

# Retomando....

# Habiamos visto...

De acuerdo a la distribución de grado las redes pueden clasificarse en

## Redes acotadas exponencialmente:

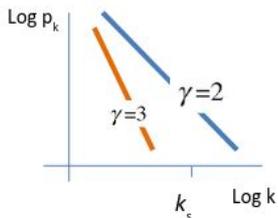
- distribución decae para altos valores de  $k$  exponencialmente o más rapido
- Ausencia de grandes fluctuaciones en el grado [  $\langle k^2 \rangle < \langle k \rangle$  ]

## Redes de cola pesada:

- redes cuya distribución decae como ley de potencia para altos valores de  $k$
- Típicamente presentan outliers (hubs)

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

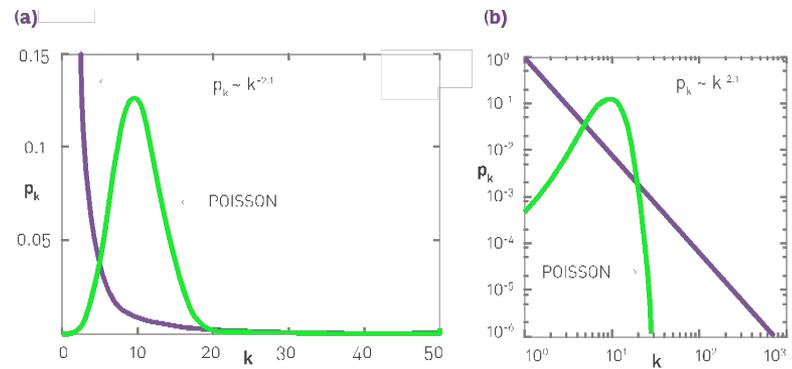
- Pueden presentarse grandes fluctuaciones en el grado [  $\langle k^2 \rangle \gg \langle k \rangle$  ]



$2 < \gamma \leq 3$   
 $\gamma > 3$

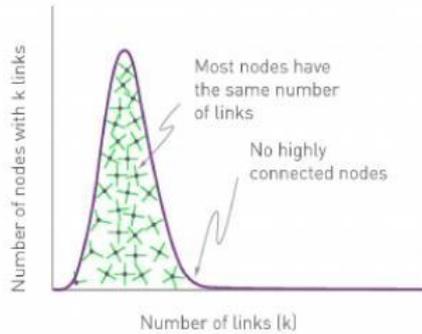
$\langle k \rangle$	$\langle k^2 \rangle$
finito	<b>infinito</b>
finito	finito

Cual es el efecto de tener **hubs** en una red?

