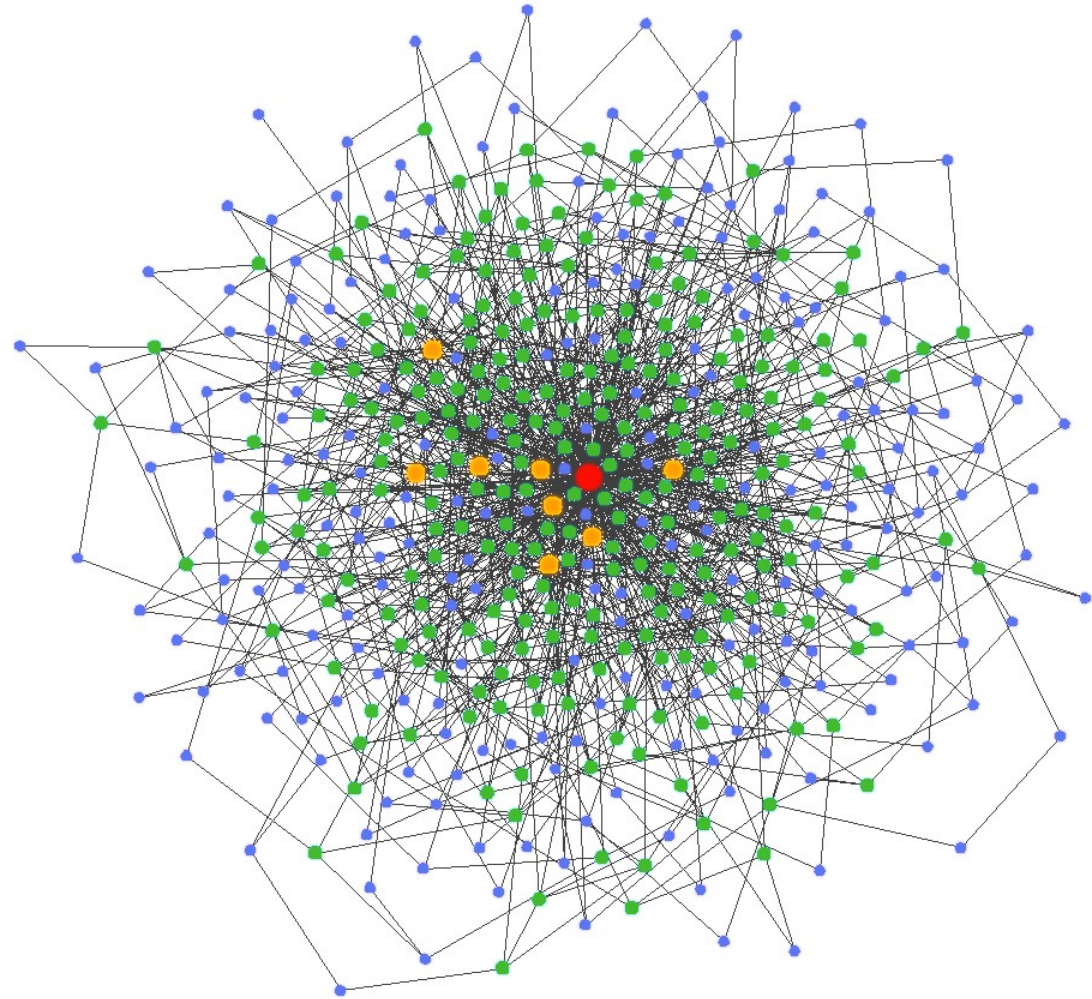
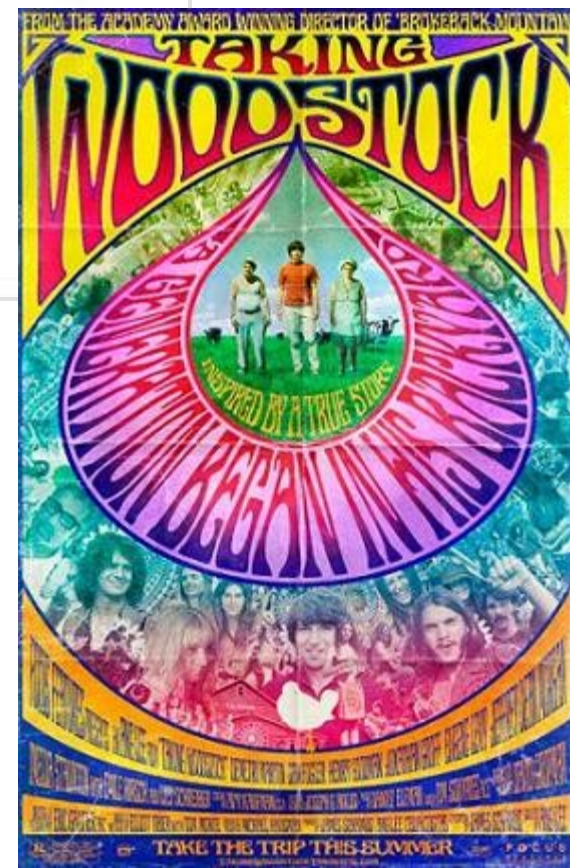
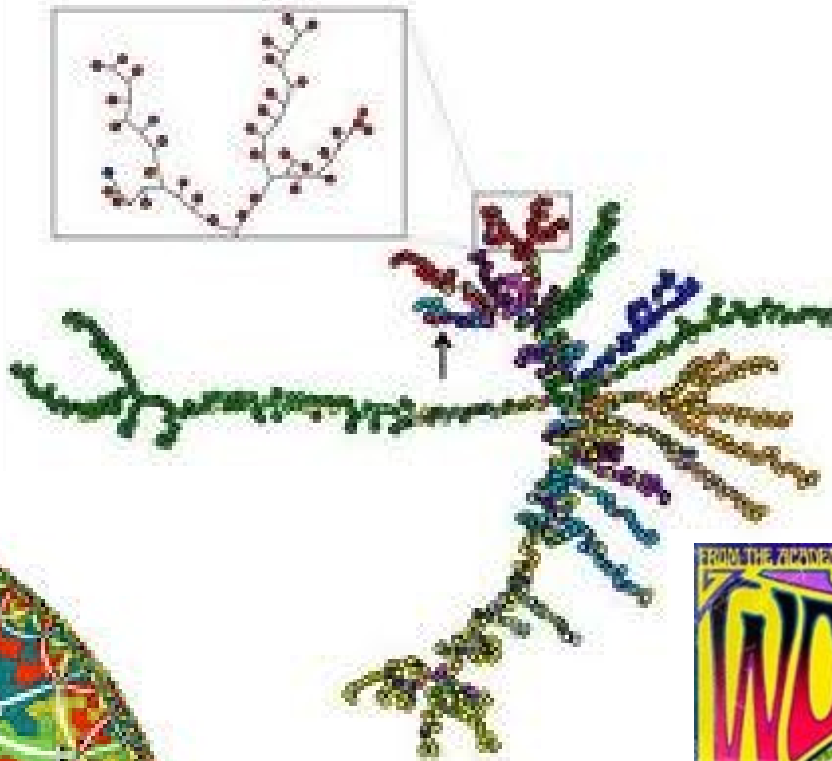
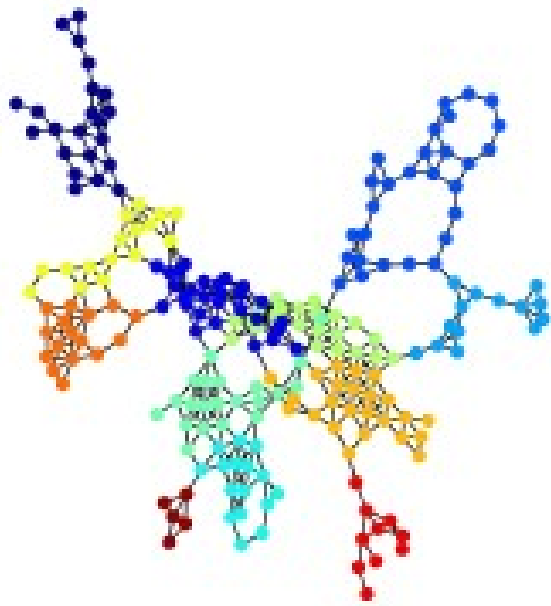


Complex Networks

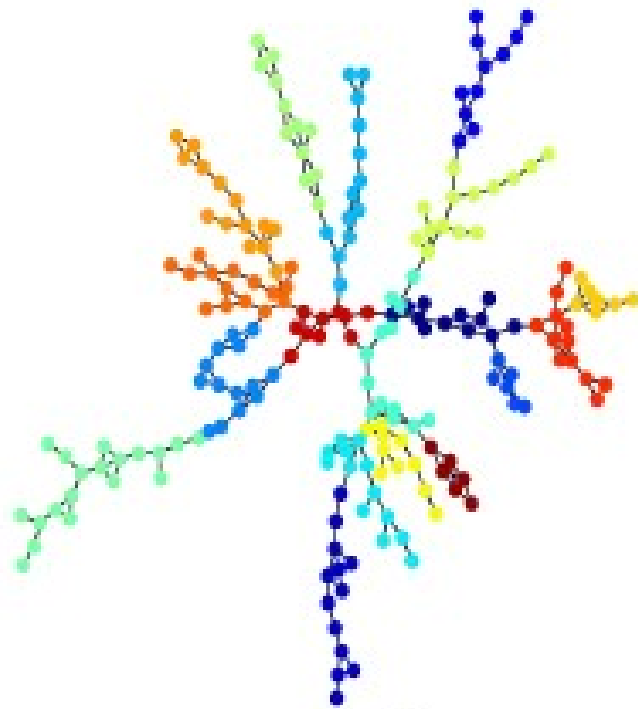




侍



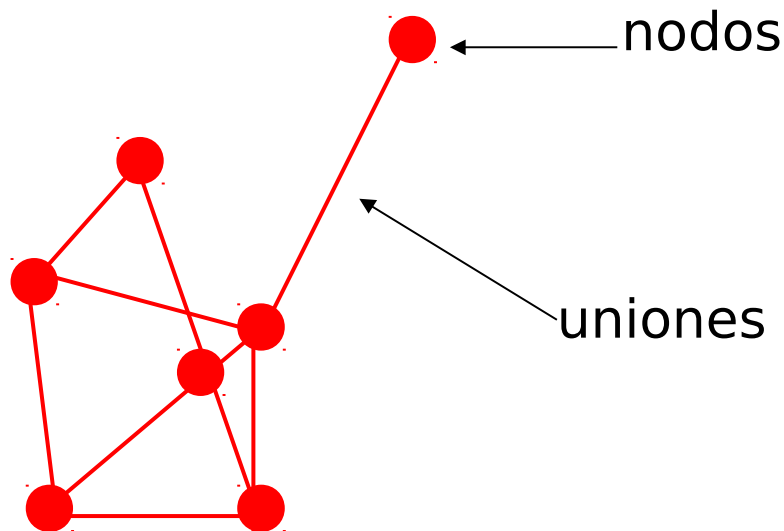
A



B

Que es una red?

