

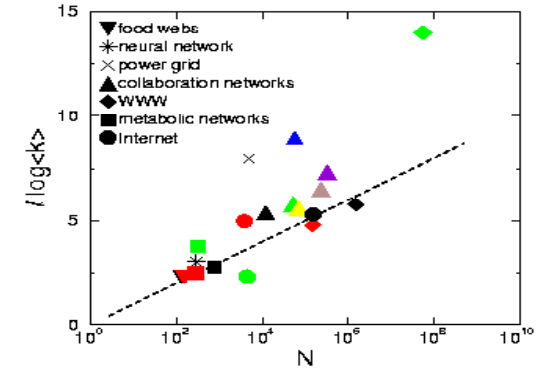
Redes Aleatorias II

Modelo configuracional y
modelos generativos

Redes ER vs Redes Reales

diametro

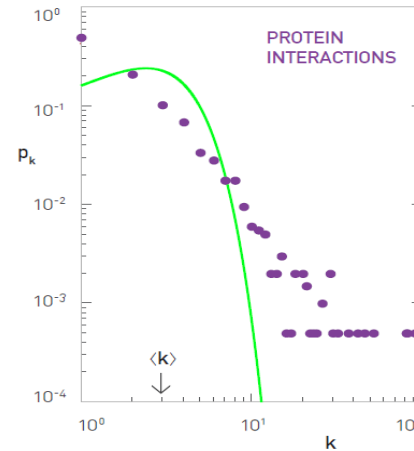
$$\langle d \rangle \sim \frac{\ln n}{n}$$



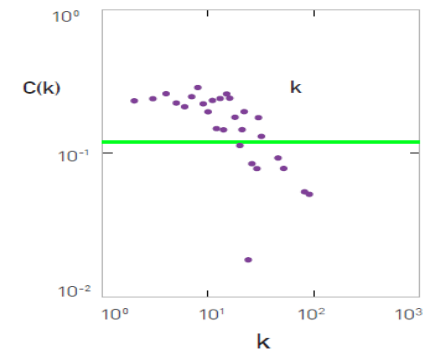
distribucion de grado

$$P_k = \binom{n-1}{k} p^k (1-p)^{n-1-k}$$

$$P_k = \frac{c^k}{k!} e^{-c}$$



(d) Protein Interactions



coef. clustering

$$C = \frac{c}{n}$$



Redes Aleatorias con distribución de grado arbitraria

Modelo configuracional:

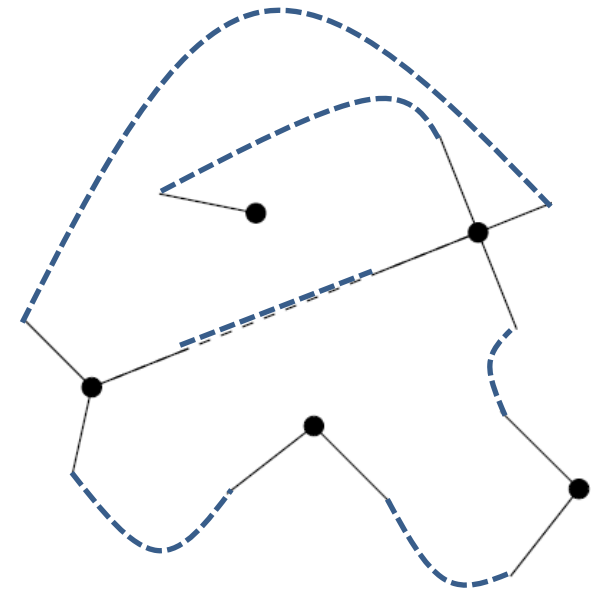
- modelo estocástico compatible con una dada **secuencia de grado** $\{k_1, k_2, \dots, k_n\}$
- Fijar una secuencia fija el número de enlaces:

$$m = \frac{1}{2} \sum_i k_i$$

por lo que sería un modelo *análogo* a $G(N, M)$

Procedimiento

- partimos de una dada secuencia de grado para n vértices:
eg $\{3, 1, 4, 2, 2\}$
- eso define, para cada nodo, una serie de **puntas** (*stubs*)
- se trata luego de **matchear** con probabilidad uniforme pares de **puntas** para terminar de definir el grafo



En el ensamble de grafos generados con el modelo configuracional cada *matching* de **puntas**, compatibles con la **secuencia de grado de interés**, ocurre con igual probabilidad. O sea, cada **punta** se conecta con cualquier otra con igual probabilidad. Esta es una característica esencial de este ensamble

Modelo configuracional

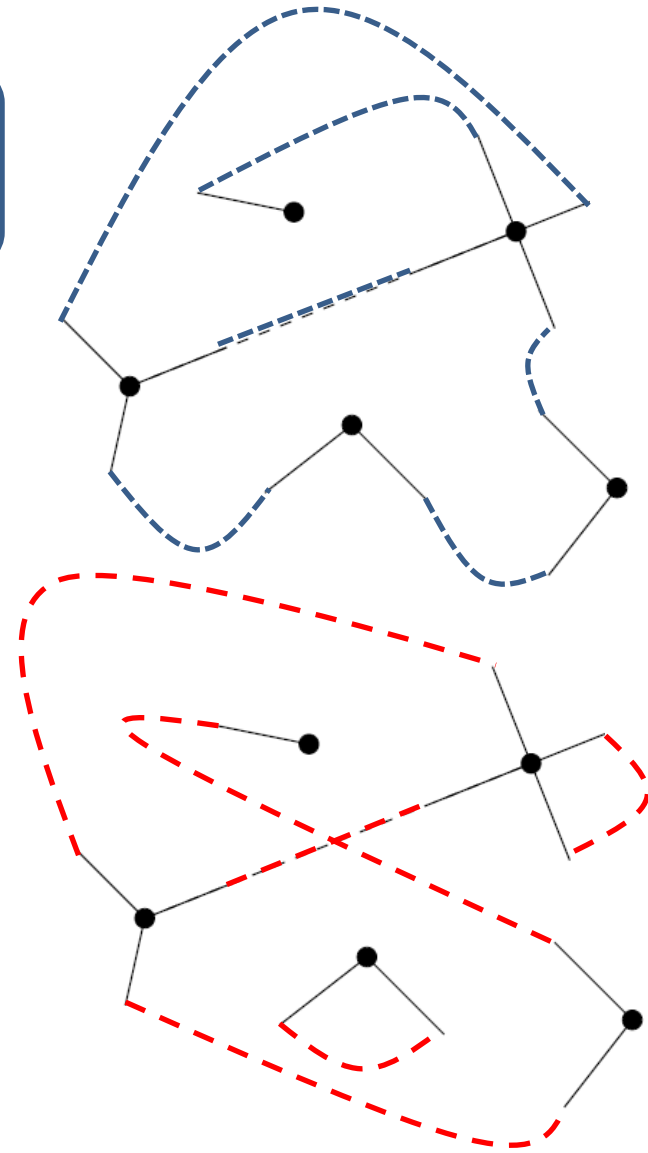
En el ensamble de grafos generados con el modelo configuracional cada *matching* de **puntas, compatibles con la secuencia de grado de interés**, ocurre con igual probabilidad. O sea, cada **punta** se conecta con cualquier otra con igual probabilidad



esto puede producir grafos con loops y/o multi-enlaces (!)

OJO

- Si uno quiere modelar grafos **simples** esto puede ser un problema: para evitar multi-enlaces y loops es necesario abandonar hipótesis de matching uniforme e independiente.
- Para redes grandes, se puede demostrar que el nro de multi-enlaces y loops es una constante, por lo que su densidad tiende a cero en ese límite
- Nosotros vamos a ignorar los efectos de su presencia y asumiremos que los resultados del modelo configuracional son directamente aplicables para grafos simples en el límite $n \gg 1$



Probabilidad de enlace

Cuál es la probabilidad p_{ij} con la que el nodo- i se conecta con el nodo- j ?



