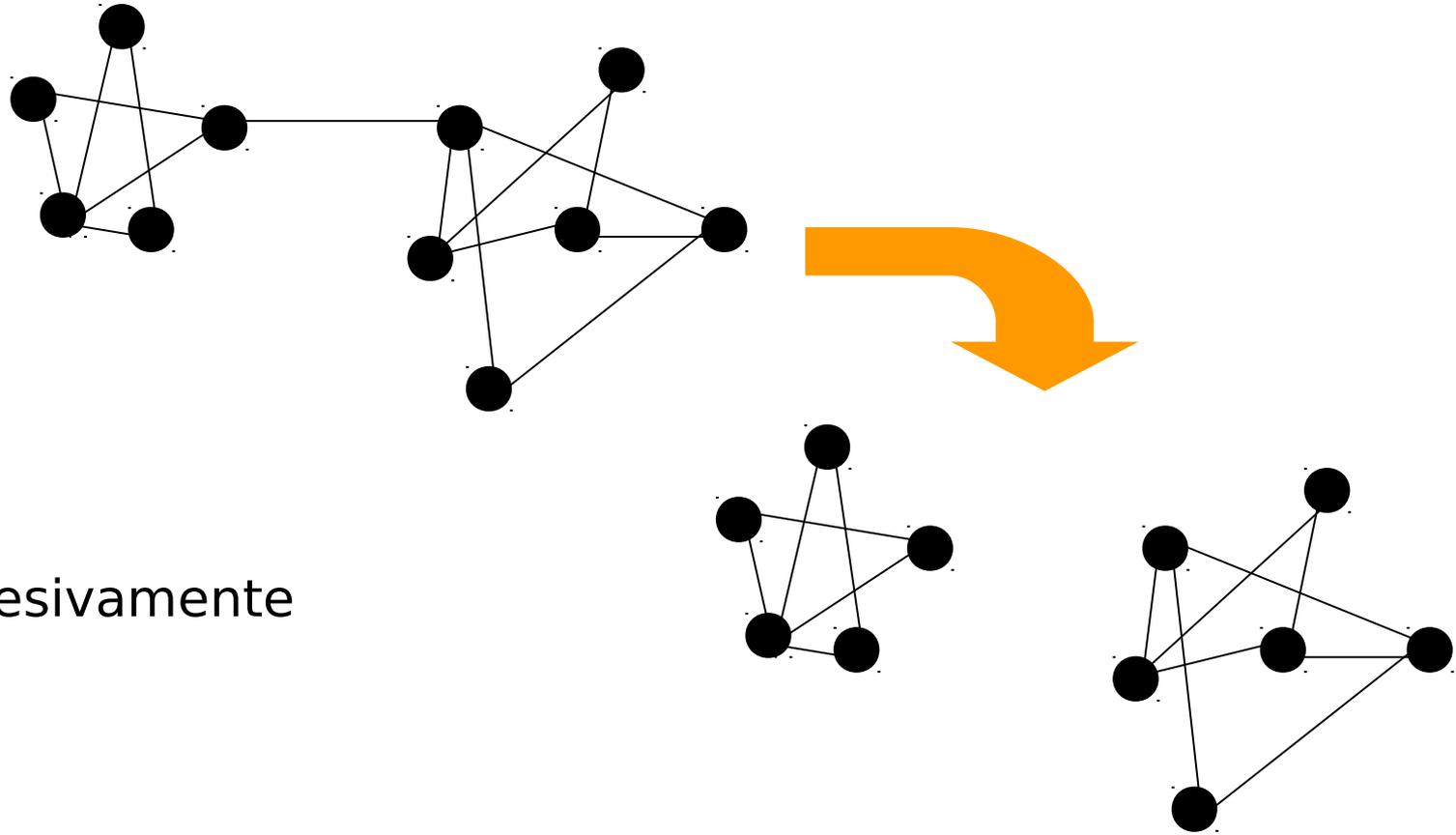
with \sum_{paths} the sum over all the path joining the n_n nodes. α_{no} is the degeneracy of the path between nodes n and o , and $\sum_{l_{km} \in path_{no}}$ is the sum over all the links l_{km} that form the path under consideration. In this way the link with highest betweenness is the one that appears most often when we study all the components of all the minimum paths between all the pairs of nodes.

According to this prescription:

(i) One calculates the betweenness of all the links in the network. (ii) The one with the highest betweenness is removed.

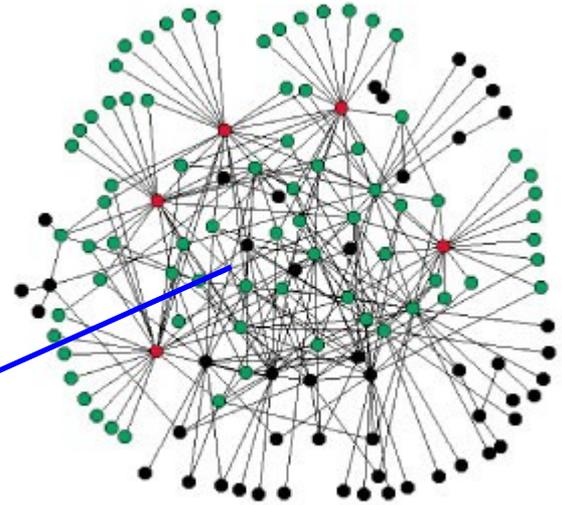
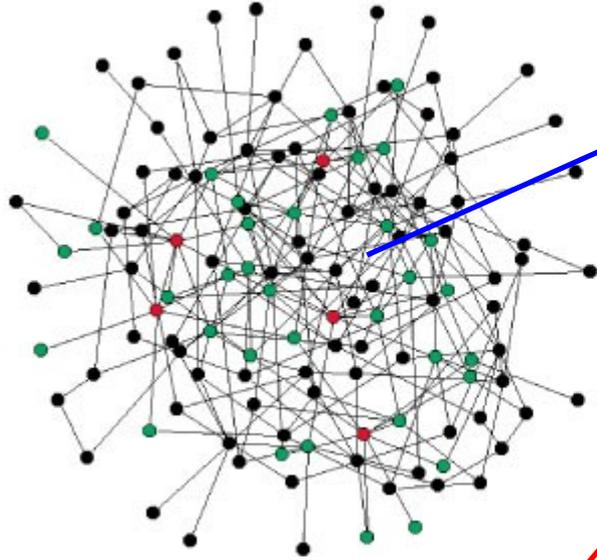
The process is continued until a disjoint cluster is obtained. Afterwards, it is applied to each of the resulting subgraphs.