Una red esta compuesta por



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

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

Graphos

Un grapho no dirigido G es definido por un par de conjuntos

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

\mathcal{V} es un conjunto no vacío y numerable de elementos llamados *nodos*

y

\mathcal{E} es un conjunto no ordenado de pares de índices (distintos) *links*

El link (i,j) une los vértices i y j .

El máximo número de links en un grapho es

$$N_{\max} = \binom{N}{2} = \frac{N(N-1)}{2}$$

Un grapho con N_{\max} links es un grapho completo o completamente conexo

Sparseness - Dilucion

Supongamos una red conexas (todo nodo puede conectarse a todo otro nodo del sistema)

El numero de links va entre $(N-1)$ a $\binom{N}{2}$

La dilucion podria definirse como (sea E el numero de links)

$$\alpha / E \propto N^\alpha$$

Es decir , el scaling del numero de links con el numero de nodos

Conectancia o densidad

$$D = \frac{E}{[N(N-1)/2]}$$

un grapho es diluido si la densidad D es $\ll 1$

Subgraphos

Un grapho $G' = (V', \mathcal{E}')$ es un subgrapho de $G = (V, \mathcal{E})$ si :

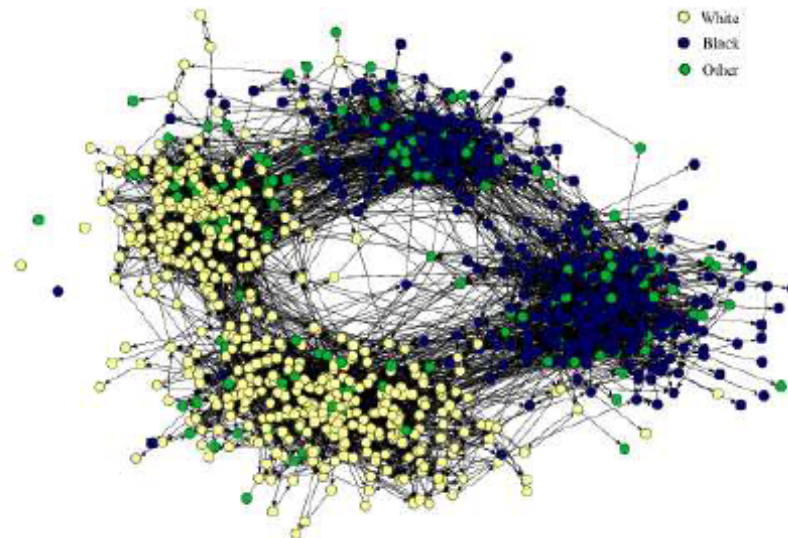
- a) todos los nodos en V' pertenecen a $V \Rightarrow V' \subset V$
- b) todos los links en \mathcal{E}' pertenecen a $\mathcal{E} \Rightarrow \mathcal{E}' \subset \mathcal{E}$

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.

os nodos :

a) en sistemas sociales suelen representar personas

b) en sistemas tecnológicos como el de transmisión de electricidad representan estaciones, subestaciones, etc.

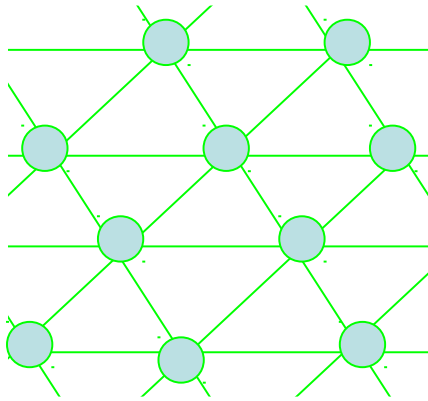
c) en sistemas de información como el WWW representan paginas

as uniones :

En a) pueden representar amistad, enemistad, lazos familiares, etc.

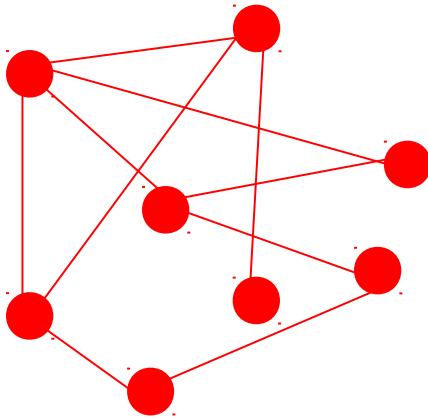
En b) representan las líneas de transmisión entre nodos

En c) los links



Red regular

-



Red "aleatoria"

Tipos de Networks

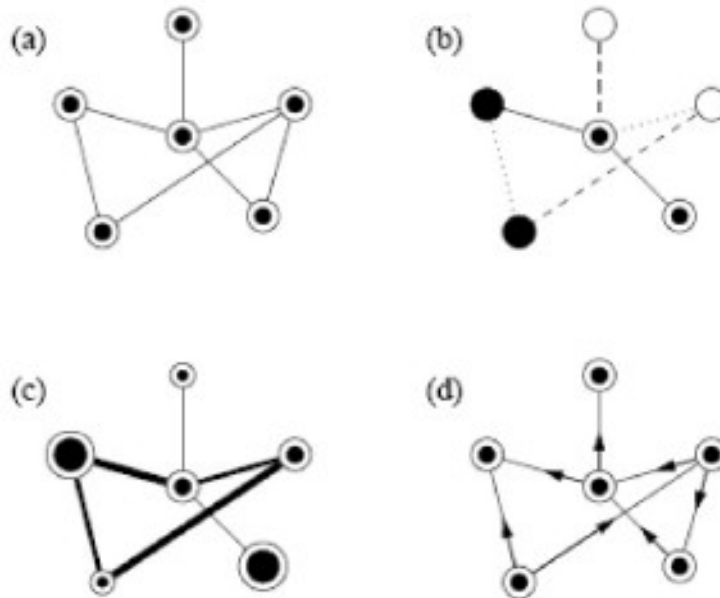


FIG. 3 Examples of various types of networks: (a) an undirected network with only a single type of vertex and a single type of edge; (b) a network with a number of discrete vertex and edge types; (c) a network with varying vertex and edge weights; (d) a directed network in which each edge has a direction.

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 $A_{ij} = A_{ji}$

Networks sociales

Networks tecnologicos

Networks de informacion

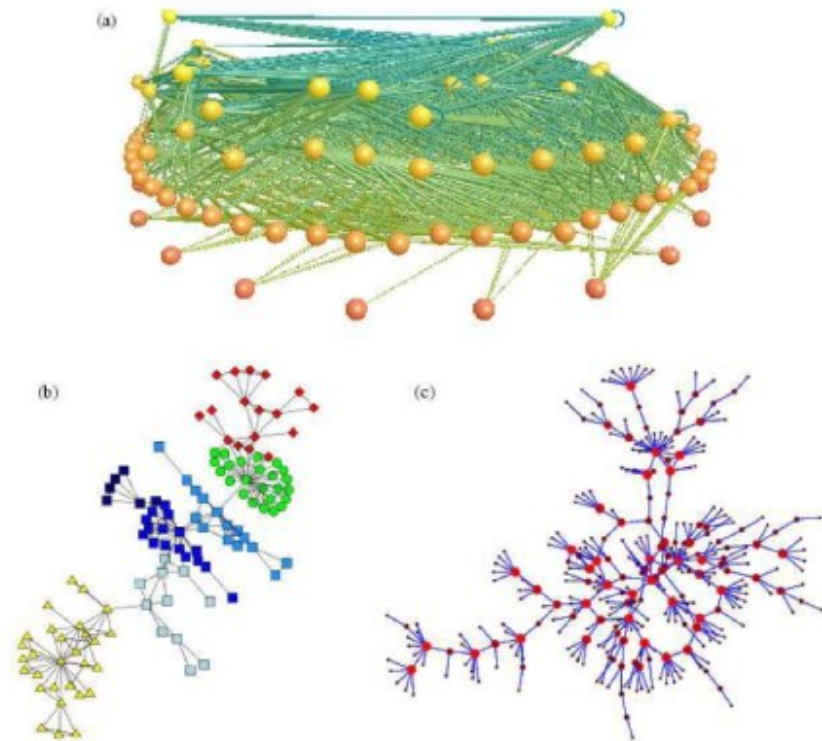
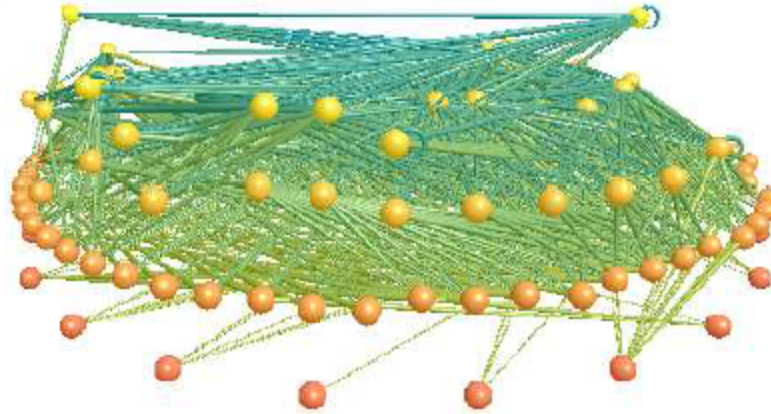
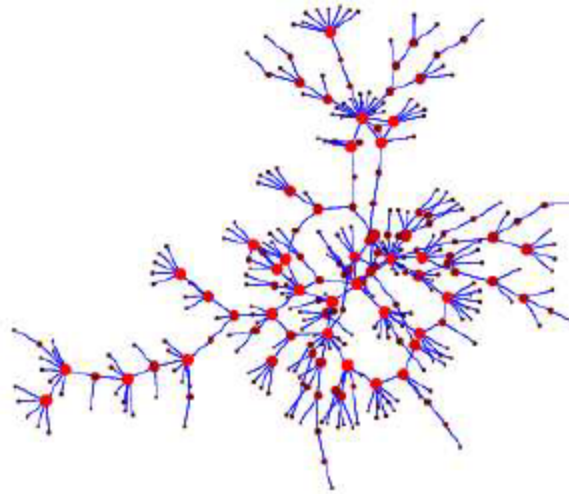


FIG. 2 Three examples of the kinds of networks that are the topic of this review. (a) A food web of predator-prey interactions between species in a freshwater lake [272]. Picture courtesy of Neo Martinez and Richard Williams. (b) The network of collaborations between scientists at a private research institution [171]. (c) A network of sexual contacts between individuals in the study by Poterat *et al.* [342].