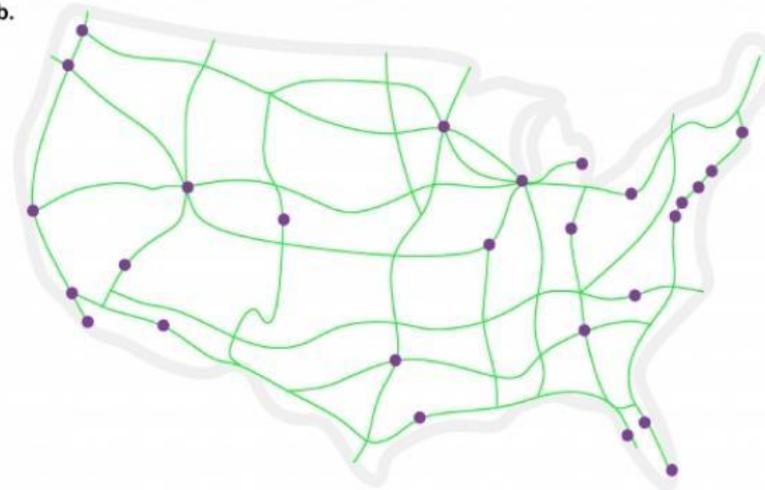


# Efecto de hubs sobre distancias

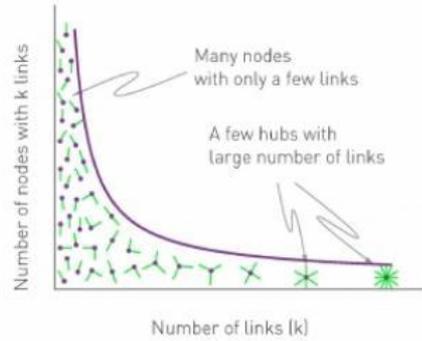
a. POISSON



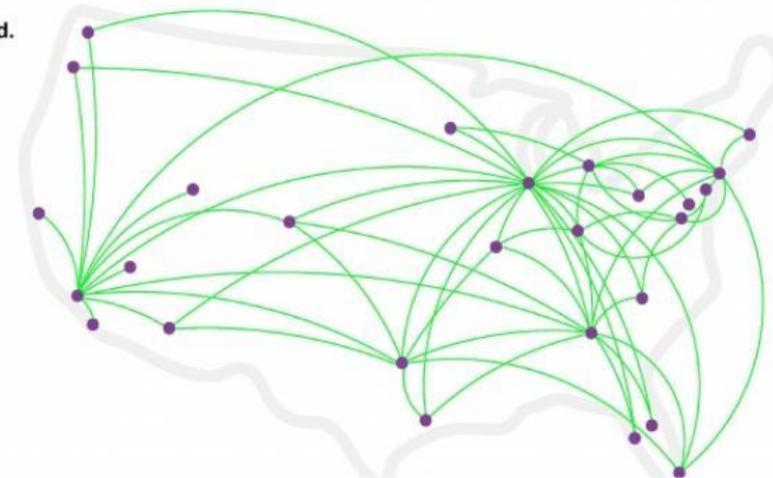
b.



c. POWER LAW



d.

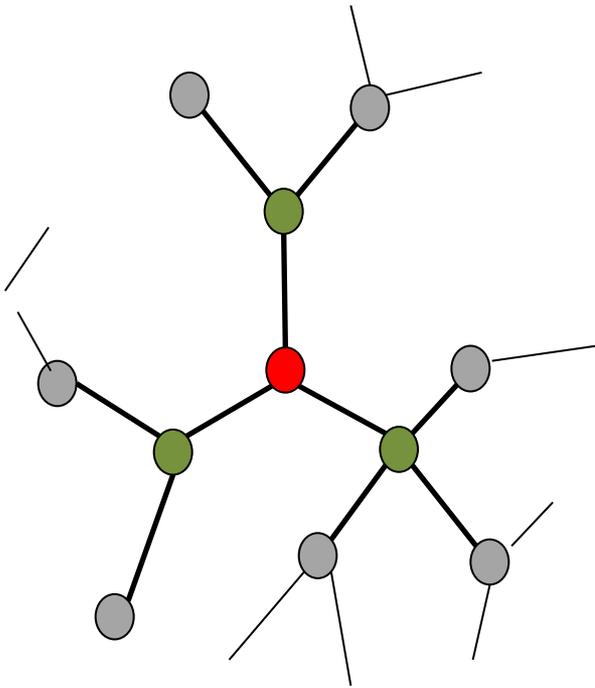


La presencia de hubs crea **atajos**

# Distancias y mundo-pequeño en redes aleatorias (idea)

En redes aleatorias (distribución Poisson) los nodos tienden a tener  $k \sim \langle k \rangle$ .

- Nro primeros vecinos  $\sim \langle k \rangle$
- Nro segundos vecinos  $\sim \langle k \rangle^2$
- Nro de vecinos a distancia  $s \sim \langle k \rangle^s$



El nro de nodos hasta una distancia  $s$  resulta

$$N(l) = 1 + \langle k \rangle + \langle k \rangle^2 + \dots + \langle k \rangle^l = \frac{\langle k \rangle^{l+1} - 1}{\langle k \rangle - 1}$$