El método de NG consiste en buscar la unión con máximo B y
Removerla de la red



Y así sucesivamente

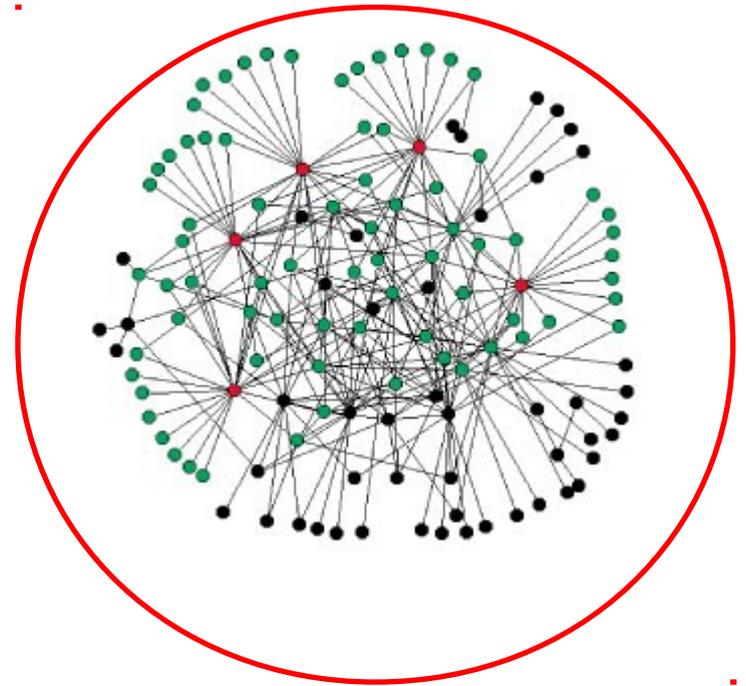
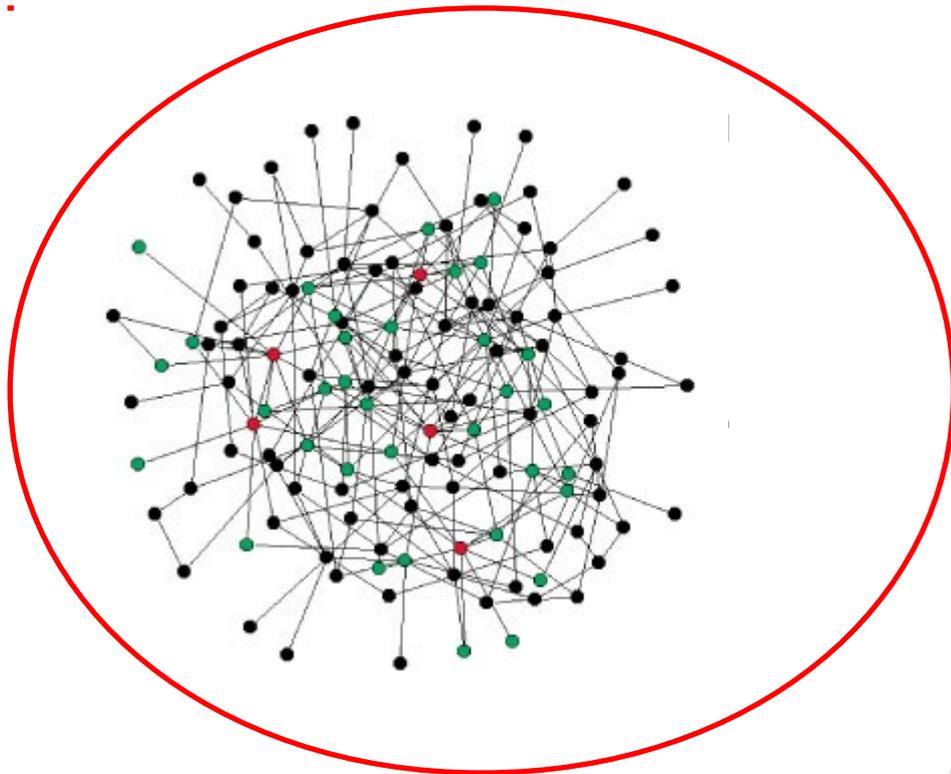
Cual es la idea de esto?



Link de alta betweenness

Gran candidato a cortar!

Cual es la idea de esto?



Q disminuye

Simulated annealing

$$q_{ij} = \min(1, \exp(-\beta(E_j - E_i))) .$$


$$q_{ij} = \min(1, \exp(-\beta'(Q_j - Q_i)))$$



Available online at www.sciencedirect.com

SCIENCE @ DIRECT®

Physica A 358 (2005) 593–604

PHYSICA A

www.elsevier.com/locate/physa

Detection of community structures in networks
via global optimization [☆]

A. Medus, G. Acuña, C.O. Dorso*

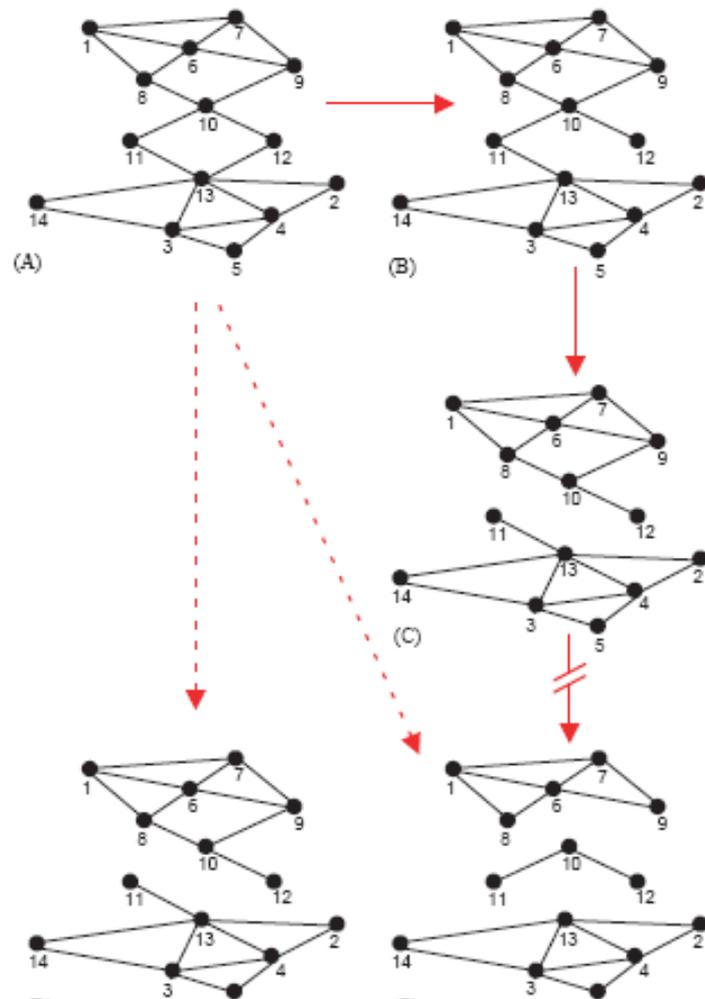
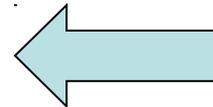
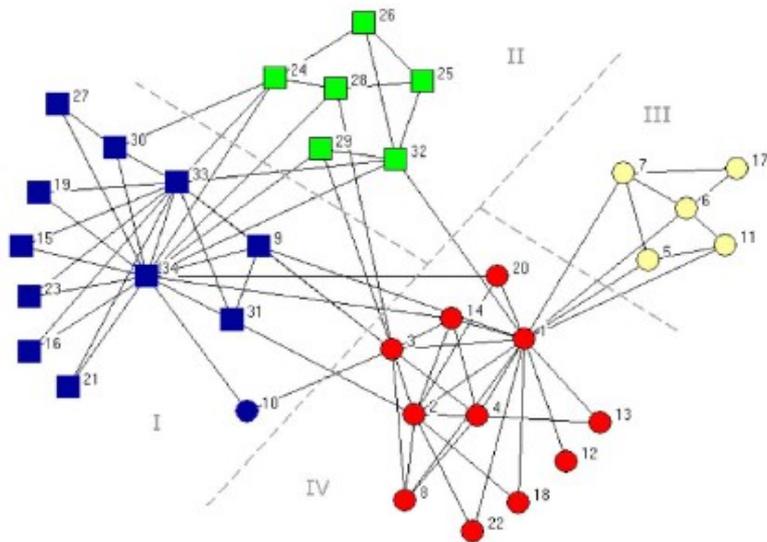


Fig. 1. Development of community structures in terms of the ER and SA analysis. Full arrows denote steps in the ER approach. Dotted arrows denote results from SA methodology. Starting from network A by applying ER methodology we first get to network B and, after the second removal of a link, to network C. On the other hand, starting from the same initial network the SA will give network D if we impose the constraint that the final configuration should display two communities. If we do not impose any constraint the result according to SA will be network E. It is important to notice that network E is unreachable from network C. This is the main drawback of the ER approach.



SA
Q=0.42

Fig. 2. Community structures for the Zachary network according to SA approach. In this figure, squares and circles denote the members of the two subsets according to observations by Zachary. Broken lines denote the partitions obtained according to SA approach.