Otros ejemplos

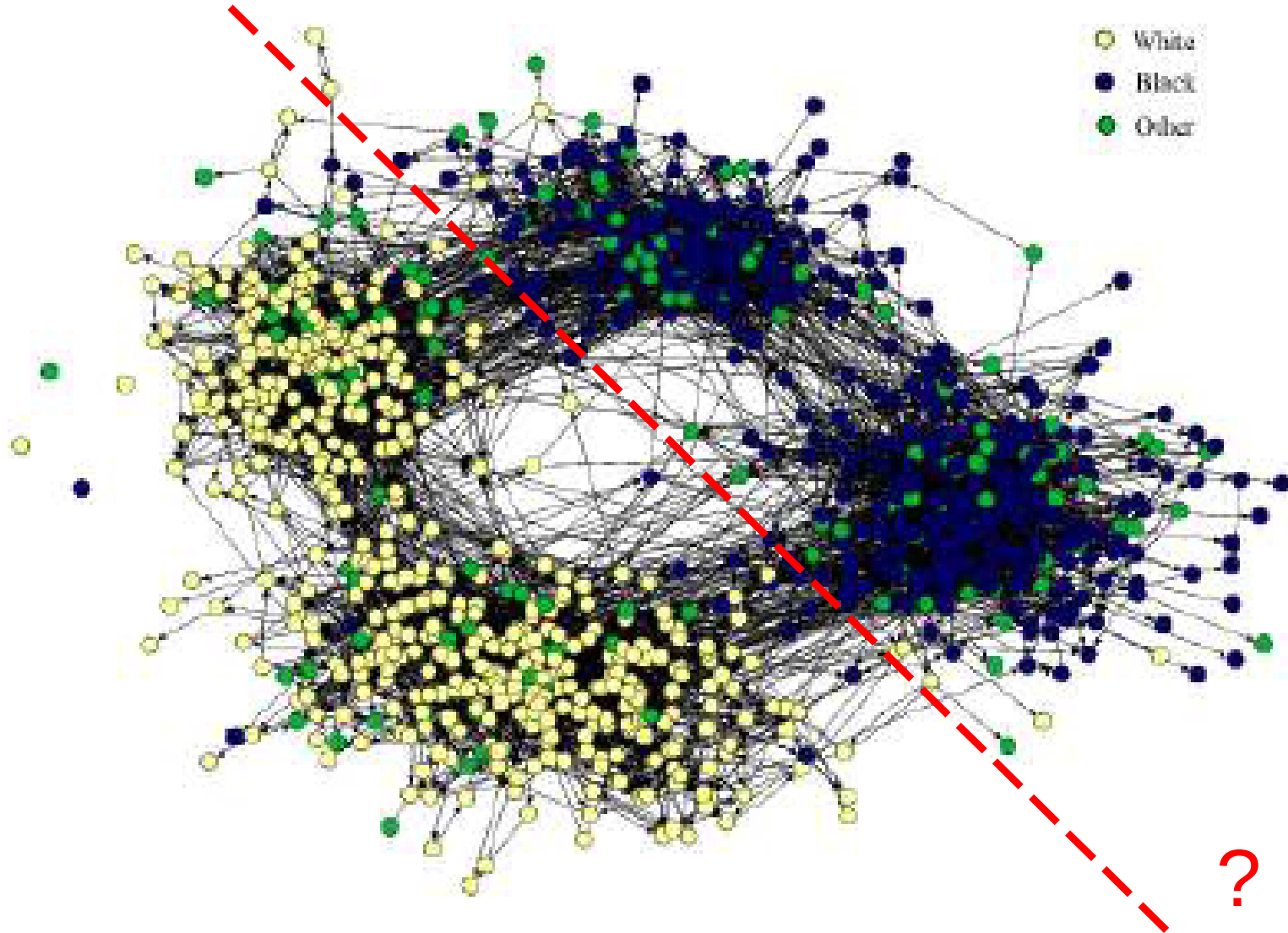


Food web
(Lago Little Rock)



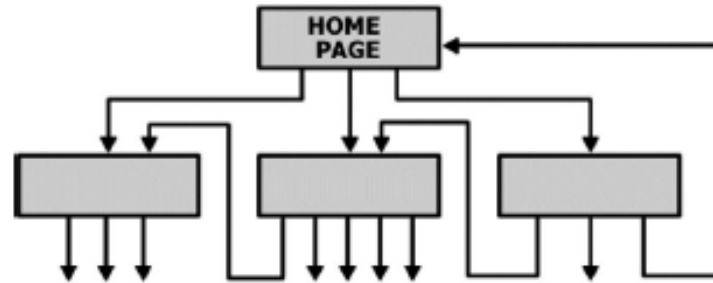
Contactos
Sexuales
(etapa epidemica
Temprana HIV Colorado
Springs)

Red de amistades en una escuela (USA)



Ejemplos familiares a todos

WORLD-WIDE WEB



INTERNET



Elementos de una red compleja

Vertex (pl. vertices): The fundamental unit of a network, also called a site (physics), a node (computer science), or an actor (sociology).

Edge: The line connecting two vertices. Also called a bond (physics), a link (computer science), or a tie (sociology).

Directed/undirected: An edge is directed if it runs in only one direction (such as a one-way road between two points), and undirected if it runs in both directions. Directed edges, which are sometimes called arcs, can be thought of as sporting arrows indicating their orientation. A graph is directed if all of its edges are directed. An undirected graph can be represented by a directed one having two edges between each pair of connected vertices, one in each direction.

Degree: The number of edges connected to a vertex. Note that the degree is not necessarily equal to the number of vertices adjacent to a vertex, since there may be more than one edge between any two vertices. In a few recent articles, the degree is referred to as the “connectivity” of a vertex, but we avoid this usage because the word connectivity already has another meaning in graph theory. A directed graph has both an in-degree and an out-degree for each vertex, which are the numbers of in-coming and out-going edges respectively.

Component: The component to which a vertex belongs is that set of vertices that can be reached from it by paths running along edges of the graph. In a directed graph a vertex has both an in-component and an out-component, which are the sets of vertices from which the vertex can be reached and which can be reached from it.

Geodesic path: A geodesic path is the shortest path through the network from one vertex to another. Note that there may be and often is more than one geodesic path between two vertices.

Diameter: The diameter of a network is the length (in number of edges) of the longest geodesic path between any two vertices. A few authors have also used this term to mean the average geodesic distance in a graph, although strictly the two quantities are quite distinct.

Propiedades fundamentales de los complex networks

Mundo pequeño

(el mundo es un pañuelo)

Experimento de Milgram (1967):

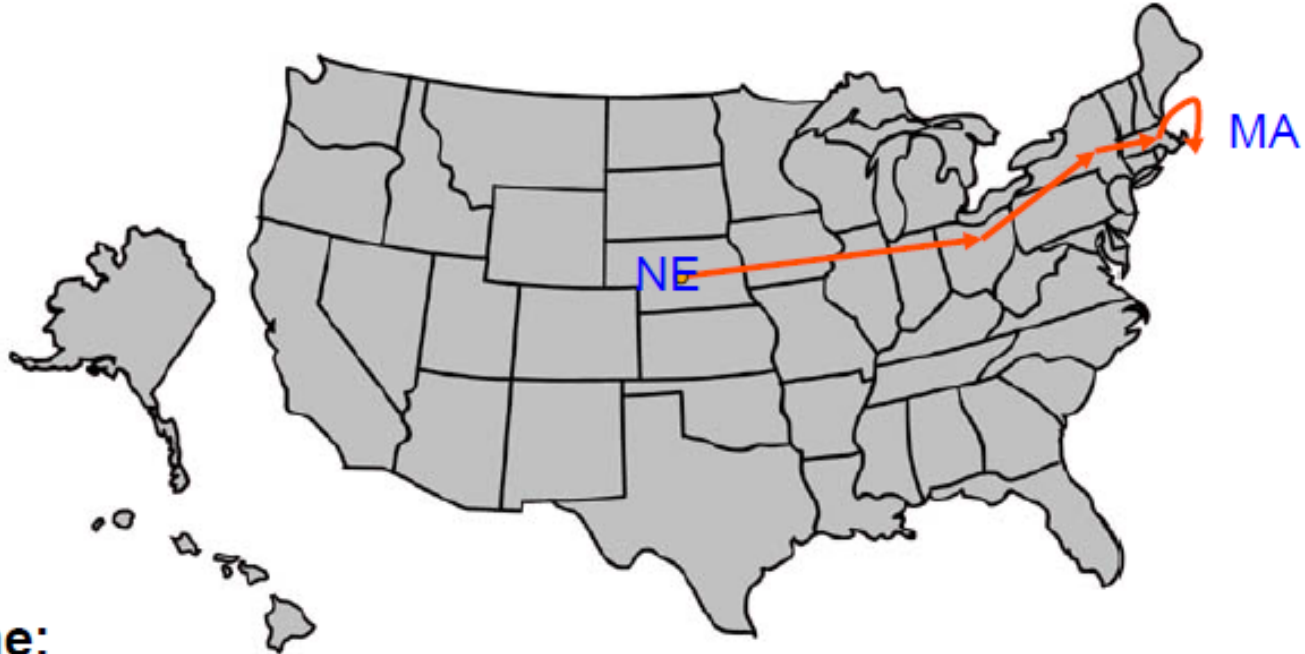
Envió paquetes a personas elegidas al azar en el medio oeste De USA y les pidió que se los hiciesen llegar a personas en Boston Pero debían hacerlo a través de personas que conociesen por el Nombre (debían llegar por amigos de amigos de amigos...).