- $p_{ij} = 0$ si $k_i=0$ o $k_j=0$
- Hay $2m$ puntas en total. De las $2m-1$ disponibles (no contamos la que utilizaremos del nodo- i) k_j son las que conducen al nodo- j
- Como los *matching* de **puntas** se realizan uniformemente al azar la probabilidad de que ocurra uno entre los nodos de interes resulta

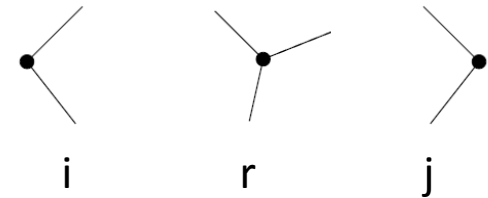
$$p_{ij} = k_i \frac{k_j}{2m - 1} \sim \frac{k_i k_j}{2m}$$

Algunas propiedades a partir de p_{ij}

Número esperado de vecinos compartidos n_{ij}

- Probabilidad de que nodo- i y nodo- j estén conectados al nodo- r :

$$\frac{k_i k_r}{2m} \frac{k_j (k_r - 1)}{2m}$$



- Número medio esperado de vecinos compartidos:

$$\begin{aligned} n_{ij} &= \sum_r \frac{k_i k_r}{2m} \frac{k_j (k_r - 1)}{2m} = \frac{k_i k_j}{2m} \sum_r \frac{k_r (k_r - 1)}{2m} \\ &= p_{ij} \frac{\sum_r k_r (k_r - 1)}{n \langle k \rangle} = p_{ij} \frac{\langle k^2 \rangle - \langle k \rangle}{\langle k \rangle} \end{aligned}$$

n_{ij} depende de p_{ij} (o sea de k_i y k_j) pero también de propiedades de la distribución de grado



Algunas propiedades más

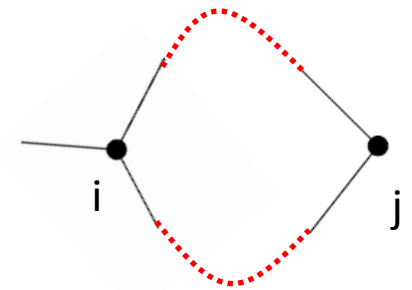
Probabilidad de tener al menos 2 enlaces:

$$p_{ij}^{(2)} = p_{ij} \frac{(k_i - 1)(k_j - 1)}{2m} = \frac{k_i k_j (k_i - 1)(k_j - 1)}{(2m)^2}$$

prob de tener un enlace



prob de tener un 2do enlace
(ya utilice dos **puntas**)



Numero medio de multi-enlaces esperados en la red:

$$\begin{aligned} n^{ME} &= \frac{1}{2} \sum_{ij} p_{ij}^{(2)} = \frac{1}{2(2m)^2} \sum_{ij} k_i k_j (k_i - 1)(k_j - 1) \\ &= \frac{1}{2\langle k \rangle^2 n^2} \sum_{ij} k_i (k_i - 1) k_j (k_j - 1) = \frac{1}{2} \left[\frac{\langle k^2 \rangle - \langle k \rangle}{\langle k \rangle} \right]^2 \end{aligned}$$

- el nro esperado no depende del tamaño de la red, se mantiene constante
- si el segundo momento de la distrib se mantiene finito: $n^{ME}/n \rightarrow 0$ cuando $n \gg 1$
- nota: para leyes de potencia n^{ME}/n puede decaer más lentamente que $1/n$

Otra propiedad...

Por qué no tenemos tantos amigos como nuestros amigos?

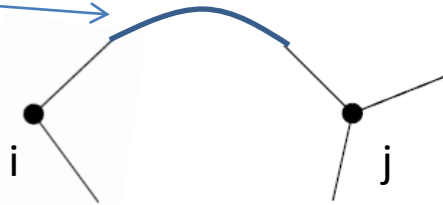


Distribución enlace-grado

- P_k es el grado esperado si elegimos un **nodo al azar** de la red
- Qué pasa si elegimos un **enlace al azar**? Sea Q_k la probabilidad de que uno de los extremos de un **enlace elegido al azar** tenga grado k .

1) elijo un enlace al azar

2) asumo que emerge del nodo- i



3) me interesa la probabilidad de que el enlace termine en un nodo- j que tenga grado k

$$\frac{k}{2m - 1}$$

← nro puntas de un nodo de grado k
← nro total de puntas (sin contar la de partida)

4) la probabilidad de que el enlace elegido termine en **cualquier** nodo de grado k

$$Q_k \sim np_k \frac{k}{2m} = \frac{kp_k}{\langle k \rangle}$$

← probabilidad de que un amigo tenga k amigos

Otra propiedad...

Por qué no tenemos tantos amigos como nuestros amigos?

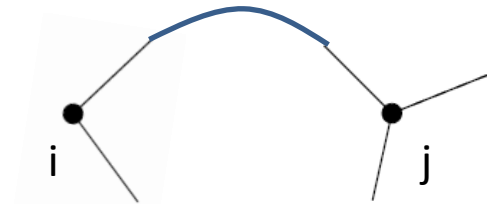


Distribución enlace-grado

- P_k es el grado esperado si elegimos un **nodo al azar** de la red
- Qué pasa si elegimos un **enlace al azar**? Sea Q_k la probabilidad de que uno de los extremos de un **enlace elegido al azar** tenga grado k .

es más probable tener amigos populares
que ser popular !!

$$Q_k = \frac{kp_k}{\langle k \rangle}$$



- Siguiendo un enlace tomado al azar es más fácil alcanzar nodos de alto grado...simplemente porque tienen más (!)
- Esto es consecuencia de la manera en que se construye las redes en el modelo configuracional. No necesariamente debe ocurrir en redes reales

Otra propiedad...

Por qué no tenemos tantos amigos como nuestros amigos?



Distribución enlace-grado

- P_k es el grado esperado si elegimos un **nodo al azar** de la red
- Qué pasa si elegimos un **enlace al azar**? Sea Q_k la probabilidad de que uno de los extremos de un **enlace elegido al azar** tenga grado k .

$$\langle k \rangle = \sum_k k P_k$$

$$\langle k \rangle_Q = \sum_k k Q_k = \sum_k k \frac{k p_k}{\langle k \rangle} = \frac{\langle k^2 \rangle}{\langle k \rangle}$$

$$\langle k \rangle_Q - \langle k \rangle = \frac{\langle k^2 \rangle}{\langle k \rangle} - \langle k \rangle = \frac{\sigma_k^2}{\langle k \rangle} > 0$$



$$\langle k \rangle_Q > \langle k \rangle$$

Otra propiedad...

Por qué no tenemos tantos amigos como nuestros amigos?



Redes de colaboración científica

Network	n	$\langle k \rangle$ Average degree	$\langle k \rangle_Q$ Average neighbor degree	$\frac{\langle k^2 \rangle}{\langle k \rangle}$
Biologists	1 520 252	15.5	68.4	130.2
Mathematicians	253 339	3.9	9.5	13.2
Internet	22 963	4.2	224.3	261.5

El modelo configuracional exagera este efecto que de todas formas puede estar presente en redes reales

Componente gigante

- Componente gigante: componente que involucra una fracción finita de la red
- Una componente gigante existe si cuando seguimos por un enlace tomado al azar hay altas chances de terminar en un nodo que tenga al menos una punta adicional que me lleve a otro nodo, que a su vez....
- La condición general para que exista CG es que