ER
Q=0.36

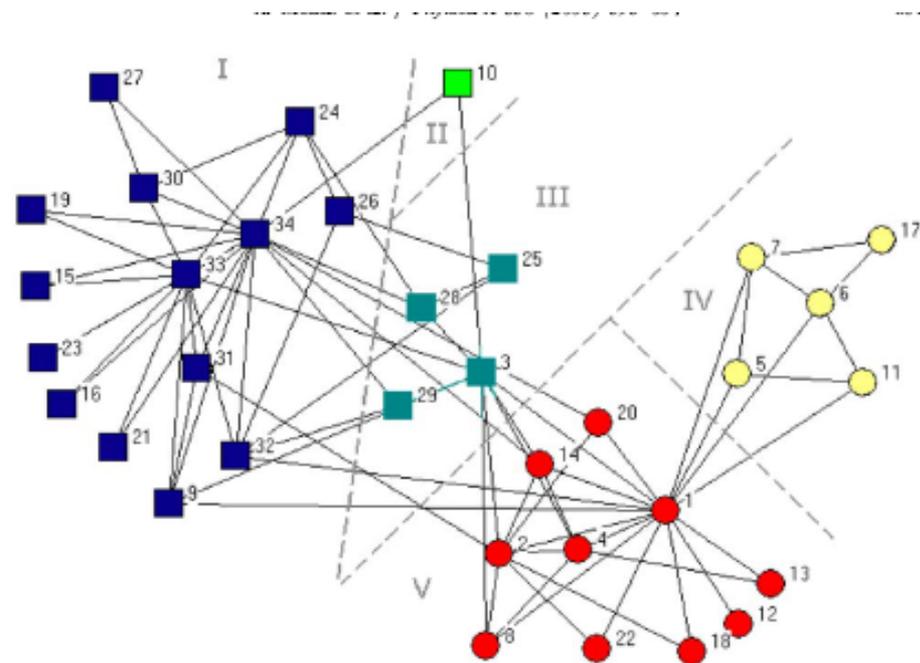


Fig. 3. Community structures for the Zachary network according to ER approach.

Alternative approach to community detection in networks

A. D. Medus^{*} and C. O. Dorso[†]

Departamento de Física, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, Pabellón 1, Ciudad Universitaria, Ciudad Autónoma de Buenos Aires 1428, Argentina

(Received 28 July 2008; revised manuscript received 21 May 2009; published 23 June 2009)

(i) *Community in strong sense.* C is a community in the strong sense if

$$k_i^{in} > k_i^{out} \quad \forall i \in C. \quad (1)$$

(ii) *Community in weak sense.* C is a community in weak sense if

$$\sum_{i \in C} k_i^{in} > \sum_{i \in C} k_i^{out}. \quad (2)$$

Given a graph G and a m -subgraphs partition $\{C_j\}_{1 \leq j \leq m}$, where each subgraph $C_j \subset G$ constitutes a community according to any of the local definitions mentioned in the previous section, we want to define a quantity that measures the “quality” of each of the resulting communities.

“community strength” (S)

$$S(C_j) = \sum_{i \in C_j} \frac{k_i^{in} - k_i^{out}}{2L(C_j)},$$

$$L(C_j) = \frac{1}{2} \sum_{i \in C_j} k_i.$$

merit factor Q_W

$$Q_W = \sum_{j=1}^m S(C_j) = \sum_{j=1}^m \sum_{i \in C_j} \frac{k_i^{in} - k_i^{out}}{2L(C_j)}$$

with the constraint that each subgraph $C_j \subset \{C_j\}_{1 \leq j \leq m}$ must satisfy the weak community definition, i.e.,

$$S(C_j) > 0 \quad \forall C_j \subset \{C_j\}_{1 \leq j \leq m}. \quad (6)$$

In the same spirit we now define a merit factor Q_S according to the strong community definition:

$$Q_S = \sum_{j=1}^m S(C_j) = \sum_{j=1}^m \sum_{i \in C_j} \frac{k_i^{in} - k_i^{out}}{2L(C_j)} \quad (7)$$

with the constraint

$$(k_i^{in} - k_i^{out}) > 0 \quad \forall i \in C_j. \quad (8)$$

Definition. The optimal m -subgraphs partition $\{C_j\}_{1 \leq j \leq m}$ of a graph G in the strong (weak) sense is that one with maximal merit factor $Q_S(Q_W)$.

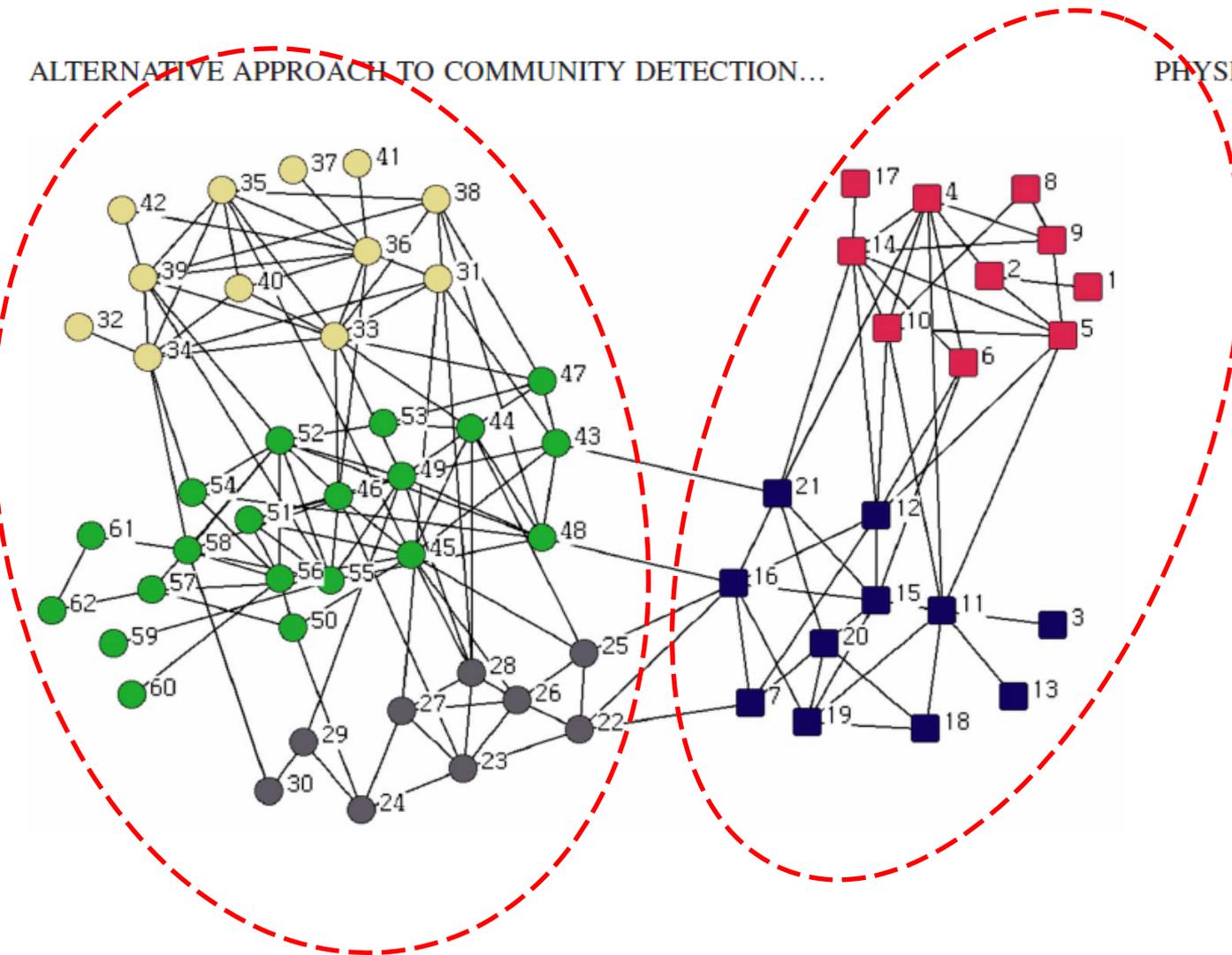


FIG. 4. (Color online) Bottleneck dolphin network. This network has a size of 62 nodes and it is known from direct observation that it has two communities. In this figure squares and circles denote the communities detected by our strong community approach and the colors (shades of gray or colors online) show the results of the weak community approach. Notice that the optimization according to Q_W merely subdivides the communities obtained through Q_S optimization [17].