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 :

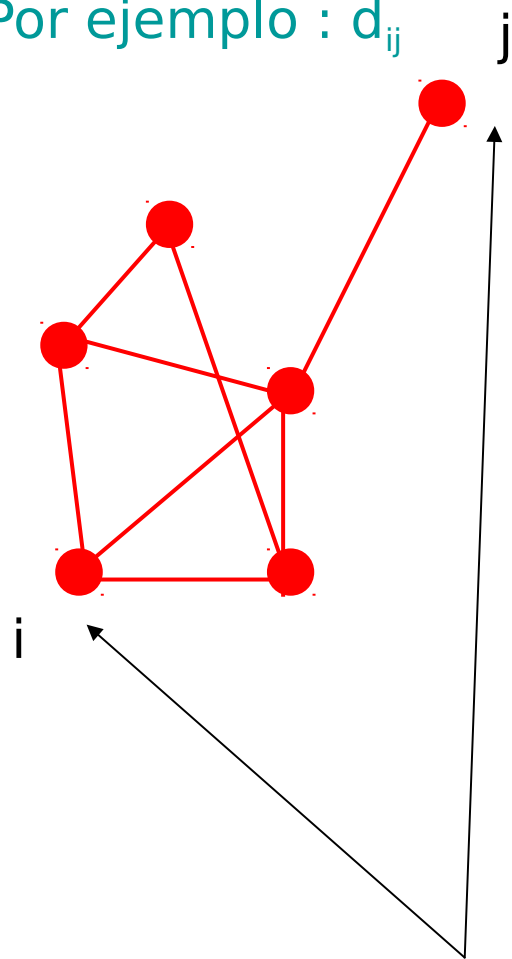
Suma sobre todos
Los pares de nodos

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

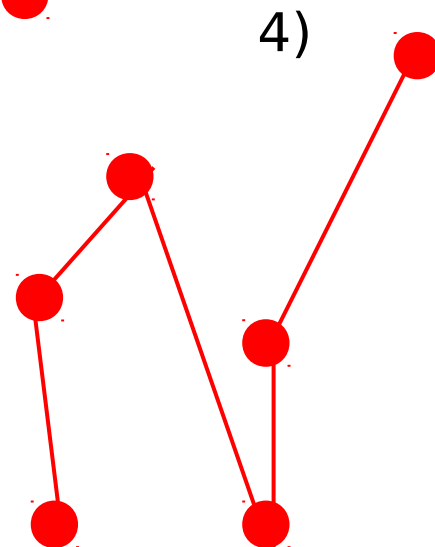
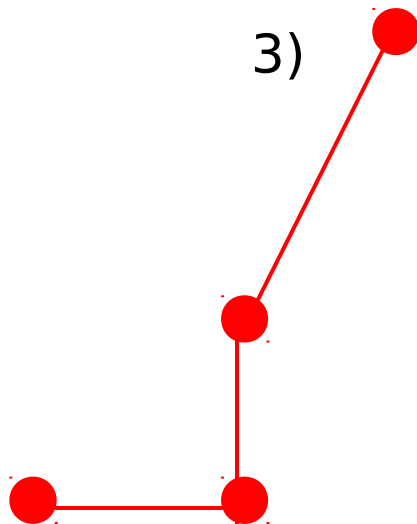
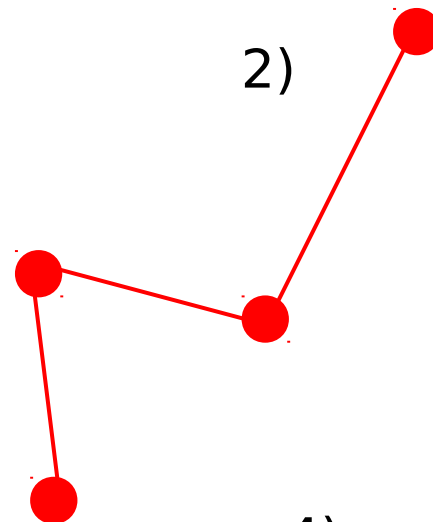
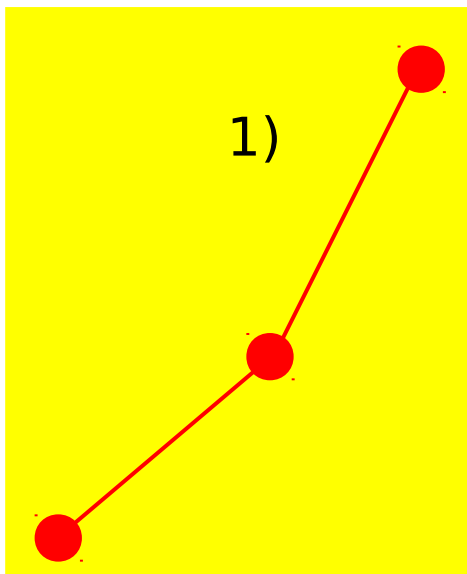
d_{ij} es el camino mínimo entre los nodos i y j

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

Por ejemplo : d_{ij}



Nodos de
interés



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?

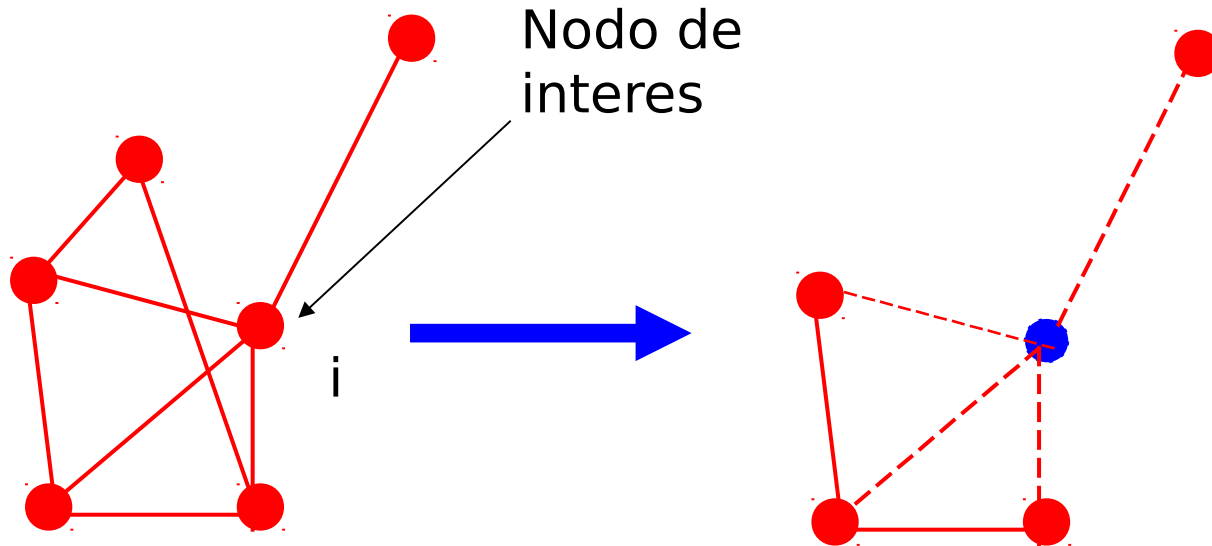
$$C_i = \frac{E_i}{k_i(k_i - 1) / 2}$$

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



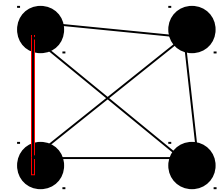
Numero de nodos vecinos al de interes : 4

Numero máximo de uniones de 4 nodos : 6

Uniones presentes

: 2

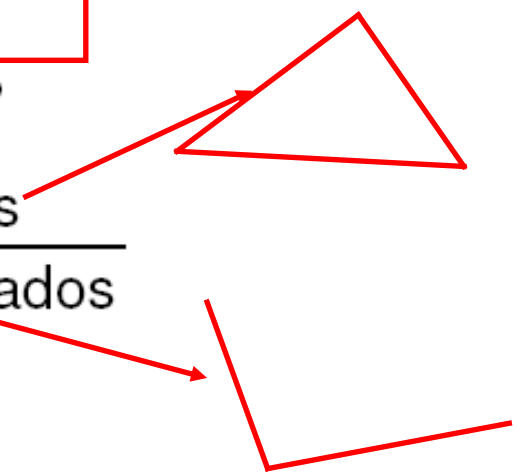
→ $C_i = 2/6$



Otras definiciones

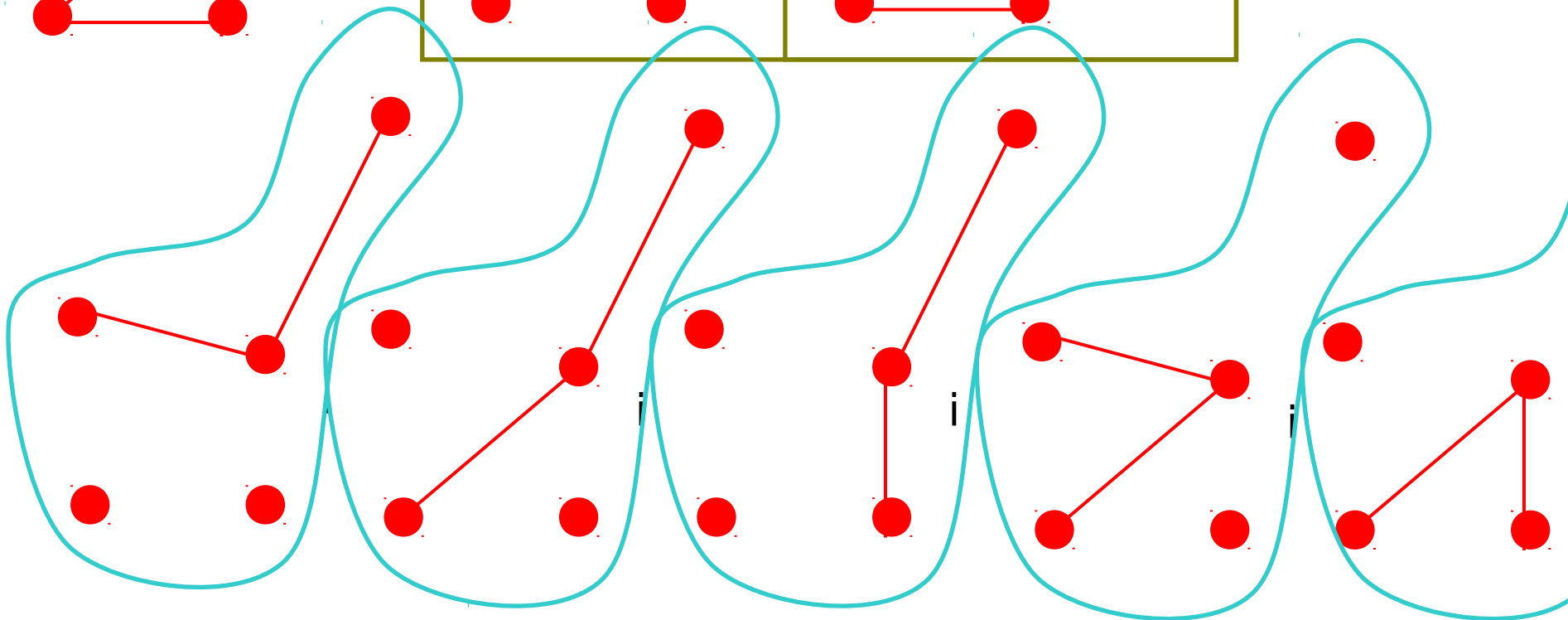
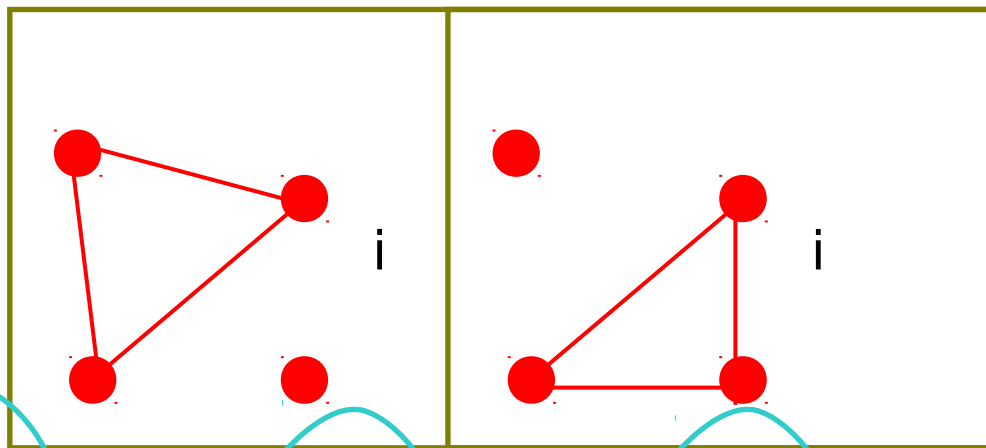
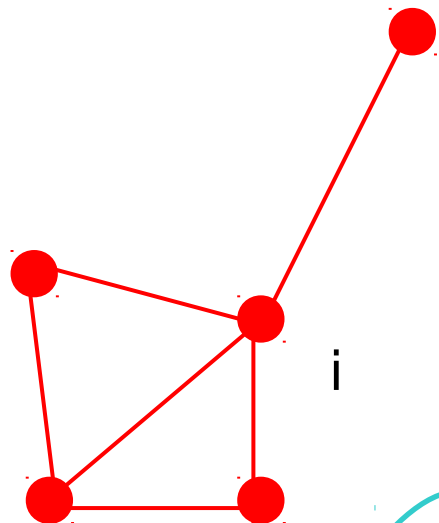
Son los amigos de mis amigos amigos entre si?