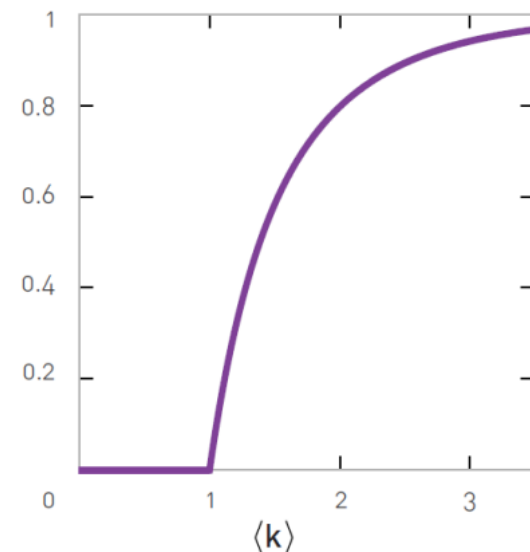
Nro medio de amigos
que tiene un amigo

$$\langle k \rangle_Q = \frac{\langle k^2 \rangle}{\langle k \rangle} \geq 2$$

Para redes ER:

$$\langle k^2 \rangle = \langle k \rangle^2 + \langle k \rangle$$

$$\langle k \rangle_Q = \frac{\langle k \rangle^2 + \langle k \rangle}{\langle k \rangle} = \langle k \rangle + 1 \geq 2 \Rightarrow \langle k \rangle \geq 1$$



Componente gigante

- Una componente gigante existe si cuando seguimos por un enlace tomado al azar hay altas chances de terminar en un nodo que tenga al menos una punta adicional que me lleve a otro nodo, que a su vez....

$$\langle k \rangle_Q = \frac{\langle k^2 \rangle}{\langle k \rangle} \geq 2$$

Para redes con cola pesada:

$$p_k = ck^{-\gamma},$$

$$\langle k^2 \rangle = c \sum_{k=1}^{\infty} k^2 k^{-\gamma} \sim \int_{x=1}^{\infty} x^{2-\gamma} dx \sim x^{3-\gamma} \Big|_{x=1}^{\infty} \stackrel{\text{si } 2 < \gamma < 3}{\downarrow} = \infty \quad (\gg \langle k \rangle)$$

Para estas redes la componente gigante **siempre existe**

Network motifs

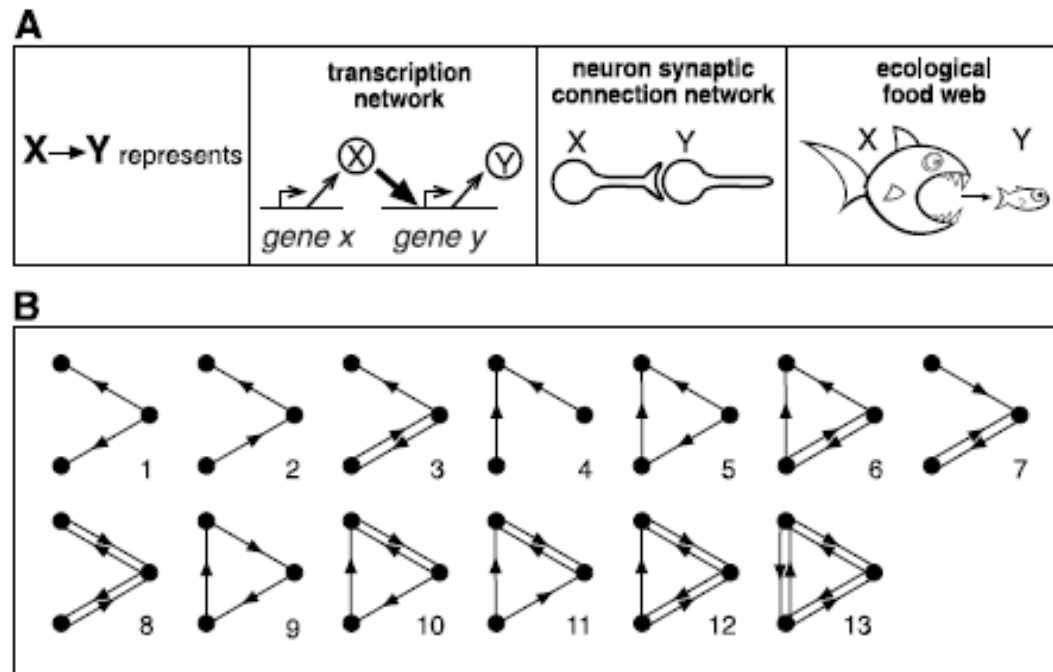
Network Motifs: Simple Building Blocks of Complex Networks

R. Milo,¹ S. Shen-Orr,¹ S. Itzkovitz,¹ N. Kashtan,¹ D. Chklovskii,²
U. Alon^{1*}

Complex networks are studied across many fields of science. To uncover their structural design principles, we defined “network motifs,” patterns of interconnections occurring in complex networks at numbers that are significantly higher than those in randomized networks. We found such motifs in networks from biochemistry, neurobiology, ecology, and engineering. The motifs shared by ecological food webs were distinct from the motifs shared by the genetic networks of *Escherichia coli* and *Saccharomyces cerevisiae* or from those found in the World Wide Web. Similar motifs were found in networks that perform information processing, even though they describe elements as different as biomolecules within a cell and synaptic connections between neurons in *Caenorhabditis elegans*. Motifs may thus define universal classes of networks. This approach may uncover the basic building blocks of most networks.

Modelando Interacciones

Fig. 1. (A) Examples of interactions represented by directed edges between nodes in some of the networks used for the present study. These networks go from the scale of biomolecules (transcription factor protein X binds regulatory DNA regions of a gene to regulate the production rate of protein Y), through cells (neuron X is synaptically connected to neuron Y), to organisms (X feeds on Y). (B) All 13 types of three-node connected subgraphs.



Son todos igual de importantes?