Red anillo

Definición débil:

Dos comunas de 10 nodos cada una.

$$Q_W(2 \text{ comunas}) = 1.2$$

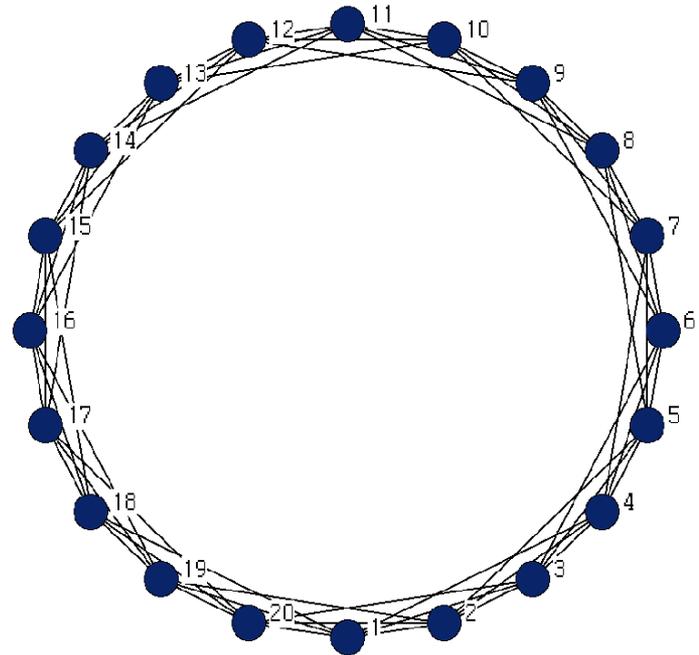
Modularidad Q de Newman:

Tres comunas: dos de 7 y una de 6 nodos.

$$Q = 0.365$$

Definición fuerte:

No hay partición.



20 nodos, k=6

Random Graph

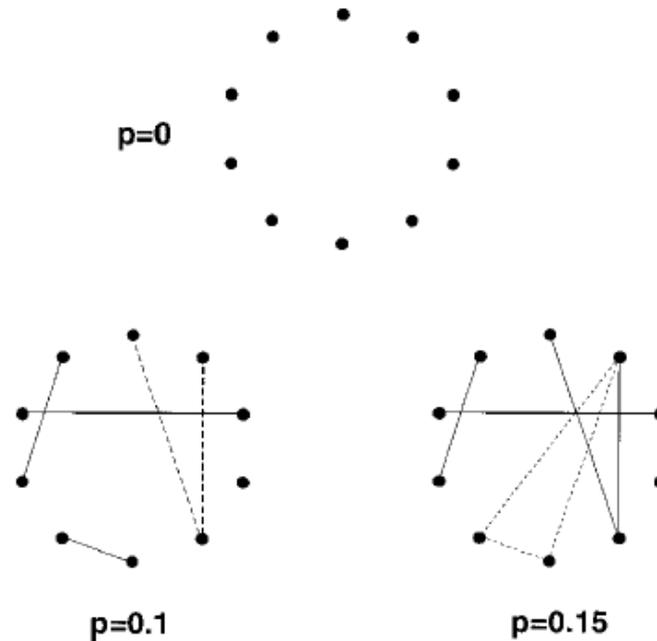


FIG. 5. Illustration of the graph evolution process for the Erdős-Rényi model. We start with $N=10$ isolated nodes (upper panel), then connect every pair of nodes with probability p . The lower panel of the figure shows two different stages in the graph's development, corresponding to $p=0.1$ and $p=0.15$. We can notice the emergence of trees (a tree of order 3, drawn with long-dashed lines) and cycles (a cycle of order 3, drawn with short-dashed lines) in the graph, and a connected cluster that unites half of the nodes at $p=0.15=1.5/N$.

Random Graphs

Erdős & Renyi

Tomemos un numero de n vertices y conectemos cada par con una probabilidad p , esto define el ensemble $G_{n,p}$ en el cual el grafo con m lados aparece con una probabilidad

donde $M = \frac{1}{2}n(n-1)$

$$G_{n,p} \rightarrow p^n (1-p)^{M-n}$$

Tambien definieron el modelo $G_{n,m}$ que es el enesemble con n vertices y exactamente m lados todos ellos con la misma probabilidad (o sea que este es del tipo microcanonico y el anterior del tipo canonico)

Distribución de grado

$$P(k_i=k) = C_{N-1}^k p^k (1-p)^{N-1-k}.$$

la proba que un dado nodo tenga k links

La expectación del numero de nodos con grado k es

$$\begin{aligned} \rightarrow E(X_k) &= NP(k_i=k) = \lambda_k, \\ \lambda_k &= NC_{N-1}^k p^k (1-p)^{N-1-k}. \end{aligned}$$

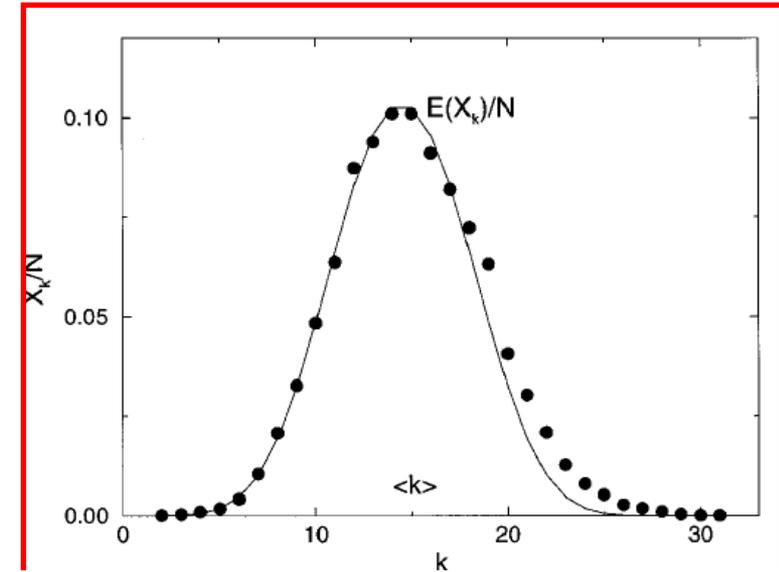


FIG. 7. The degree distribution that results from the numerical simulation of a random graph. We generated a single random graph with $N=10\,000$ nodes and connection probability $p=0.0015$, and calculated the number of nodes with degree k, X_k . The plot compares X_k/N with the expectation value of the Poisson distribution (13), $E(X_k)/N=P(k_i=k)$, and we can see that the deviation is small.

Que tiende a

$$P(X_k=r) = e^{-\lambda_k} \frac{\lambda_k^r}{r!}.$$

En este caso a medida que vamos incrementando p a partir de un valor bajo empiezan a aparecer "estructuras" conexas.

El resultado fundamental es que (no resulta insperado)