$$C = \frac{3 \cdot \text{numero de triangulos}}{\text{numero de tripletes conectados}}$$



En terminos de variables locales se define

$$C_i = \frac{\text{numero de triangulos conectados al nodo } i}{\text{numero de tripletes conectados al nodo } i}$$



+ 1 que no entra

O tambien: sea el nodo i , sea k_i el numero de nodos conectados a i . Si estos nodos forman un clique (están totalmente conectados entre si) tendríamos $k_i(k_i - 1)/2$ links entonces

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

Con E_i el numero de links presentes.

De aqui el C esta definido por

$$C = \frac{1}{n} \sum C_i$$

operativamente

Sea N =numero de nodos

Sea k =numero total de links que salen de un dado nodo

Sea d_s =grado total de un grafo

Calcularemos C

$$C = \frac{1}{n_g} \sum_g \left\{ \frac{1}{N_g} \sum_i \frac{\sum_{i \neq j} \sum_{k \neq \{i,j\}} a_{ij} \cdot a_{ik} \cdot a_{jk}}{d_i(d_i - 1)} \right\}$$

Donde a_{ij} es el elemento de matriz de la matriz de adyacencias \bar{A}_i resultante de la particion considerada.

Donde d_i es el grado del nodo i

Donde N_g es el numero de nodos en la particion g

Donde \sum_g es la suma sobre las particiones

Luego es el promedio sobre todos los nodos de la cantidad

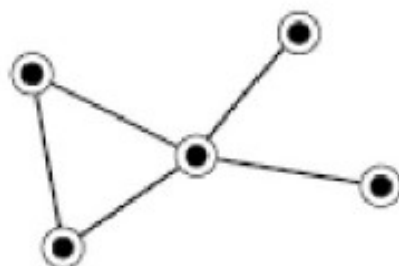


FIG. 5 Illustration of the definition of the clustering coefficient C , Eq. (3). This network has one triangle and eight connected triples, and therefore has a clustering coefficient of $3 \times 1/8 = \frac{3}{8}$. The individual vertices have local clustering coefficients, Eq. (5), of 1, 1, $\frac{1}{2}$, 0 and 0, for a mean value, Eq. (6), of $C = \frac{13}{30}$.

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

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 values in the ℓ and C 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}
MEDLINE co-authorship	1 520 251	18.1	4.6	4.91	0.066	1.1×10^{-5}
SPIRES co-authorship	56 627	173	4.0	2.12	0.726	0.003
NCSTRL co-authorship	11 994	3.59	9.7	7.34	0.496	3×10^{-4}
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
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

Ejemplo

Respecto de los actores:

<http://oracleofbacon.org/>

Bacon number : distancia mínima de un actor a Kevin Bacon



jorge porcel tiene un Bacon number de 2.

Jorge Porcel actuó en *Carlito's Way* (1993) con Sean Penn (I)

Sean Penn (I) actuó en *Mystic River* (2003) con Kevin Bacon

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

c) Distribuciones de grado

El grado de un nodo es el numero de links que estan conectados a un dado nodo.

Sea p_k la fraccion de vertices (nodos) que tienen grado k en un grafo con N nodos (vertices)

$$p_k = \frac{\# \text{ nodos con grado } k}{N}$$

Dado \vec{p}_k podemos binarlo (de alguna forma) y graficarlo. Pero los k grandes tendran poca estadistica. Ademas las colas a la derecha son largas(mas largas que Poisson), por eso es dificil medirlos (remember percolation) , hay dos posibles manera de tratar de resolver esto

- a) usar binses 1, 2 – 3, 4 – 7, 8 – 15 , que es 1, 2, 4, 8 o sea 2^0 , 2^1 , 2^2 , 2^3 pues lo vamos a estudiar en una escala logaritmica
- b) otra forma es estudiar la cumulativa

$$P_k = \sum_{k'=k}^{\infty} P_{k'}$$

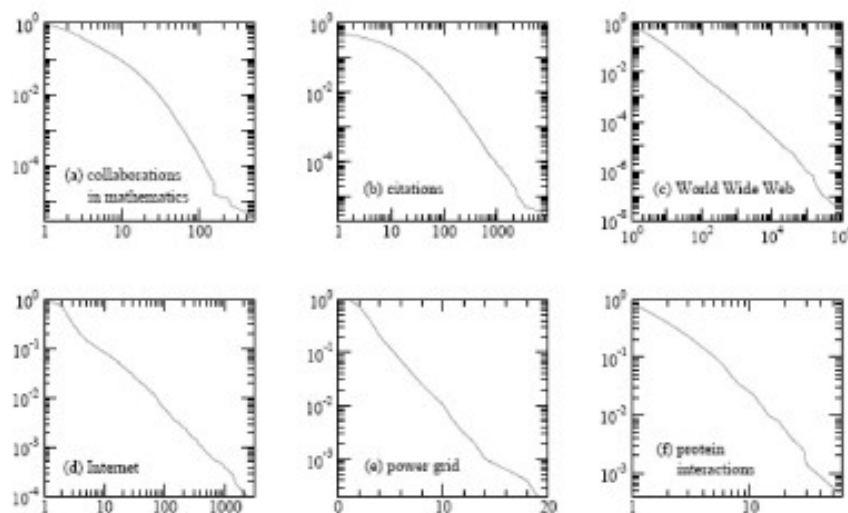


FIG. 6 Cumulative degree distributions for six different networks. The horizontal axis for each panel is vertex degree k (or in-degree for the citation and Web networks, which are directed) and the vertical axis is the cumulative probability distribution of degrees, i.e., the fraction of vertices that have degree greater than or equal to k . The networks shown are: (a) the collaboration network of mathematicians [182]; (b) citations between 1981 and 1997 to all papers cataloged by the Institute for Scientific Information [351]; (c) a 300 million vertex subset of the World Wide Web, circa 1999 [74]; (d) the Internet at the level of autonomous systems, April 1999 [86]; (e) the power grid of the western United States [416]; (f) the interaction network of proteins in the metabolism of the yeast *S. Cerevisiae* [212]. Of these networks, three of them, (c), (d) and (f), appear to have power-law degree distributions, as indicated by their approximately straight-line forms on the doubly logarithmic scales, and one (b) has a power-law tail but deviates markedly from power-law behavior for small degree. Network (e) has an exponential degree distribution (note the log-linear scales used in this panel) and network (a) appears to have a truncated power-law degree distribution of some type, or possibly two separate power-law regimes with different exponents.

Algunas de estas distribuciones tiene power laws (c, d, f parecer ser de este tipo)

Otras son exponenciales (ver las escalas en e)

Pero estas cosas no son del todo "limpias"

Tomar en cuenta que

$$P_k \sim \sum_{k'=k}^{\infty} P_{k'} \sim k^{-(\alpha-1)}$$

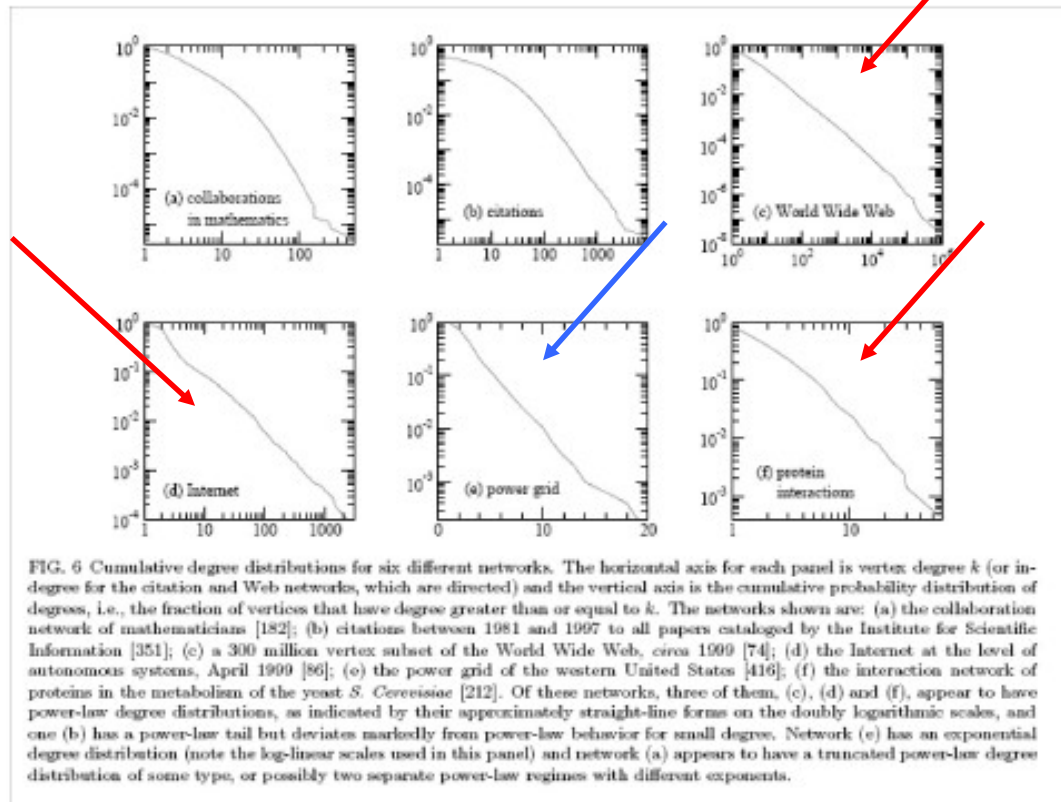
$$\int_{x_0}^{\infty} x^{-\alpha} dx =$$