Detección de motivos en la red

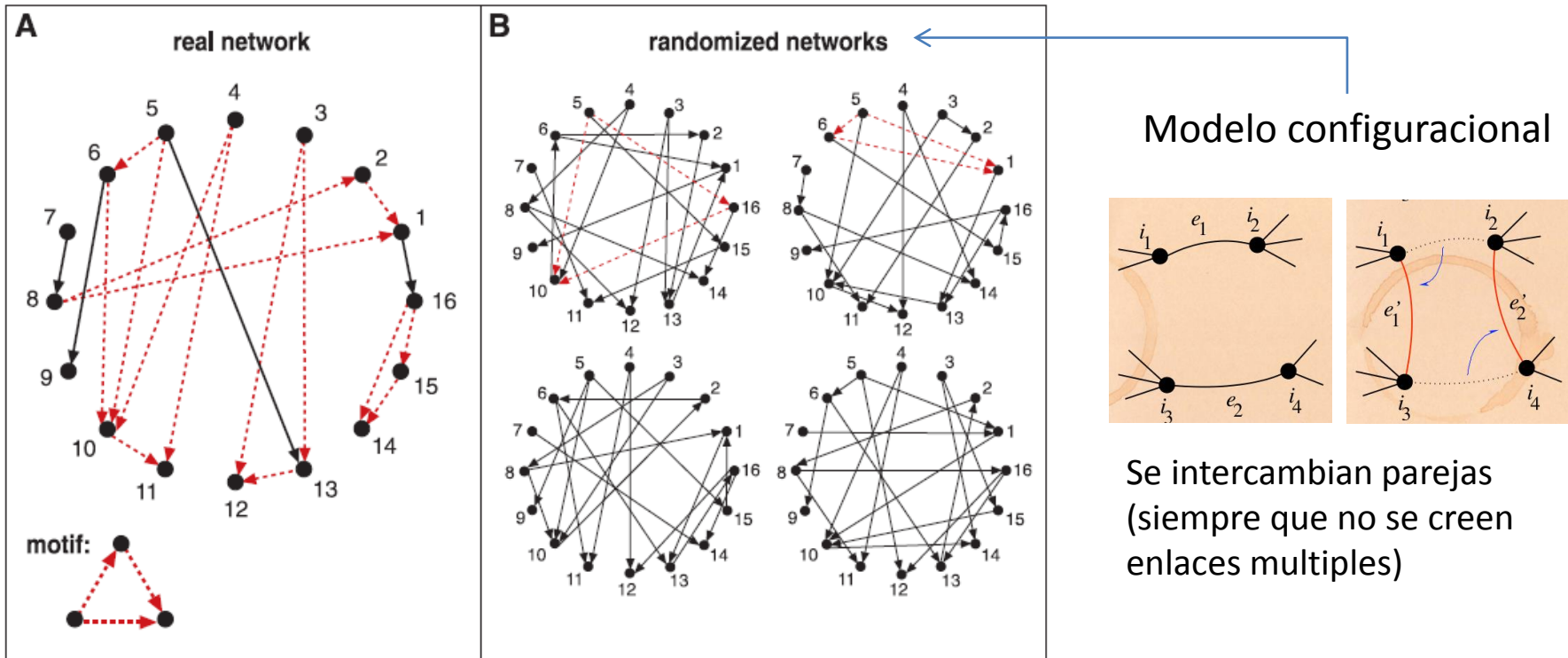


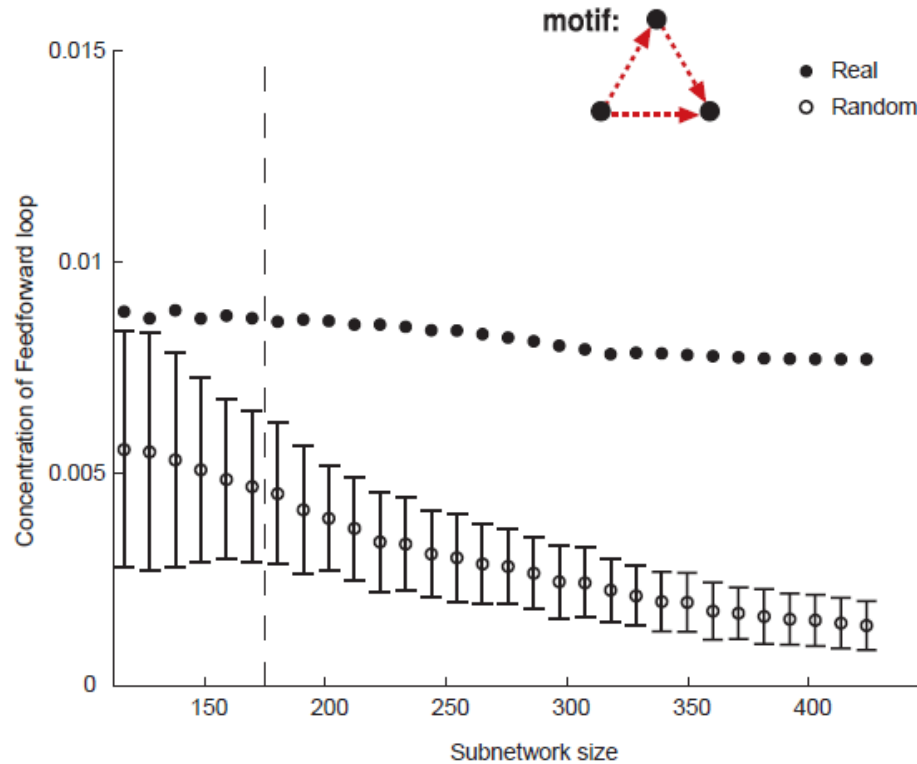
Fig. 2. Schematic view of network motif detection. Network motifs are patterns that recur much more frequently (**A**) in the real network than (**B**) in an ensemble of randomized networks. Each node in the randomized networks has the same number of incoming and outgoing edges as does the corresponding node in the real network. Red dashed lines indicate edges that participate in the feedforward loop motif, which occurs five times in the real network.

Sobrerrepresentación feed-forward

Fig. 3. Concentration C of the feedforward loop motif in real and randomized subnetworks of the *E. coli* transcription network (17).

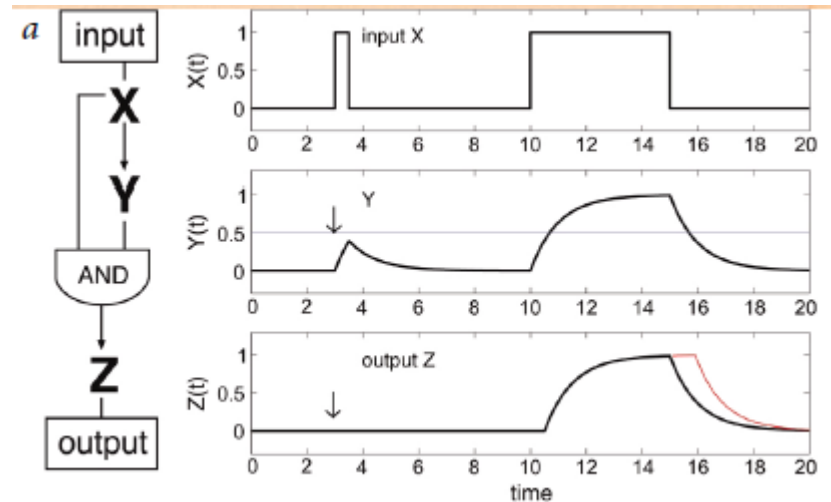
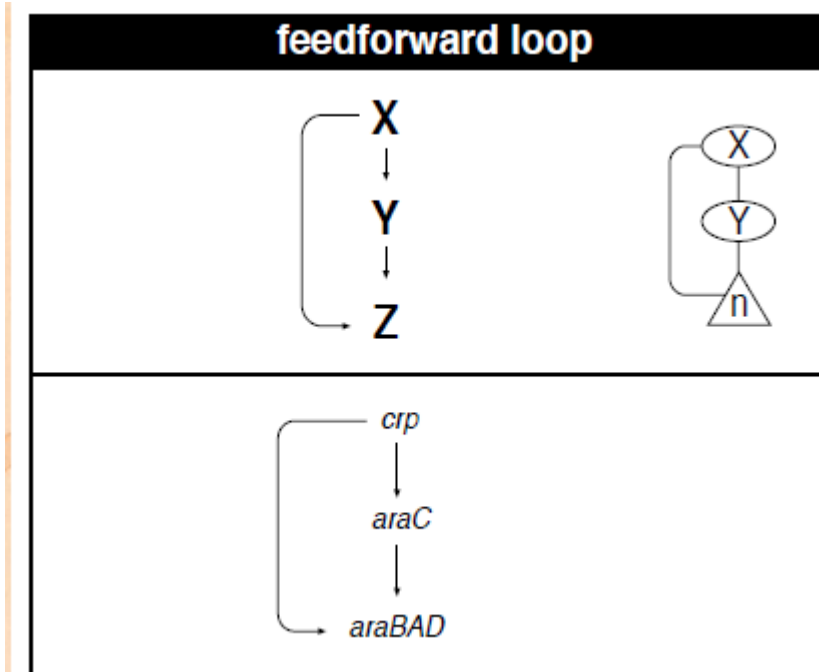
C is the number of appearances of the motif divided by the total number of appearances of all connected three-node subgraphs (Fig.

1B). Subnetworks of size S were generated by choosing a node at random and adding to it nodes connected by an incoming or outgoing edge, until S nodes were obtained, and then including all of the edges between these S nodes present in the full network. Each of the subnetworks was randomized (17, 18) (shown are mean and SD of 400 subnetworks of each size).