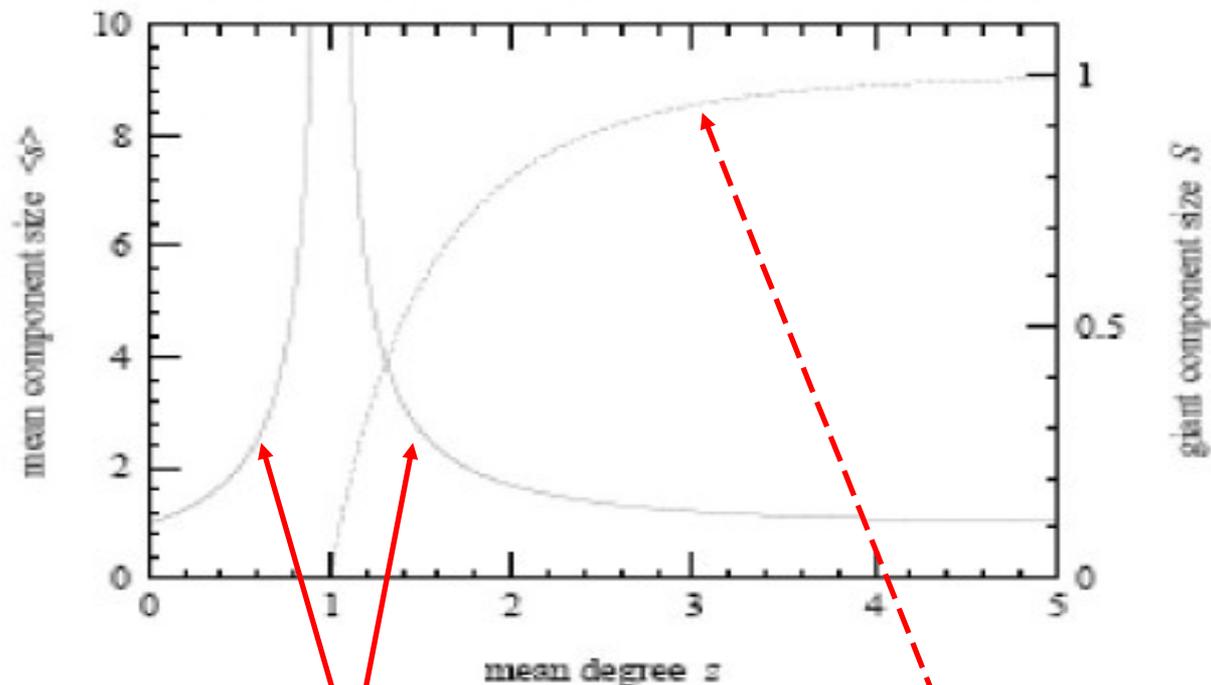


FIG. 10 The mean component size (solid line), excluding the giant component if there is one, and the giant component size (dotted line), for the Poisson random graph, Eqs. (20) and (21).

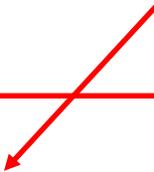
Es decir que existe una probabilidad crítica a partir de la cual una fracción macroscópica de nodos se agrupan en lo que se llama "el componente gigante"

Sea u la fracción de nodos que no pertenecen al componente gigante.

la probabilidad de que un nodo no pertenezca a la componente gigante es igual a al proba que ninguno de sus vecinos pertenezca a dicha componente

Si el nodo tiene grado k esta probabilidad es u^k

Entonces

$$u = \sum_{k=0}^{\infty} p_k u^k = e^{-z} \sum_{k=0}^{\infty} (zu)^k \frac{1}{k!} = e^{-z} e^{zu} = \exp(z(u-1))$$


pues $\sum_{k=0}^{\infty} \frac{x^k}{k!} = e^x$

Entonces la fracción S de nodos que pertenecen a la componente gigante es

$$S = (1 - u) = (1 - e^{-zS})$$

Se demuestra que

$$\langle s \rangle = \frac{1}{1 - z - zS}$$

Entonces estamos en algo que se parece a una transición de fase donde S juega el rol de un parámetro de orden y $\langle s \rangle$ de las fluctuaciones.

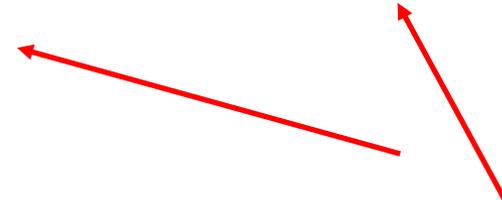
Aparecen entonces exponentes críticos

$$S \sim (z - 1)^\beta$$

$$\langle s \rangle \sim |z - 1|^{-\gamma}$$

con β y $\gamma = 1$

La transición ocurre en $z = 1$ y además esto es fijar p pues
 $z = p(n - 1)$



Si se estudia la distribución de fragmentos en $z = 1$ se encuentra que están distribuidos con $\tau = 3/2$ cuando uno elige nodos del modo que hicimos en percolación.

El random graph describe solo el camino mínimo medio de los grafos reales.

Camino mínimo

Network	Size	$\langle k \rangle$	ℓ	ℓ_{rand}
WWW, site level, undir.	153 127	35.21	3.1	3.35
Internet, domain level	3015–6209	3.52–4.11	3.7–3.76	6.36–6.18
Movie actors	225 226	61	3.65	2.99
LANL co-authorship	52 909	9.7	5.9	4.79
MEDLINE co-authorship	1 520 251	18.1	4.6	4.91
SPIRES co-authorship	56 627	173	4.0	2.12
NCSTRL co-authorship	11 994	3.59	9.7	7.34
Math. co-authorship	70 975	3.9	9.5	8.2
Neurosci. co-authorship	209 293	11.5	6	5.01
<i>E. coli</i> , substrate graph	282	7.35	2.9	3.04
<i>E. coli</i> , reaction graph	315	28.3	2.62	1.98
Ythan estuary food web	134	8.7	2.43	2.26
Silwood Park food web	154	4.75	3.40	3.23
Words, co-occurrence	460.902	70.13	2.67	3.03
Words, synonyms	22 311	13.48	4.5	3.84
Power grid	4941	2.67	18.7	12.4
<i>C. Elegans</i>	282	14	2.65	2.25

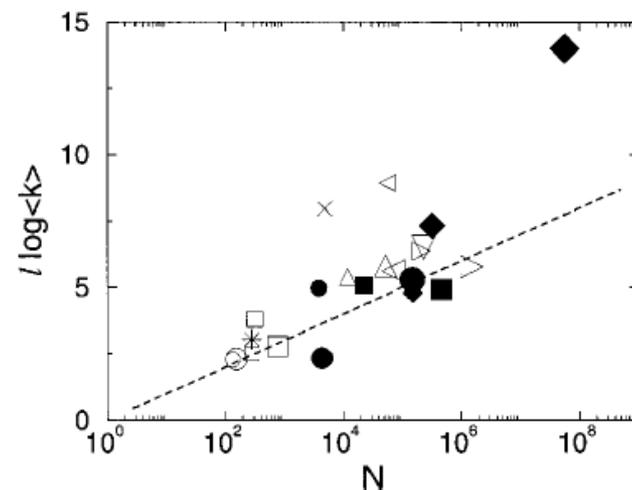


FIG. 8. Comparison between the average path lengths of real networks and the prediction (17) of random-graph theory (dashed line). For each symbol we indicate the corresponding number in Table I or Table II: small \circ , I.12; large \circ , I.13; \star , I.17; small \square , I.10; medium \square , I.11; large \square , II.13; small \bullet , II.6; medium \bullet , I.2; \times , I.16; small \triangle , I.7; small \blacksquare , I.15; large \triangle , I.4; small \triangleleft , I.5; large \triangleleft , I.6; large \bullet , II.6; small \blacklozenge , I.1; small \triangleright , I.7; ∇ , I.3; medium \blacklozenge , II.1; large \blacksquare , I.14; large \triangleright , I.5; large \blacklozenge , II.3.

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

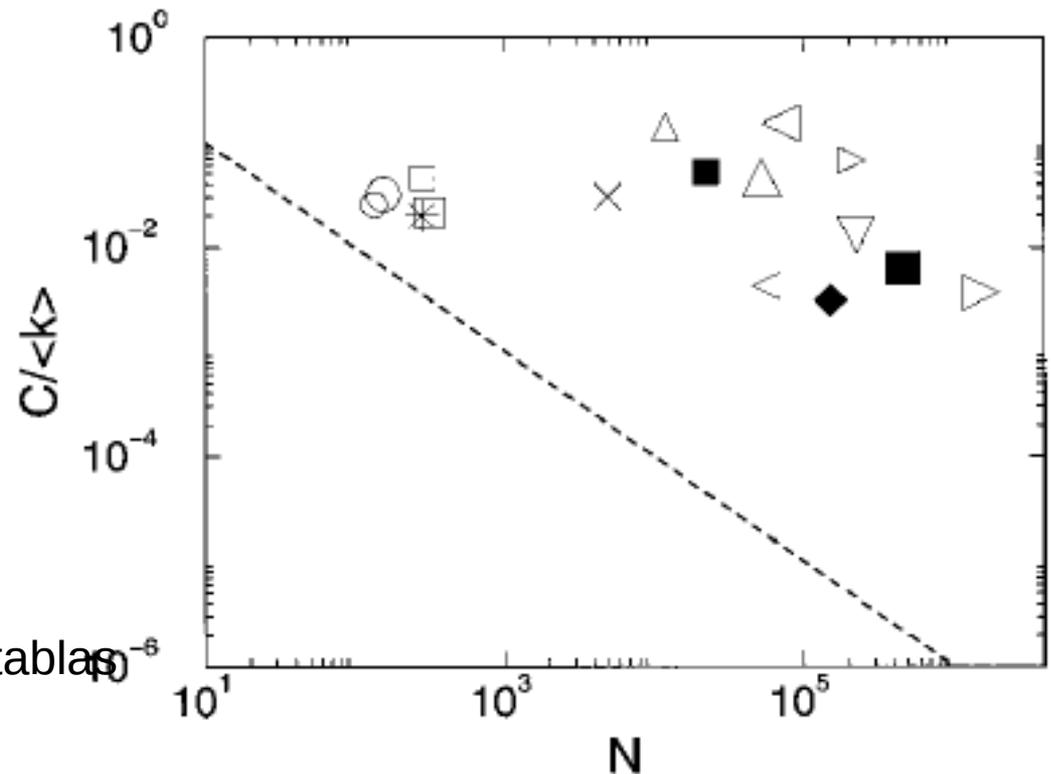
$$\ell_{rand} \sim \frac{\ln(N)}{\ln(\langle k \rangle)}.$$