Para la exponencial

$$P_k \sim \sum_{k'=k}^{\infty} P_{k'} \sim \sum_{k'=k}^{\infty} \exp(-k'/\kappa) \sim \exp(-k'/\kappa)$$

- a) usar binses 1, 2 – 3, 4 – 7, 8 – 15 , que es 1, 2, 4, 8 o sea 2^0 , 2^1 , 2^2 , 2^3 pues lo vamos a estudiar en una escala logaritmica
- b) otra forma es estudiar la cumulativa

$$P_k = \sum_{k'=k}^{\infty} P_{k'}$$



Algunas de estas distribuciones tiene power laws (c, d, f parecer ser de este tipo)

Otras son exponenciales (ver las escalas en e)

Pero estas cosas no son del todo "limpias"

Tomar en cuenta que

$$P_k \sim \sum_{k'=k}^{\infty} P_{k'} \sim k^{-(\alpha-1)}$$

$$\int_{x_0}^{\infty} x^{-\alpha} dx =$$

Para la exponencial

$$P_k \sim \sum_{k'=k}^{\infty} P_{k'} \sim \sum_{k'=k}^{\infty} \exp(-k'/\kappa) \sim \exp(-k'/\kappa)$$

d) Resistencia de un Network

Cuan robusto es un network ante la eliminacion de links/nodes?

Como es la remocion de links/nodes

- a) random
- b) por alguna propiedad relevante
 - i) grado
 - ii) betweenness

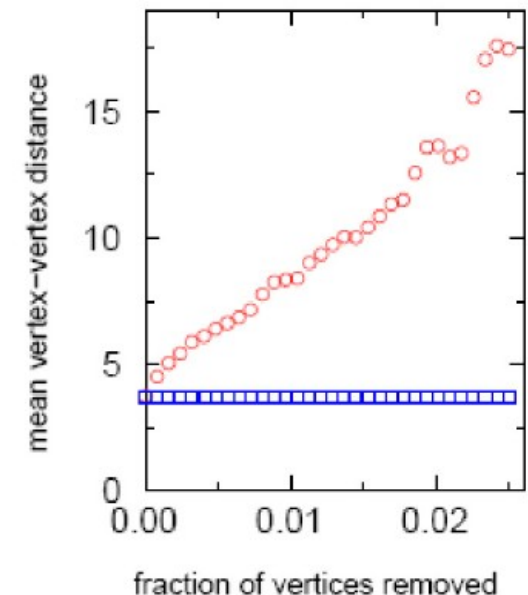
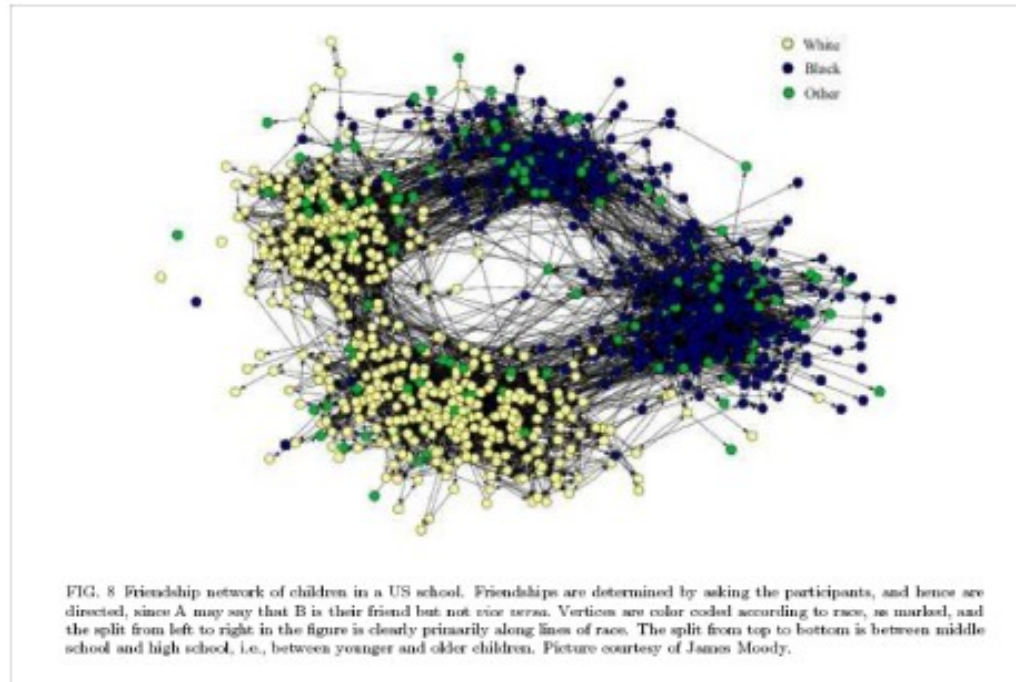


FIG. 7 Mean vertex-vertex distance on a graph representation of the Internet at the autonomous system level, as vertices are removed one by one. If vertices are removed in random order (squares), distance increases only very slightly, but if they are removed in order of their degrees, starting with the highest degree vertices (circles), then distance increases sharply. After Albert *et al.* [15].

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



Respecto de las comunidades

El analisis 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

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

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

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 $\frac{k_\omega}{2m}$

Given a partition \mathbf{P} comprising g subsets, a matrix \mathbf{e} (of dimension $g * g$) is defined such that the corresponding component e_{ij} is the fraction of edges in the original network that connect nodes in subset i with nodes in subset j . In I the *modularity* Q is defined as

$$Q = \sum_i e_{ii} - \sum_{ijk} e_{ij}e_{ki} = \text{Tr}\mathbf{e} - |\mathbf{e}^2|. \quad (1)$$

Q stands for the difference between the relative quantity of links within subsets and the expected relative number of links that would result from a random placement of links when no attention is given to the community structure of the network under consideration [9].

If the network under consideration has no community structure, Q equals 0. On the other hand, if the network under consideration does have a community structure, the closer the chosen partition is to the actual community structure of the network, the larger the modularity Q will be.

In this way, the search of community structures in networks is reduced to finding the partition \mathbf{P} which maximizes the modularity Q .

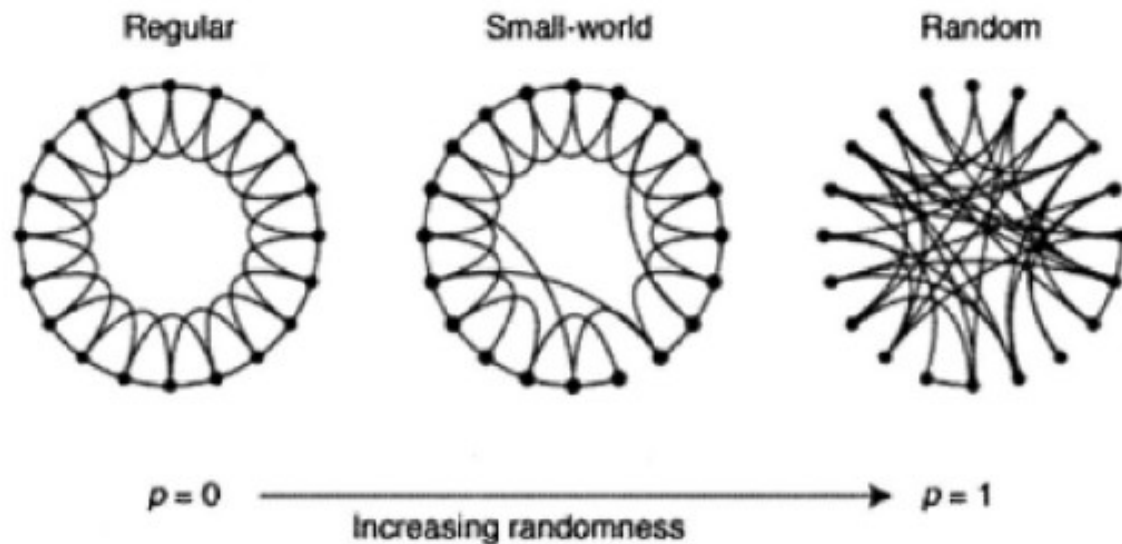


FIG. 15. The random rewiring procedure of the Watts-Strogatz model, which interpolates between a regular ring lattice and a random network without altering the number of nodes or edges. We start with $N=20$ nodes, each connected to its four nearest neighbors. For $p=0$ the original ring is unchanged; 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 $= \frac{l_s}{L} = \frac{kN}{kN} = 1$