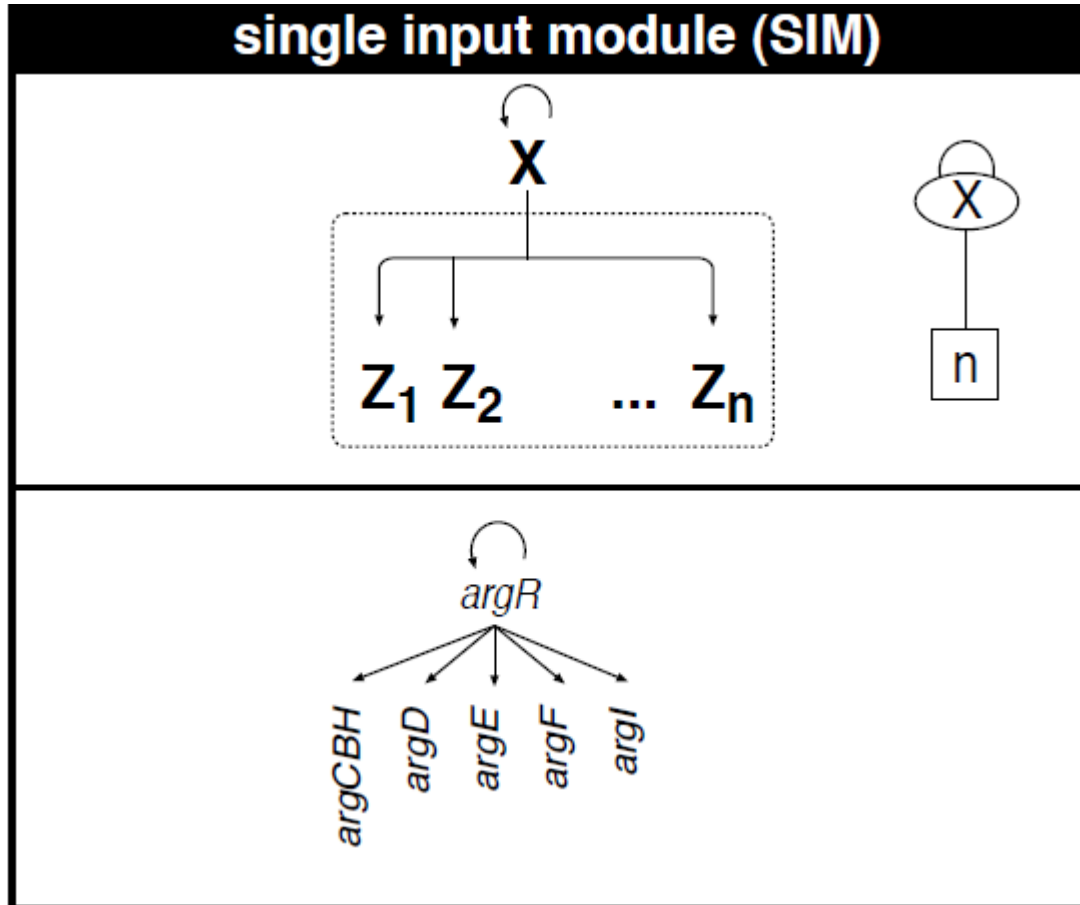
- C_{FFW} se mantiene cte con el tamaño de la red (subredes de la red real)

Motifs: Búsqueda de función desde el patrón de interconexión de pocos nodos

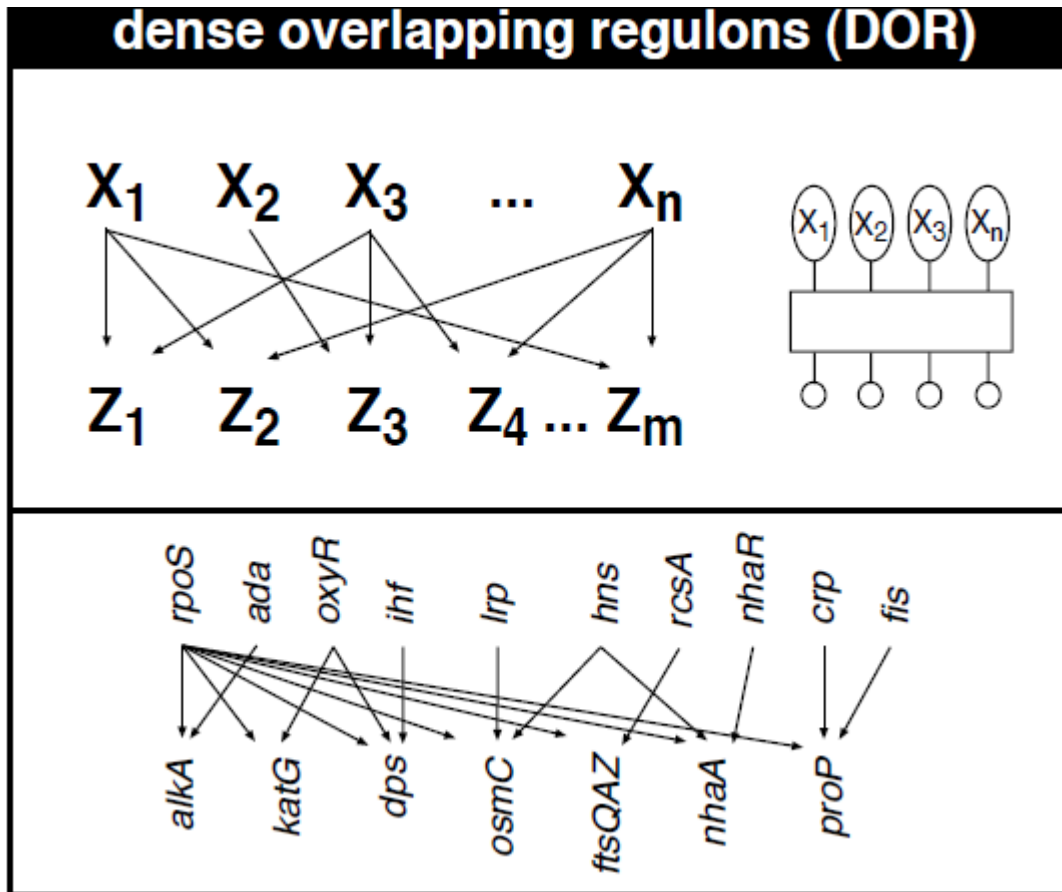




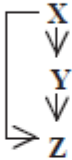

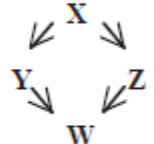

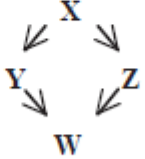
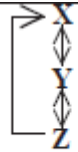
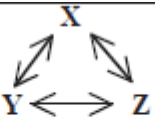
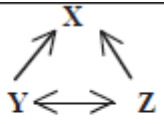
- Z se activa ante una respuesta **sostenida** del estímulo x
- Cuando desaparece x , Z se desactiva rápidamente

Master switch



DOR



Network	Nodes	Edges	N_{real}	$N_{\text{rand}} \pm \text{SD}$	Z score	N_{real}	$N_{\text{rand}} \pm \text{SD}$	Z score	N_{real}	$N_{\text{rand}} \pm \text{SD}$	Z score
Gene regulation (transcription)				Feed-forward loop		Bi-fan					
<i>E. coli</i>	424	519	40	7 ± 3	10	203	47 ± 12	13			
<i>S. cerevisiae</i> *	685	1,052	70	11 ± 4	14	1812	300 ± 40	41			
Neurons				Feed-forward loop		Bi-fan		Bi-parallel			
<i>C. elegans</i> †	252	509	125	90 ± 10	3.7	127	55 ± 13	5.3	227	35 ± 10	20
Food webs				Three chain		Bi-parallel					
Little Rock	92	984	3219	3120 ± 50	2.1	7295	2220 ± 210	25			
Ythan	83	391	1182	1020 ± 20	7.2	1357	230 ± 50	23			
St. Martin	42	205	469	450 ± 10	NS	382	130 ± 20	12			
Chesapeake	31	67	80	82 ± 4	NS	26	5 ± 2	8			
Coachella	29	243	279	235 ± 12	3.6	181	80 ± 20	5			
Skipwith	25	189	184	150 ± 7	5.5	397	80 ± 25	13			
B. Brook	25	104	181	130 ± 7	7.4	267	30 ± 7	32			
World Wide Web				Feedback with two mutual dyads		Fully connected triad		Uplinked mutual dyad			
nd.edu§	325,729	1.46e6	1.1e5	2e3 ± 1e2	800	6.8e6	5e4 ± 4e2	15,000	1.2e6	1e4 ± 2e2	5000

Modelando redes que evolucionan

- Modelos como $G(N,M)$, $G(N,p)$, Modelo Configuracional permiten reportar propiedades de ensambles de redes "completas"
- **Modelos generativos:** modelan los **mecanismos** por los cuales las redes se generan y adquieren determinadas características.

$$G(t) = G(N(t), M(t))$$

- Enlace preferencial (*preferential attachment*)
- Modelo de replicado de vértices

Modelo Barabasi-Albert

Dos ingredientes principales

- **Crecimiento** de la red: La red se modela en continuo crecimiento agregando en cada paso un nuevo nodo de grado m

$$G(t) = G(N = t + m_0, M = m_0 + m * t)$$

- **Enlace preferencial**: El nuevo nodo se enlaza *preferentemente* a nodos de alta conectividad. La probabilidad de enlazarse al nodo- i

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

Plausibilidad:

- Links en la www. Tendemos a conocer (linkear) sitios que conocemos que son, en gran medida típicamente conocidos.
- Redes de citas bibliográficas. Cuanto más citado es un paper, más alta es la posibilidad de que lo conozcamos. Y podemos incrementar aún más sus citas.
- Etc...