Chung, F., and L. Lu, 2001, Adv. Appl. Math. 26, 257.

Coeficiente de clusterización

Si considero un nodo al azar en un random graph y sus vecinos inmediatos, la probabilidad de que 2 vecinos estén conectados es la misma que para un de nodos arbitrarios

$$C_{rand} = P = \frac{\langle k \rangle}{N}$$



Los datos experimentales de las tablas

Como se calculan estas cosas?

Calculo de grados y distribuciones de grado

Calculo de clusterizacion

Calculo de camino mínimo medio

Calculo de grados y distribucio

la densidad de grados
la cumulativa de grados
la correlación de grados

Son todas triviales

son power

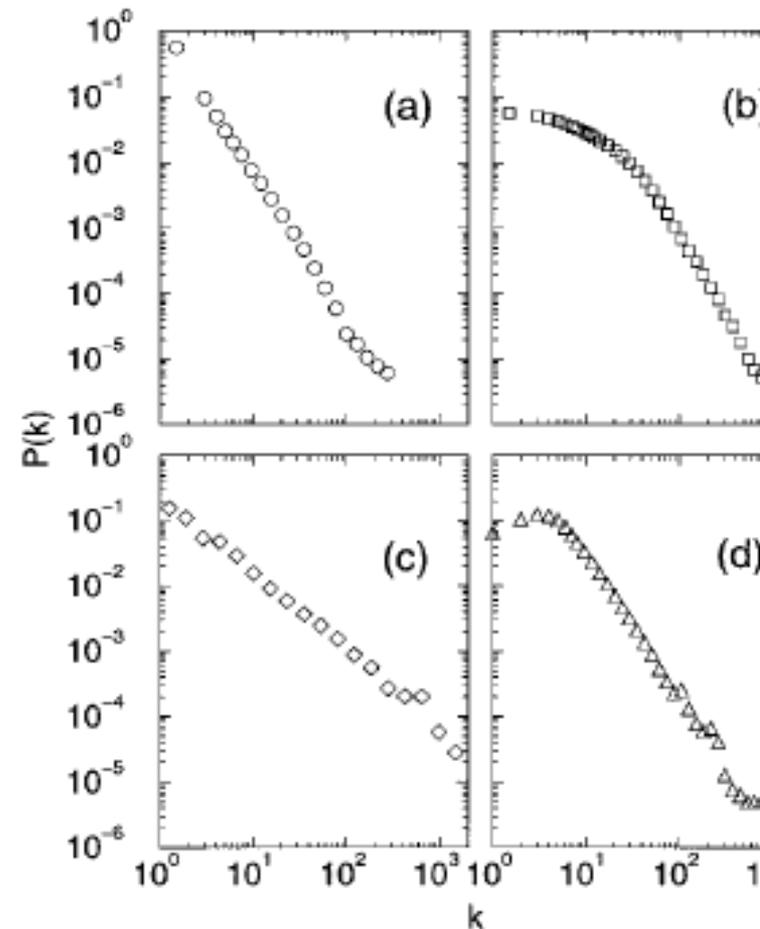


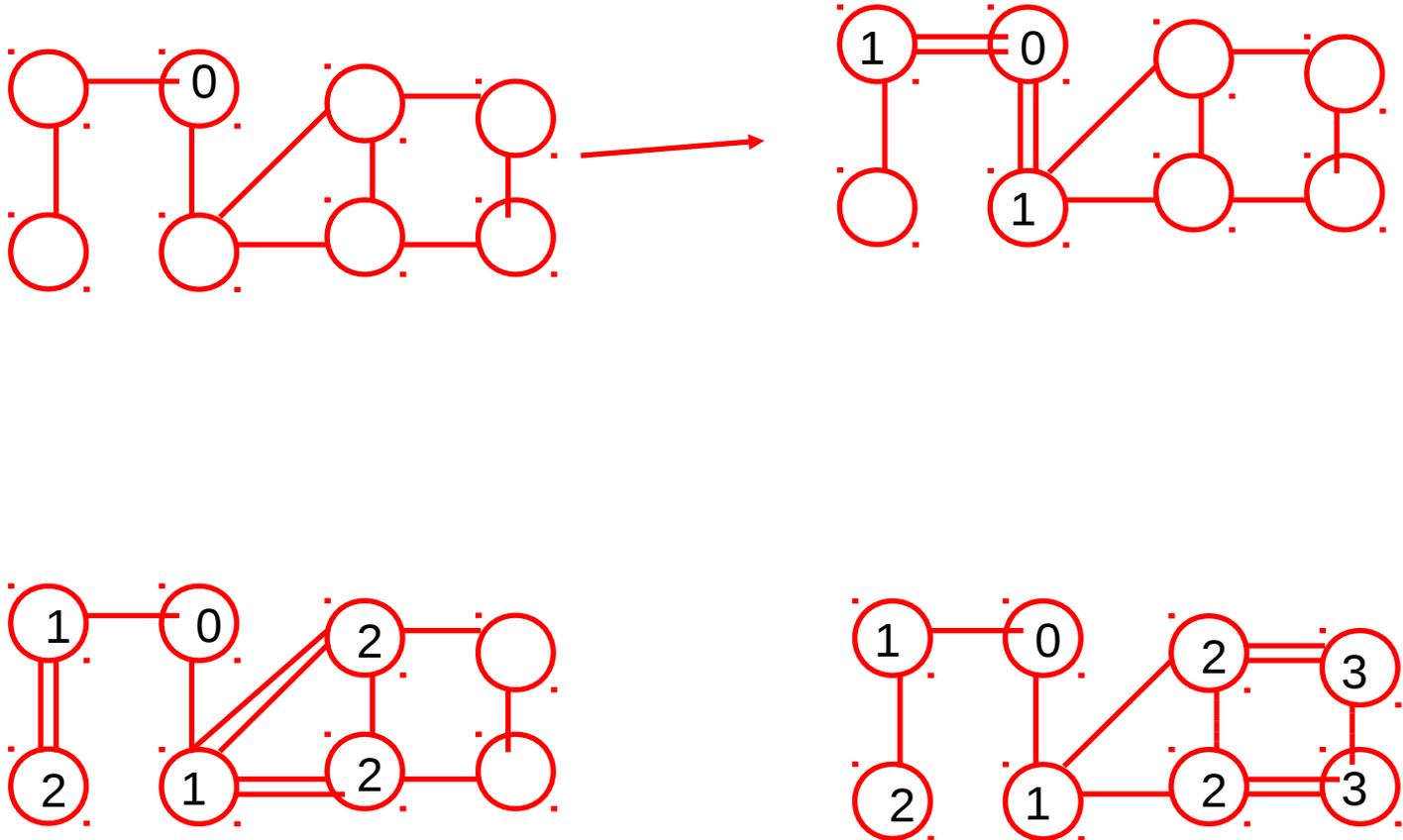
FIG. 3. The degree distribution of several real networks: (a) Internet at the router level. Data courtesy of Ramezanpour and Albert 1999; (b) movie actor collaboration network. After Ramezanpour and Albert 1999. Note that if TV series are included, an exponential cutoff emerges for large k (Amaral *et al.*, 2000); (c) co-authorship network of high-energy physicists. After Newman and Watts 2001b; (d) co-authorship network of neuroscientists. After Barabási *et al.* (2001).