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

$$Q \simeq 0.5$$

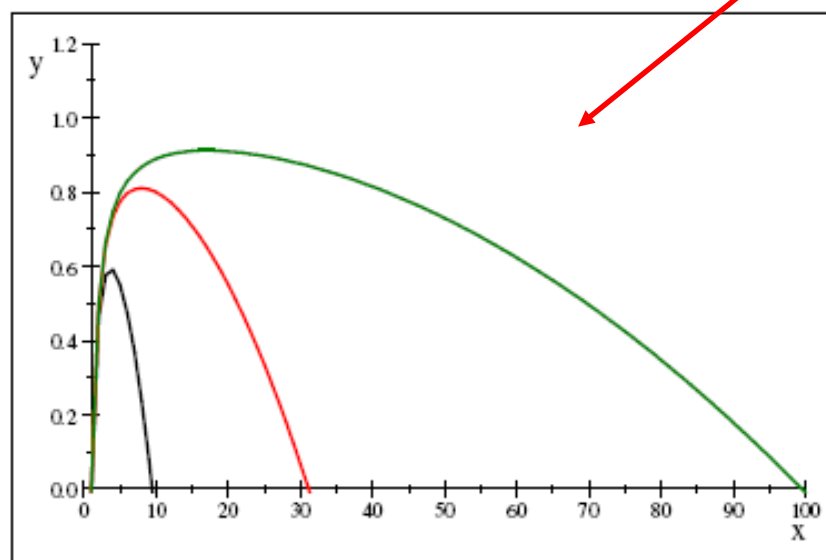
Luego Q crece mal

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

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

Simulated annealing

otro approach



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*

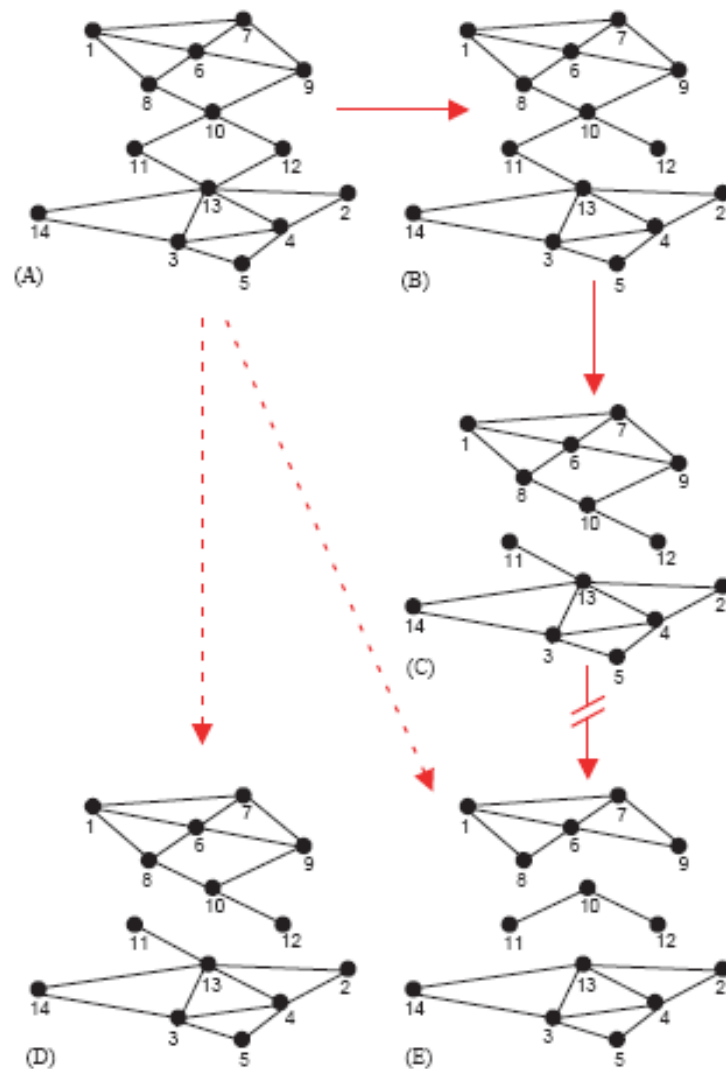
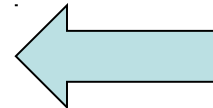
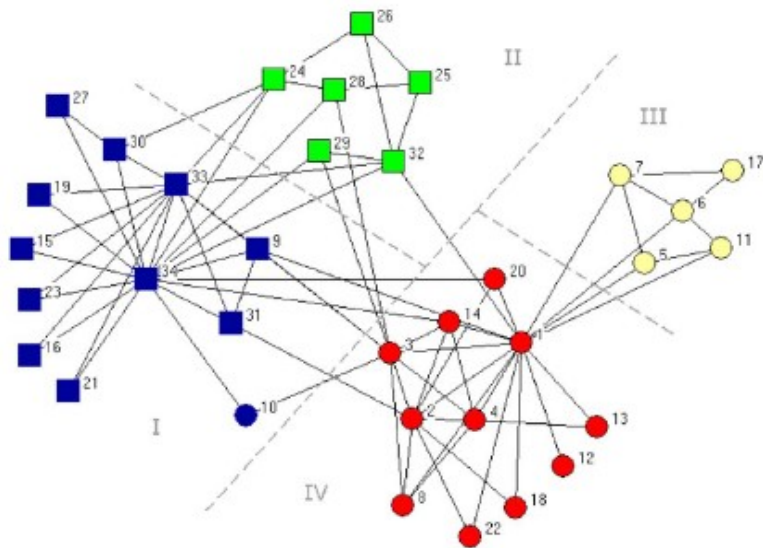


Fig. 1. Development of community structures in terms of the ER and SA analysis. Full arrows denote steps in the ER approach. Dotted arrows denote results from SA methodology. Starting from network A by applying ER methodology we first get to network B and, after the second removal of a link, to network 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 communes. 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

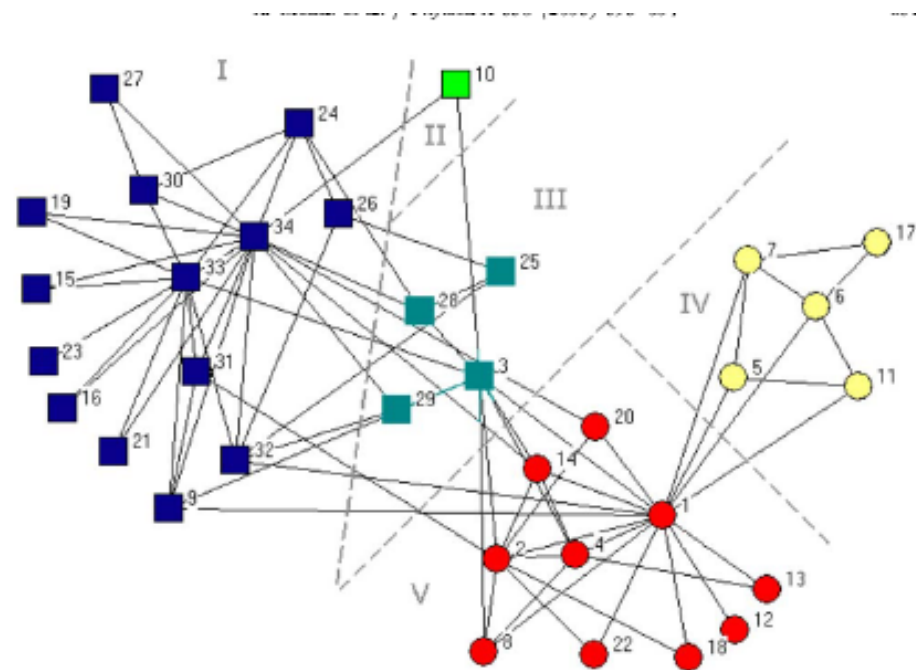


Fig. 3. Community structures for the Zachary network according to ER approach.

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

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

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

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)

Supongamos que calculamos las cosas en el límite de n muy grande, manteniendo el grado medio constante $z = p(n - 1)$, luego estamos en el caso de Poisson

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

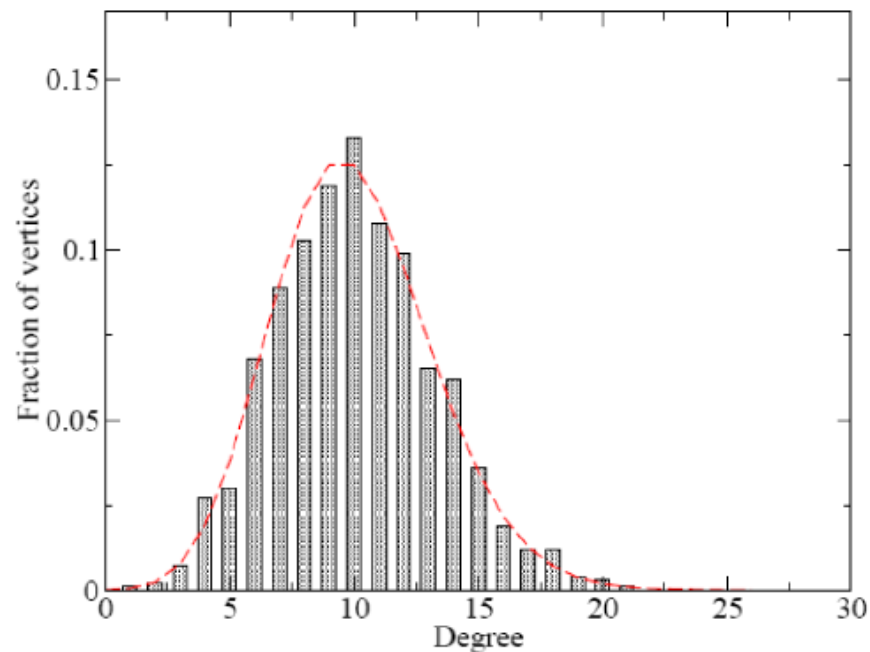


Figure 1.9. The degree distribution for an ER random network where $n = 1000$ and $p = 0.1$, with the real distribution plotted as a bar graph and the Poisson approximation plotted as the dashed line.

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)

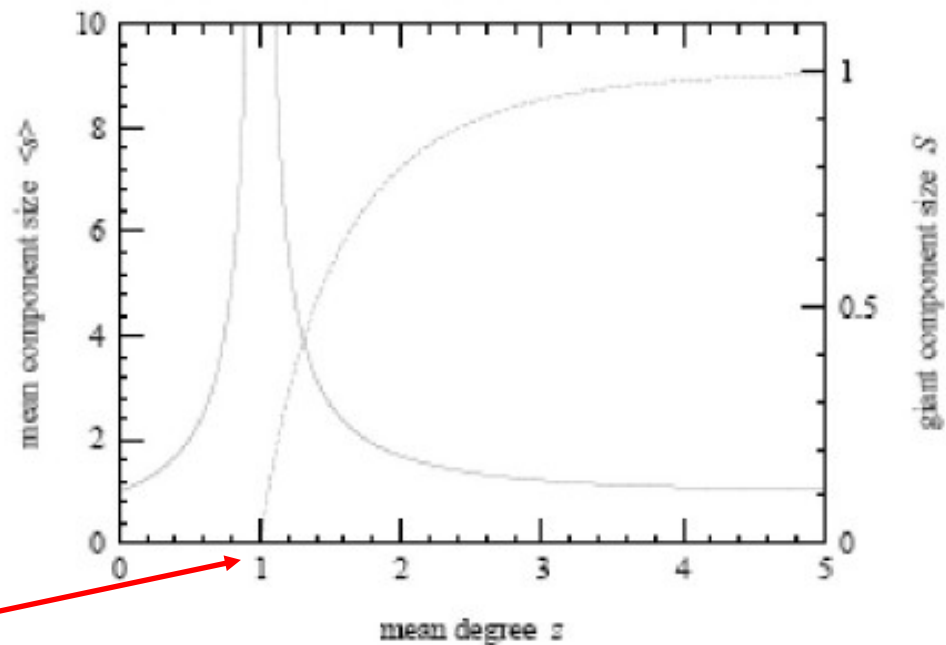


FIG. 10 The mean component size (solid line), excluding the giant component if there is one, and the giant component size (dotted line), for the Poisson random graph, Eqs. (20) and (21).

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.

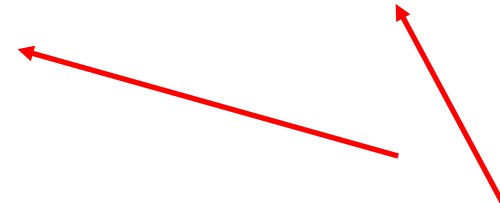
Aparecen entonces exponentes críticos

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

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

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.

Random graph generalizado

El modelo configuracional!!!!!!