Modelo Barabasi-Albert

Emergence of Scaling in Random Networks

Albert-László Barabási* and Réka Albert

Systems as diverse as genetic networks or the World Wide Web are best described as networks with complex topology. A common property of many large networks is that the vertex connectivities follow a scale-free power-law distribution. This feature was found to be a consequence of two generic mechanisms: (i) networks expand continuously by the addition of new vertices, and (ii) new vertices attach preferentially to sites that are already well connected.

A model based on these two ingredients reproduces the observed stationary scale-free distributions, which indicates that the development of large networks is governed by robust self-organizing phenomena that go beyond the particulars of the individual systems.

The inability of contemporary science to describe systems composed of nonidentical elements that have diverse and nonlocal inter-

Department of Physics, University of Notre Dame, Notre Dame, IN 46556, USA.

*To whom correspondence should be addressed. E-mail: alb@nd.edu

actions currently limits advances in many disciplines, ranging from molecular biology to computer science (*1*). The difficulty of describing these systems lies partly in their topology: Many of them form rather complex networks whose vertices are the elements of the system and whose edges represent the interactions between them. For example, liv-

Modelo Barabasi-Albert (MBA)

Se comienza con m_0 nodos, unidos mediante m_0 enlaces de manera que formen una componente conexa

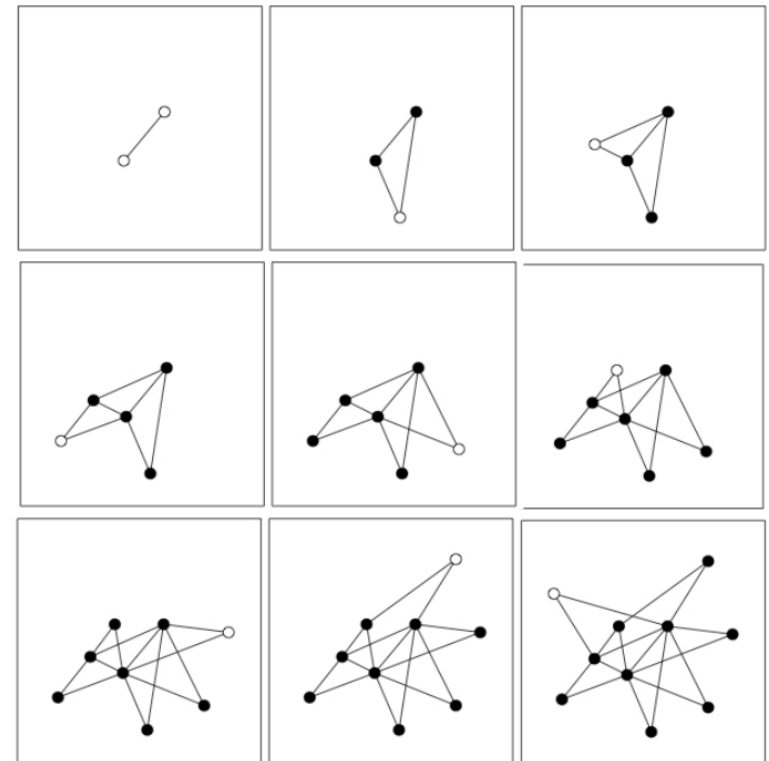
Dos ingredientes principales

- **Crecimiento** de la red: La red se modela en continuo crecimiento agregando en cada paso un nuevo nodo de grado **m**

$$G(t) = G(N = t + m_0, M = m_0 + m * t)$$

- **Enlace preferencial**: El nuevo nodo se enlaza *preferentemente* a nodos de alta conectividad. La probabilidad de enlazarse al nodo- i

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



Distribucion de grado en el MBA

Dos ingredientes principales

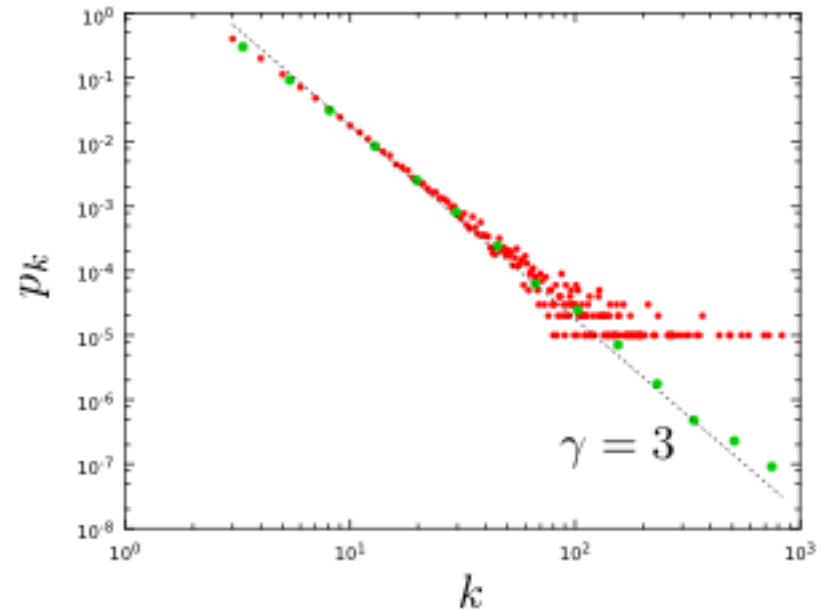
- **Crecimiento** de la red: La red se modela en continuo crecimiento agregando en cada pas un nuevo nodo de grado m

$$G(t) = G(N = t + m_0, M = m_0 + m * t)$$

- **Enlace preferencial**: El nuevo nodo se enlaza *preferentemente* a nodos de alta conectividad. La probabilidad de enlazarse al nodo- i


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

$N=100000, m=3$



Origen de los hubs en el MBA

- Como varía el grado de un nodo con el tiempo?
- Usamos aproximación k continuo (el grado se reparte de 'a diferenciales' entre nodos competidores).

$$\frac{dk_i(t)}{dt} = m \frac{k_i(t)}{\sum_j^{N-1} k_j(t)} = m \frac{k_i(t)}{2mt - m} = \frac{k_i(t)}{2t - 1} \sim \frac{k_i(t)}{2t}$$
$$\sum_j^{N-1} k_j(t) = 2mt - m$$


Origen de los hubs en el MBA

- Como varía el grado de un nodo con el tiempo?
- Usamos aproximación k continuo: el grado se reparte de a diferenciales entre nodos competidores.