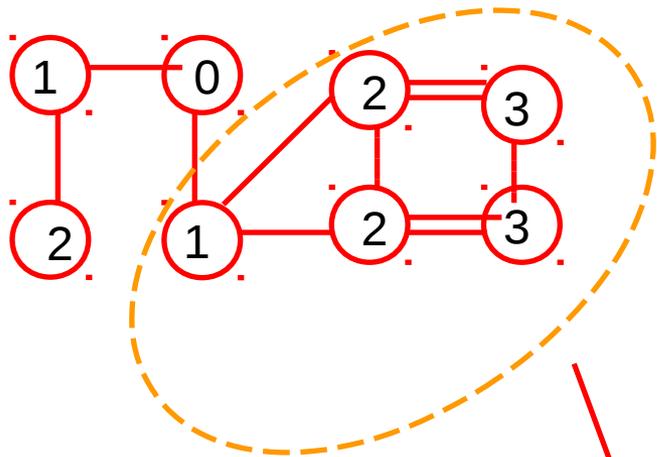
Calculo de clusterizacion

$$C_i = \frac{(\text{numero de pares de vecinos de } i \text{ que estan conectados})}{(\text{numero de pares de vecinos de } i)}$$

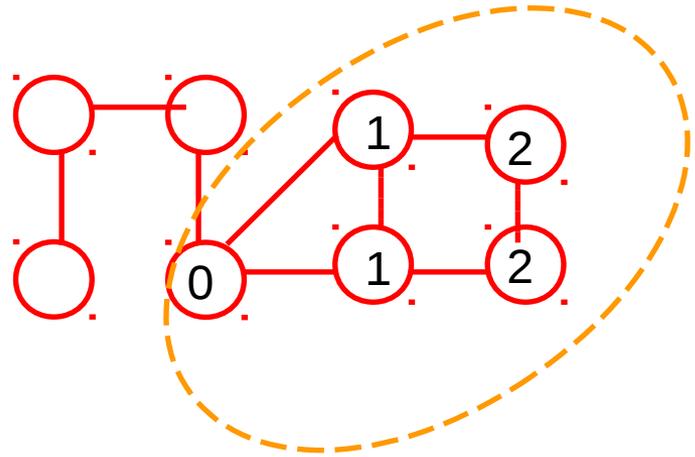
Calculo de camino mínimo medio

breadth first



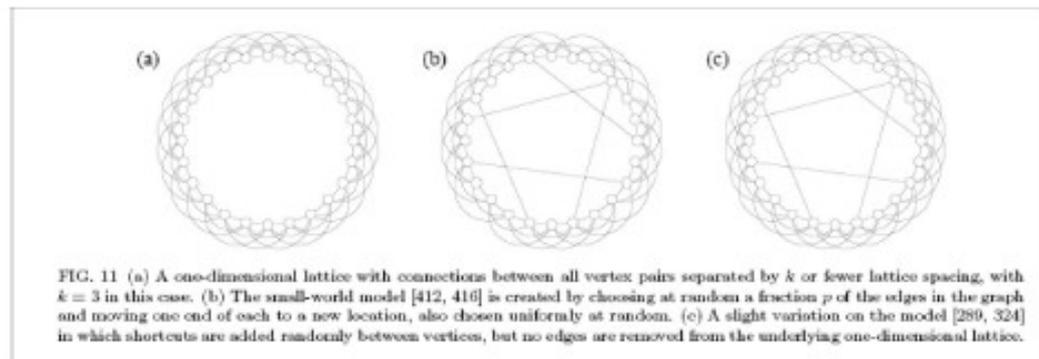


pero tambien tenemos un monton de informacion extra



Grafos de mundo pequeño

Los grafos de Watts - Strogatz



Empezamos con uno unidimensional

1) empezamos con orden

Para el caso con orden tendremos

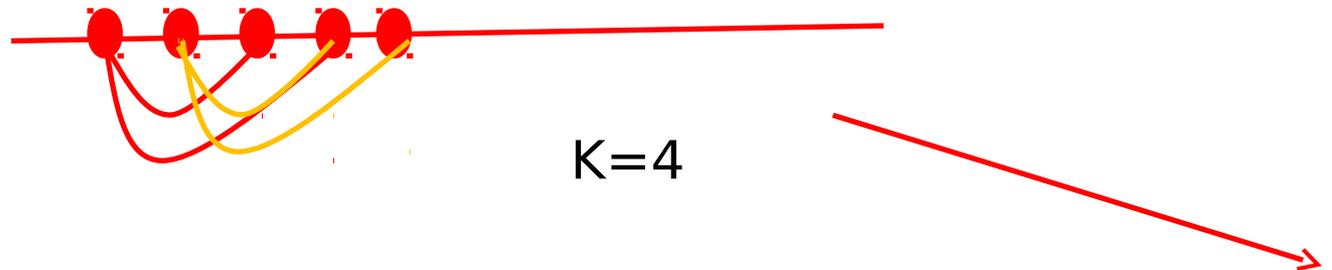
Cada vertice tiene $2k$ vecinos

El numero de links entre estos vecinos es $3k(k - 1)/2$

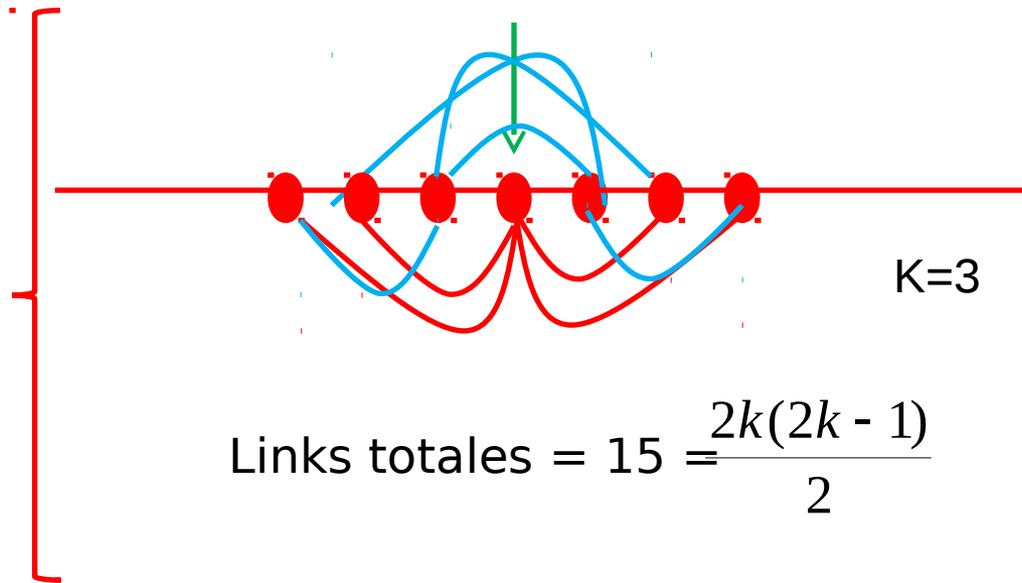
El numero total de links es $2k(2k - 1)/2$