- i) especificamos una distribución de grados p_k
- ii) elegimos una secuencia de grados k_i con $i = 1, 2, 3, 4, \dots$

Cada vertice i tendrá entonces k_i "palitos" saliendo

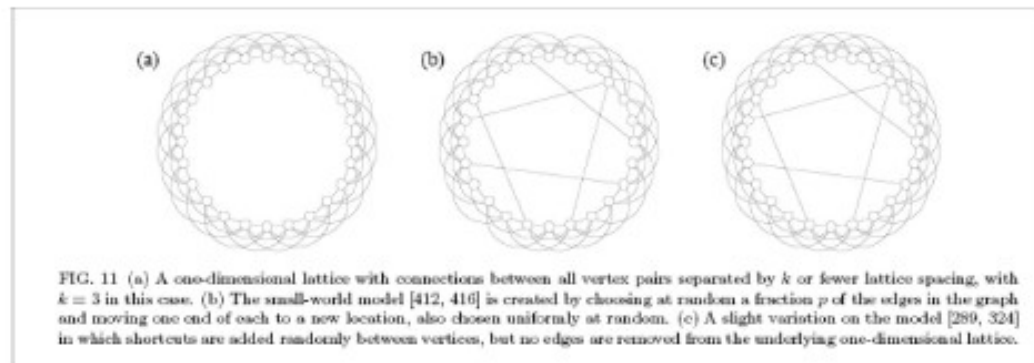
- iii) elegimos pares de "palitos" al azar y los unimos entre sí.

Se demuestra que este proceso genera todo posible grafo

El modelo configuracional queda entonces definido por el ensemble de grafos con igual probabilidad.

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

2) randomizamos con rewiring

Clustering coefficient

Entonces

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

Que es la proba de que los 3 links sobrevivan

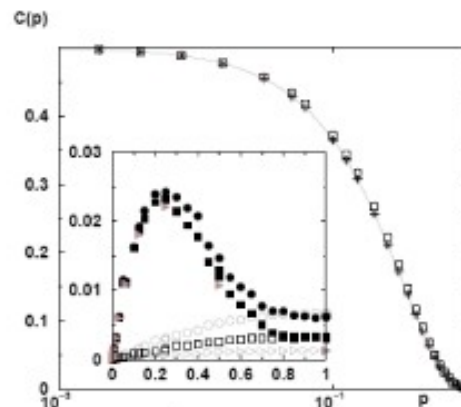


FIG. 9. $C(p)$ and $\hat{C}(p)$ versus p , for $k=2$ ($C(0) = \hat{C}(0) = 0.5$), $N = 1000, 2000, 5000$: open symbols are for $C(p)$, and the crosses are for $\hat{C}(p)$; the line is $C(0)(1-p)^3$. Inset: corrections $C(p) - C(0)(1-p)^3$ (filled symbols) for $N = 1000$ (circles), $N = 2000$ (squares)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

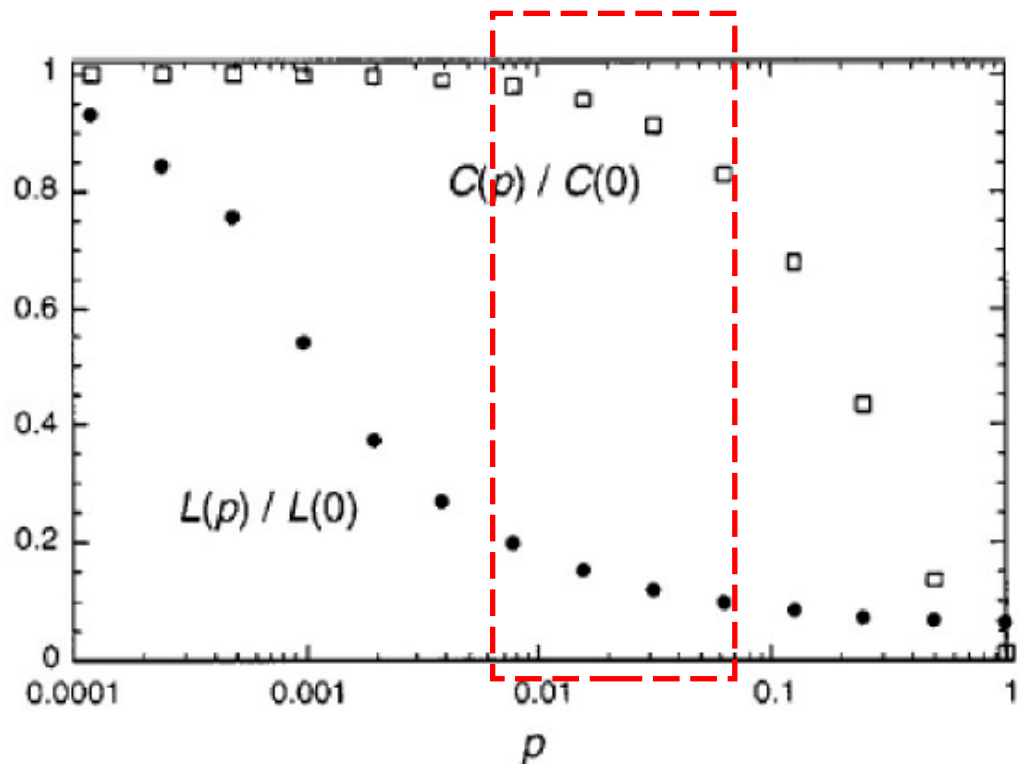


FIG. 16. Characteristic path length $\ell(p)$ and clustering coefficient $C(p)$ for the Watts-Strogatz model. The data are normalized by the values $\ell(0)$ and $C(0)$ for a regular lattice. A logarithmic horizontal scale resolves the rapid drop in $\ell(p)$, corresponding to the onset of the small-world 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 $\tau = 1$

Modelos de crecimiento de networks

Hasta ahora vimos como eran las propiedades de networks con características dadas

Ademas vimos como eran los networks reales

Como es que surgen los networks reales?

Focalizamos en el power law distribution of grades

a) Herbert Simon "rich get richer" (1955)

[69] Bornholdt, S. and Ebel, H., World Wide Web scaling exponent from Simon's 1955 model, *Phys. Rev. E* **64**, 035104 (2001).

b) Price (1965) aplico las ideas de Simon al crecimiento de networks

[344] Price, D. J. de S., A general theory of bibliometric and other cumulative advantage processes, *J. Amer. Soc. Inform. Sci.* **27**, 292–306 (1976).

Price appears to have been the first to discuss cumulative advantage specifically in the context of networks, and in particular in the context of the network of citations between papers and its in-degree distribution. His idea was that the rate at which a paper gets new citations should be proportional to the number that it already has. This is easy to justify in a qualitative way. The probability that one comes across a particular paper whilst reading the literature will presumably increase with the number of other papers that cite it, and hence the probability that you cite it yourself in a paper that you write will increase similarly. The same argument can be applied to other networks also, such as the Web. It is not clear that the dependence of citation probability on previous citations need be strictly linear, but certainly this is the simplest assumption one could make and it is the one that Price, following Simon, adopts. We now describe in detail Price's model and his exact solution of it, which uses what we would now call a *master-equation* or *rate-equation* method.

Sea un grafo dirigido

a) con n vertices,

b) con p_k la fraccion de vertices con de k 'in-degree'

$$\sum_k p_k = 1$$

c) vertices nuevos se agregan continuamente

Los vertices nuevos tendran un 'out-degree' con un valor medio m

Como deben unirse a algo, se satisface

$$\sum_k k \cdot p_k = m$$

Dado un nuevo vertice con un cierto 'out-degree' lo uniremos a los viejos vertices con una probabilidad proporcional al 'in-degree' de los mismos

Como cuando se empieza el 'in-degree' es cero se supone que el proceso anterior sera proporcional a $k + k_0$, (tomo $k_0 = 1$)

Entonces la proba de que se una a un vertice de 'in-degree' k es

$$\begin{aligned} \frac{(k+1)n_k}{\sum_k (k+1)n_k} \frac{n}{n} &= \frac{(k+1)p_k}{\sum_k (k+1)p_k} \\ &= \frac{(k+1)p_k}{\sum_k kp_k + \sum_k p_k} \\ &= \frac{(k+1)p_k}{m+1} \end{aligned}$$

El numero medio de nuevas citas a vertices con 'in-degree' k es

$$\frac{(k+1)p_k}{(m+1)} \cdot m = \frac{m}{(m+1)} \cdot (k+1)p_k$$

y esta es la perdida de nodos del tipo 'in-degree' k .

Sea $p_{k,n}$ la proba al paso con n vertices

Con esto podemos plantear para la variacion del numero de nodos con 'in-degree' k

al ir de (n) a $(n+1)$ vertices y tomando en cuenta los aportes desde $(k-1)$ y los aportes a $(k+1)$

para $k \geq 1$

$$(n + 1)p_{k,n+1} - np_{k,n} = [kp_{k-1,n} - (k + 1)p_{k,n}] \frac{m}{m + 1}$$

para $k = 0$

$$(n + 1)p_{0,n+1} - np_{0,n} = [1 - p_{0,n}] \frac{m}{m + 1}$$

La solución estacionaria corresponde a $p_{k,n+1} = p_{k,n}$

De donde

$$(n + 1)p_k - np_k = [kp_{k-1} - (k + 1)p_k] \frac{m}{m + 1}$$

$$(n + 1)p_0 - np_0 = [1 - p_0] \frac{m}{m + 1}$$

$$p_k = [kp_{k-1} - (k+1)p_k] \frac{m}{m+1}$$

$$p_0 = 1 - p_0 \frac{m}{m+1}$$

De donde para el primer caso

$$p_k = [kp_{k-1} - (k+1)p_k] \frac{m}{m+1} \Rightarrow p_k \left(\frac{m+1}{m} (k+1) \right) = kp_{k-1} \Rightarrow$$

$$p_k = p_{k-1} \frac{k}{(k+2+1/m)}$$

Para el otro

$$p_0 = \frac{m+1}{(2m+1)}$$

De este modo a partir de p_0 se construyen los otros p_k y entonces

$$p_1 = p_0 \frac{1}{(3+1/m)} = \frac{m+1}{(2m+1)} \frac{m}{(3m+1)}$$

en general

$$\begin{aligned} p_k &= p_0 \frac{k(k-1)\dots 1}{(k+2+1/m)\dots(3+1/m)} \\ &= \left(1 + \frac{1}{m}\right) B(k+1, 2+1/m) \end{aligned}$$

con $B(k+1, 2+1/m) = B(a, b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$

Si a crece y b se mantiene fijo, asintóticamente va a a^{-b} o sea

$$p_k \sim k^{-(2+1/m)}$$

qed

Observar que k_0 no aparece en la solución con lo que se confirma que haberlo elegido =1...

Albert & Barabasi

Siguen la linea de Price, pero el grafo es no dirigido
Comienzan con nodos con grado m