$$\frac{dk_i(t)}{dt} = \frac{k_i(t)}{2t}$$

$$\frac{dk_i(t)}{k_i(t)} = \frac{1}{2} \frac{dt}{t}$$

Integro desde
 t_i (tiempo de insercion
en la red del nodo-i)

$$\log \frac{k_i(t)}{k_i(t_i)} = \frac{1}{2} \log \frac{t}{t_i}$$

m

$$k_i(t) = m \left(\frac{t}{t_i} \right)^{1/2}$$

Origen de los hubs en el MBA

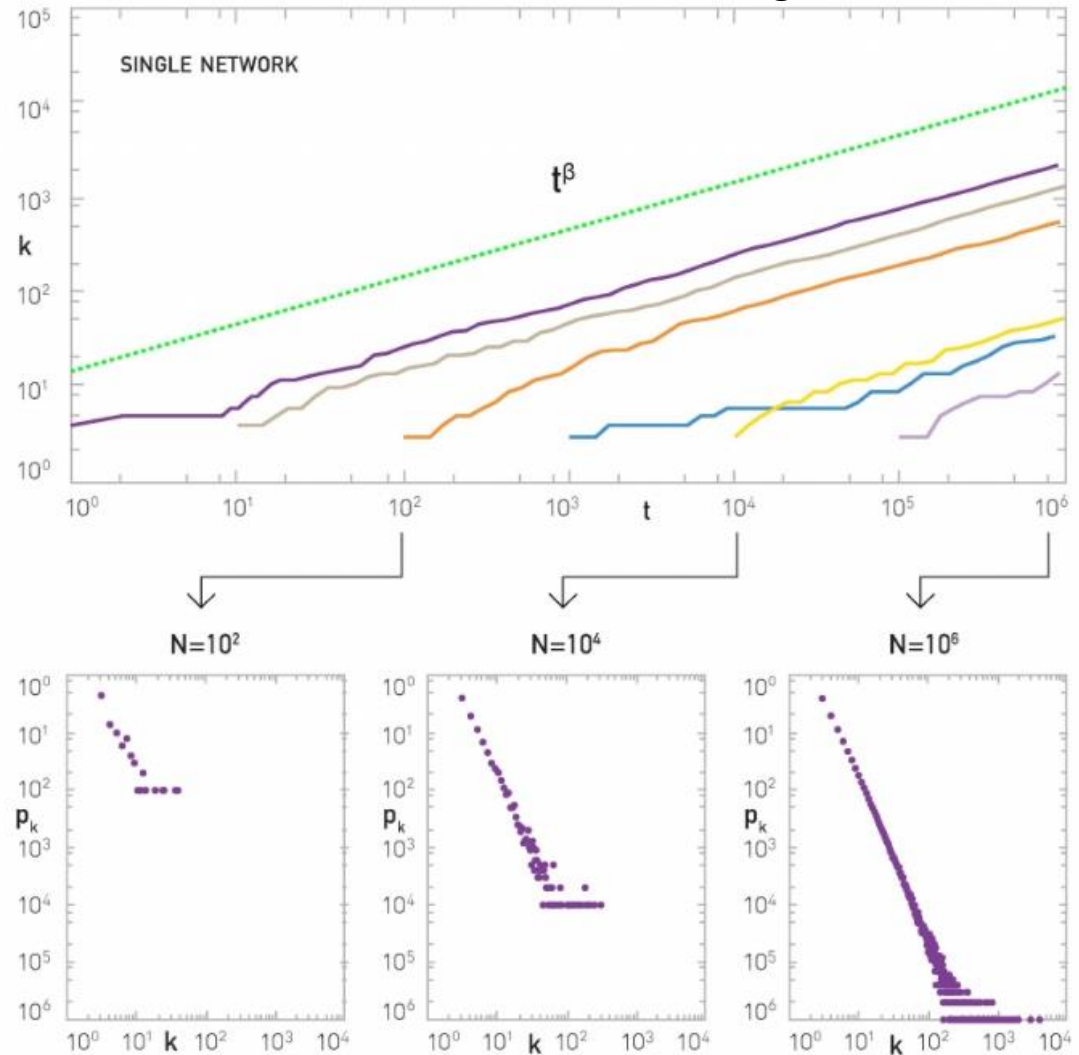
$$k_i(t) = m \left(\frac{t}{t_i} \right)^{1/2}$$

- Todos los nodos siguen la misma ley de crecimiento (!)
- El grado de c/nodo crece como ley de potencias con exponente crítico $\beta = 1/2$ (el crecimiento es sub-lineal, o sea $\beta < 1$)
- La dependencia de la ley de crecimiento con el nodo- i se da a través de t_i . Nodos **antiguos** tienen **mas grado** (solo por haber participado del juego por mas tiempo)

Origen de los hubs en el MBA

Fig 5.6 Barabasi

$$k_i(t) = m \left(\frac{t}{t_i} \right)^{1/2}$$



Notar que la distribución de grado resultante es libre de escala



Distribucion de grado del MBA

Cuantos nodos con $k_i < k$ hay en la red a tiempo t ?

$$k_i(t) = m \left(\frac{t}{t_i} \right)^\beta < k \longrightarrow t_i > t \left(\frac{m}{k} \right)^{1/\beta}$$

Los nodos con $k_i < k$ son los que ingresaron *recientemente*

Como a cada paso temporal adiciono **un** nodo a la red, el nro de nodos que ingresaron a tiempo t_i resulta:

$$t \left(\frac{m}{k} \right)^{1/\beta}$$

Todos ellos tienen más de k vecinos

A tiempo t hay un total de $N(t) = m_0 + t \sim t$ (con $t \gg 1$)

Fraccion de nodos con $k_i > k$ $\frac{t \left(\frac{m}{k} \right)^{1/\beta}}{t} = \left(\frac{m}{k} \right)^{1/\beta}$

Fraccion de nodos con $k_i < k$ $P(k) = 1 - \left(\frac{m}{k} \right)^{1/\beta}$

$$p(k) = \frac{dP(k)}{dk} = \frac{1}{\beta} \frac{m^{1/\beta}}{k^{\frac{1}{\beta}+1}}$$

$$p(k) = 2m^2 k^{-3}$$

Replicación de vértices y modelo de duplicación y divergencia de genes

A model of large-scale proteome evolution

Ricard V. Solé^{1,2,3}, Romualdo Pastor-Satorras¹, Eric D. Smith², and Thomas Kepler²

¹ *Complex Systems Research Group, FEN*

Universitat Politècnica de Catalunya, Campus Nord B4, 08034 Barcelona, Spain

² *Santa Fe Institute, 1399 Hyde Park Road, New Mexico 87501, USA*

³ *NASA-associated Astrobiology Institute, INTA/CSIC, Carretera del Ajalvir Km 4, Madrid*

The next step in the understanding of the genome organization, after the determination of complete sequences, involves proteomics. The proteome includes the whole set of protein-protein interactions, and two recent independent studies have shown that its topology displays a number of surprising features shared by other complex networks, both natural and artificial. In order to understand the origins of this topology and its evolutionary implications, we present a simple model of proteome evolution that is able to reproduce many of the observed statistical regularities reported from the analysis of the yeast proteome. Our results suggest that the observed patterns can be explained by a process of gene duplication and diversification that would evolve proteome networks under a selection pressure, favoring robustness against failure of its individual components.

- Se elige un nodo al azar y se **duplica** (y sus enlaces)
- Con proba δ se eliminan sus enlaces
- Con proba α se recablean

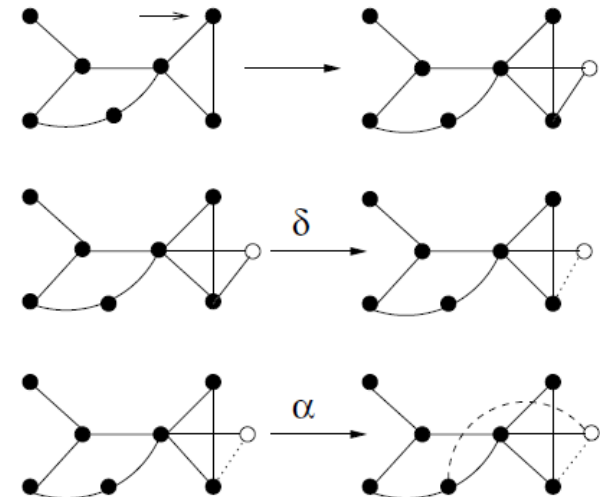
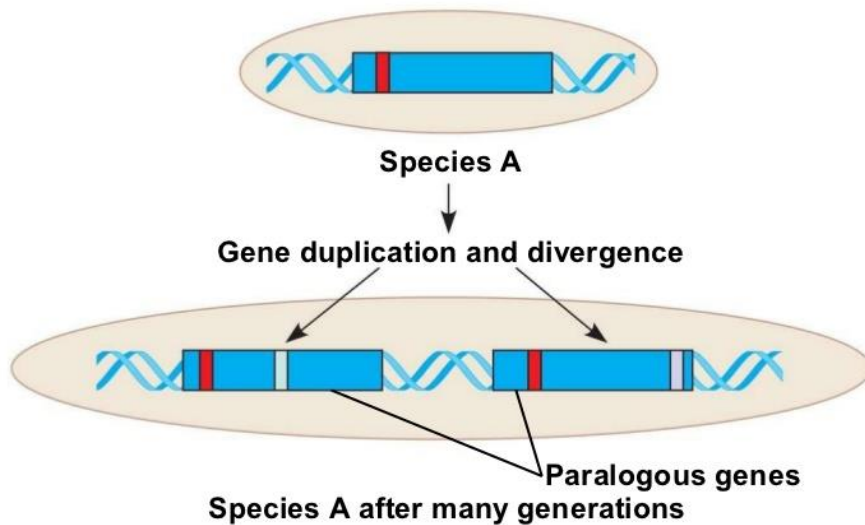