$$C = \frac{3(k - 1)}{2(2k - 1)}$$

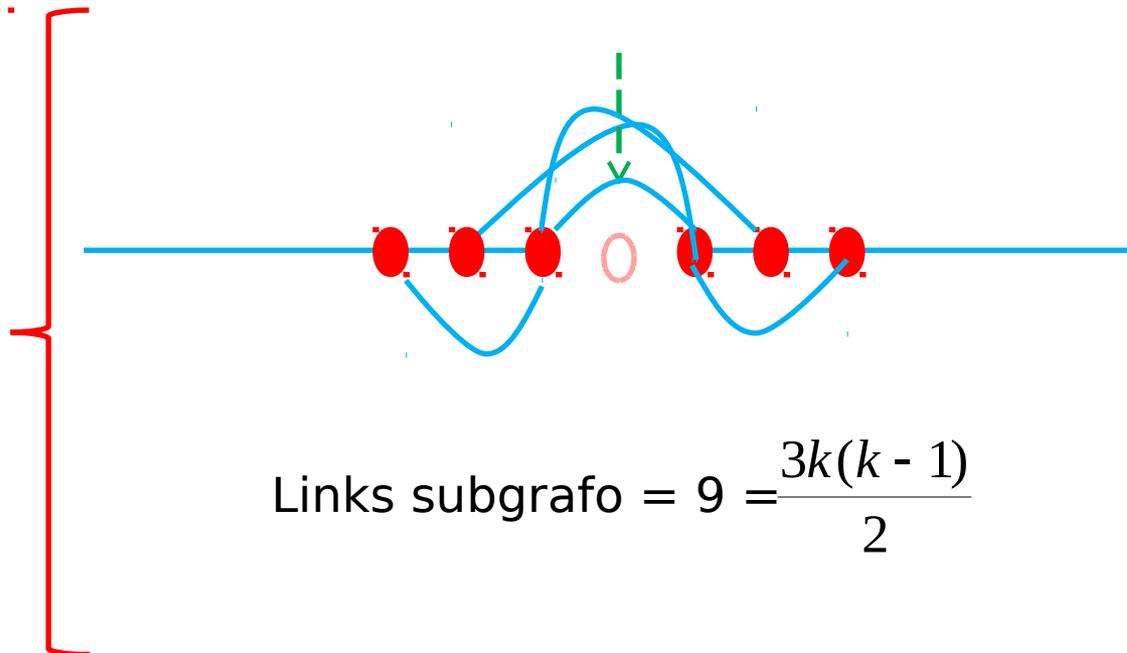
Donde K es el numero de vecinos



Nos fijamos
En un dado
nodo



Para el
subgrafo



Clustering coefficient

triangulos

Entonces

$$C = \frac{3(k-1)}{2(2k-1)}(1-p)^3$$

Que es la proba de que los 3 links sobrevivan

(rewiring)

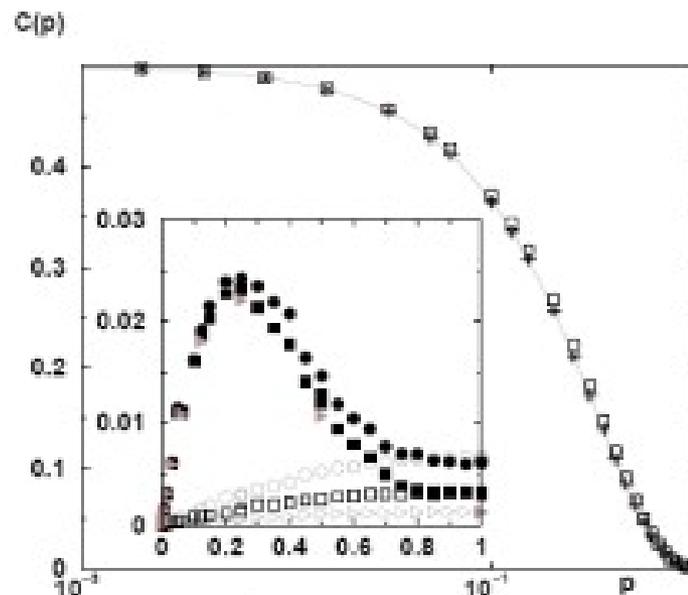


FIG. 9. $C(p)$ and $\tilde{C}(p)$ versus p , for $k = 2$ ($C(0) = \tilde{C}(0) = 0.5$), $N = 1000, 2000, 5000$: open symbols are for $C(p)$, and the crosses are for $\tilde{C}(p)$; the line is $C(0)(1-p)^3$. Inset: corrections $C(p) - C(0)(1-p)^3$ (filled symbols) for $N = 1000$ (circles), $N = 2000$ (squares) and $N = 5000$ (triangles), and $\tilde{C}(p) - C(0)(1-p)^3$ (open symbols) for $N = 1000$ (circles), $N = 2000$ (squares) and $N = 5000$ (triangles). We see that the corrections go to zero as $1/N$ for $\tilde{C}(p)$; the corrections for $C(p)$ are larger, but anyway very small.

MUNDO
PEQUEÑO !!!!

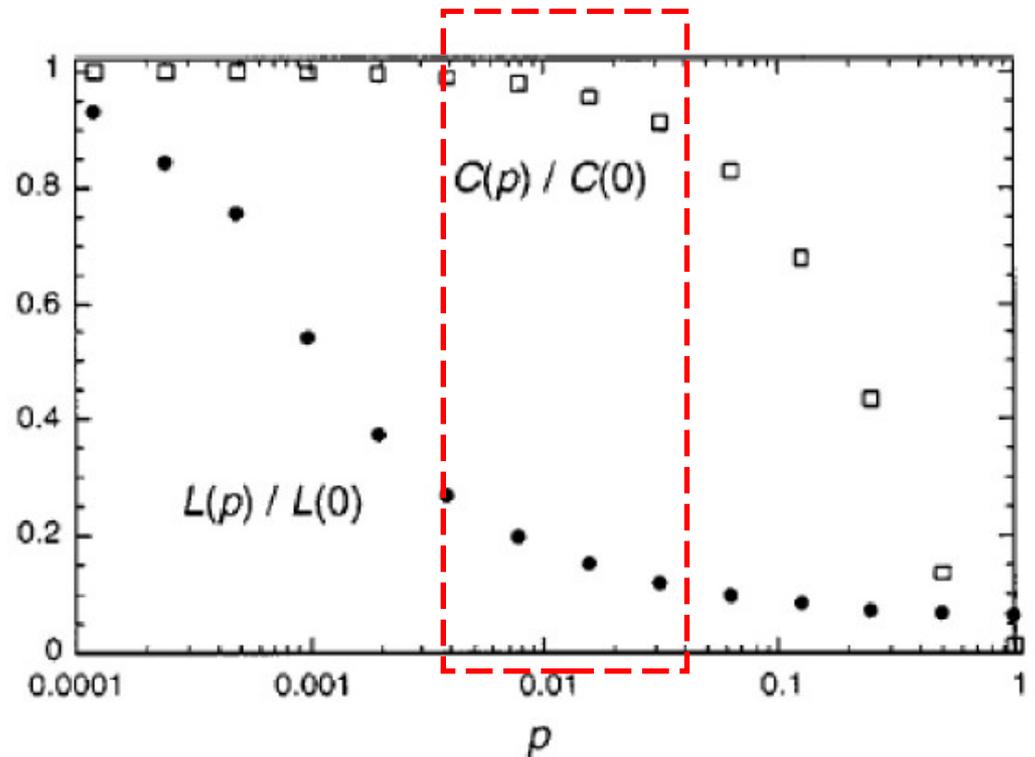


FIG. 16. Characteristic path length $\ell(p)$ and clustering coefficient $C(p)$ for the Watts-Strogatz model. The data are normalized by the values $\ell(0)$ and $C(0)$ for a regular lattice. A logarithmic horizontal scale resolves the rapid drop in $\ell(p)$, corresponding to the onset of the small-world phenomenon. During this drop $C(p)$ remains almost constant, indicating that the transition to a small world is almost undetectable at the local level. After Watts and Strogatz, 1998.

Ha sido conjeturado que el camino mínimo medio satisface la siguiente relacion de escala (recuerdos?):

[42] Barthélemy, M. and Amaral, L. A. N., Small-world networks: Evidence for a crossover picture, *Phys. Rev. Lett.* **82**, 3180–3183 (1999).

$$l = \xi g\left(\frac{l}{\xi}\right)$$

con

$$g(x) \sim \begin{cases} x & \text{para } x \gg 1 \\ \log(x) & \text{para } x \ll 1 \end{cases}$$

y ademas ξ diverge cuando $p \rightarrow 0$

Esto se compró "experimentalmente"

También conjeturaron que $\xi \sim p^{-\tau}$ con $\tau = 2/3$ pero se encontró que $\tau = 1$