FIG. 2. Growing network by duplication of nodes. First (a) duplication occurs after randomly selecting a node (arrow). The links from the newly created node (white) now can experience deletion (b) and new links can be created (c); these events occur with probabilities δ and α , respectively.

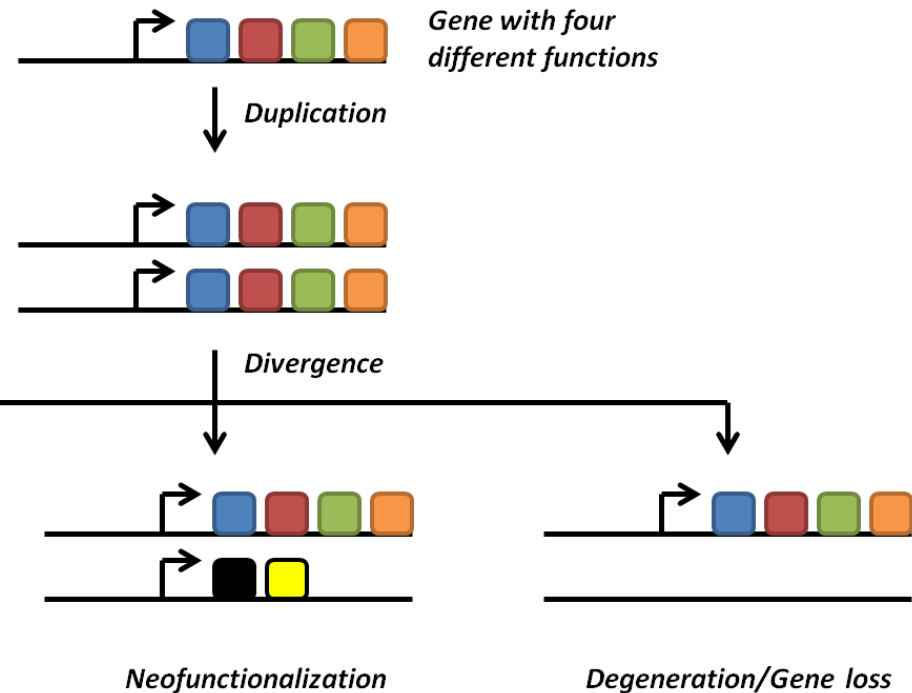
Todo lo que ud quiso saber sobre duplicacion genica y genomica funcional en una remera

Fig. 26-18b

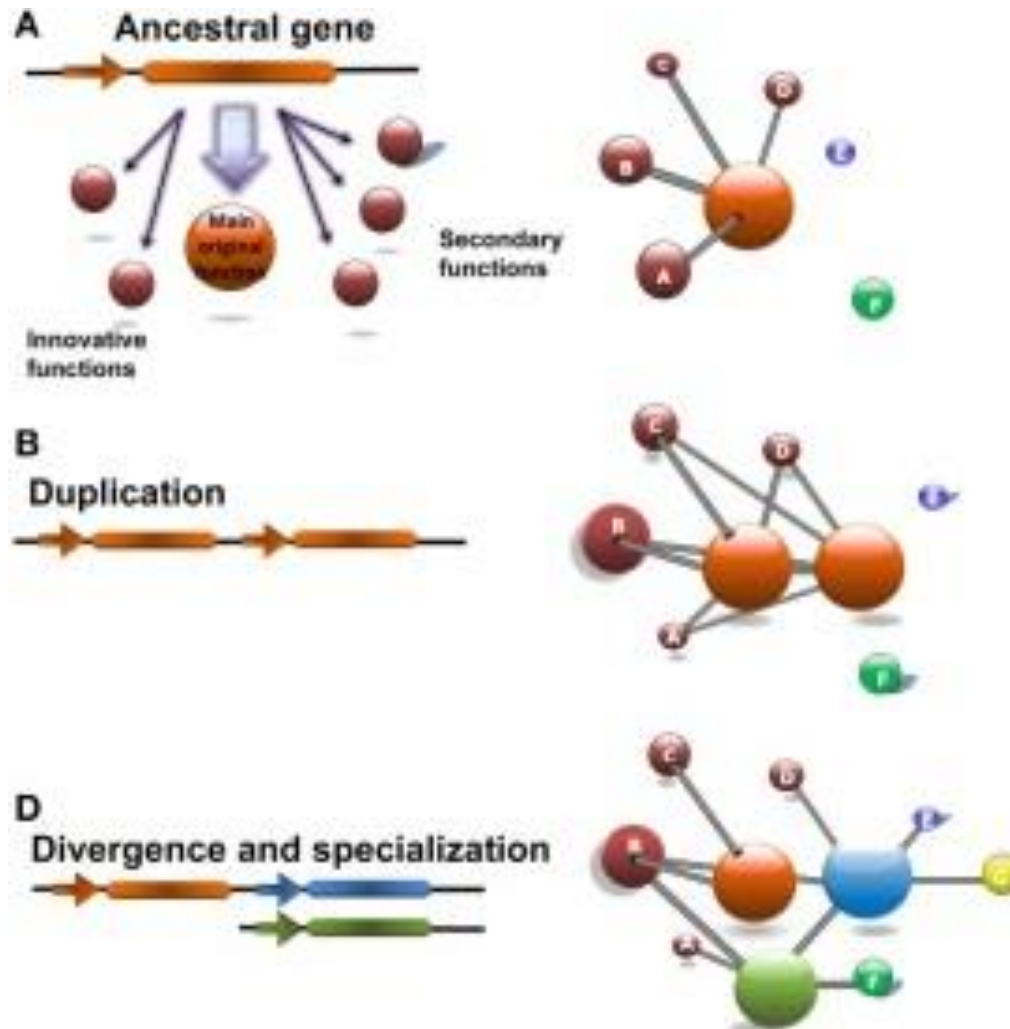


(b) Paralogous genes

Copyright © 2008 Pearson Education, Inc., publishing as Pearson Benjamin Cummings.



Replicación de vértices y modelo de duplicación y divergencia de genes



Modelo de replicado de vértices

- I. Comenzamos con n_0 nodos. Todos de grado c , conectados entre sí de manera aleatoria
- II. A cada paso temporal se agrega un nodo nuevo. Sus c enlaces salientes se fijan
 - a. Copiando (γc) enlaces salientes de un nodo de referencia elegido al azar.
 - b. Asignando el nro restante, $(1-\gamma)c$, de enlaces eligiendo como blancos nodos de la red de manera aleatoria

Este procedimiento conduce a una red que asintóticamente presenta una distribución de grado tipo ley de potencia:

$$P_k \sim k^{-(1+\frac{1}{\gamma})} \left\{ \begin{array}{l} -2: \text{ copia perfecta} \\ -\infty: \text{ copia aleatoria} \end{array} \right.$$

Atenti....A pesar de poder generar la misma P_k que el modelo-BA, este modelo posee una estructura muy diferente: presenta correlaciones entre los conexiones de los vértices.