Como máximo  $N(l_{\max}) = N$

$$N \sim \langle k \rangle^{l_{\max}}$$

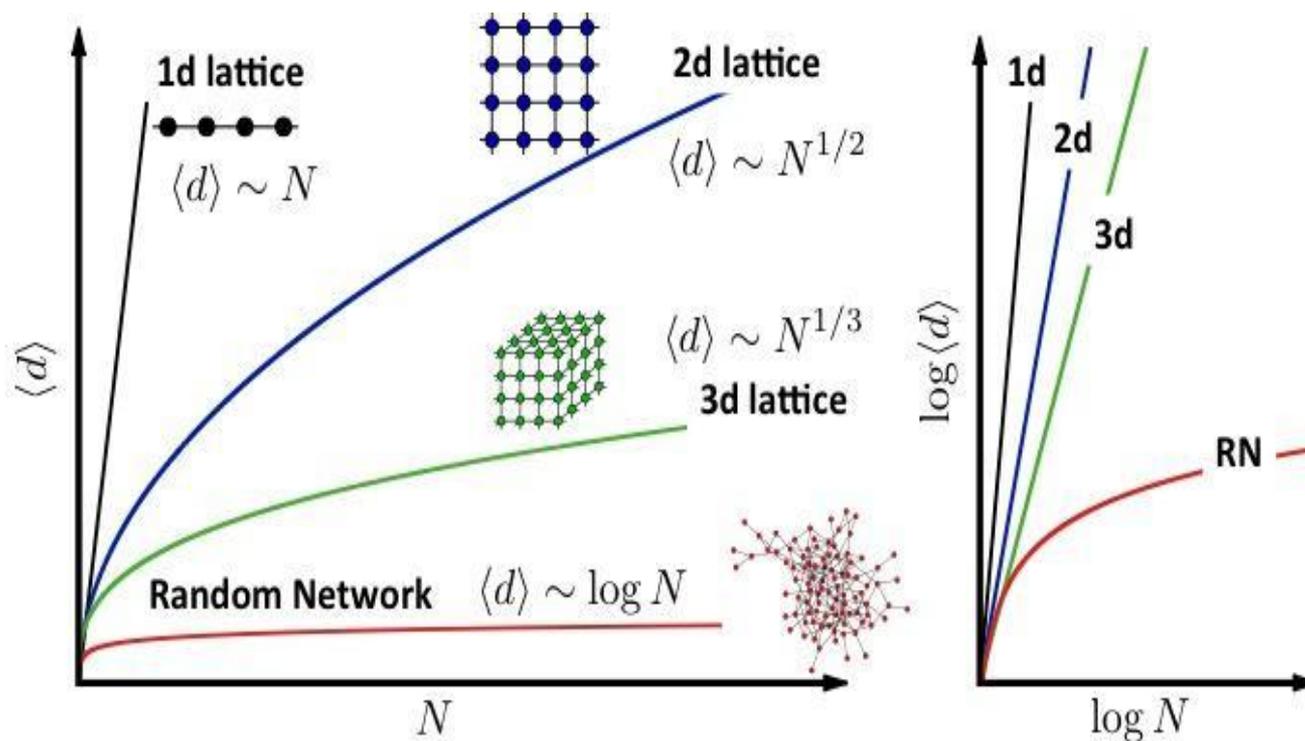
$$l_{\max} \sim \frac{\log N}{\log \langle k \rangle} \quad \leftarrow \text{Efecto mundo pequeño}$$

# Vivir en mundos-pequeños

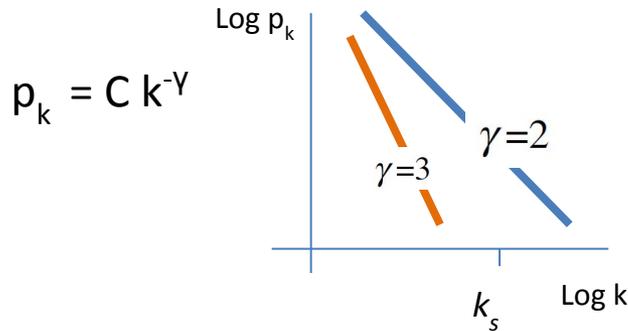
Distancias 'mundos-pequeños'  $\leftrightarrow$  Distancias 'mundo físico'

$$l_{max} \sim \log N$$

$$l \sim N^{1/d}$$



# Distancias en redes libres de escala

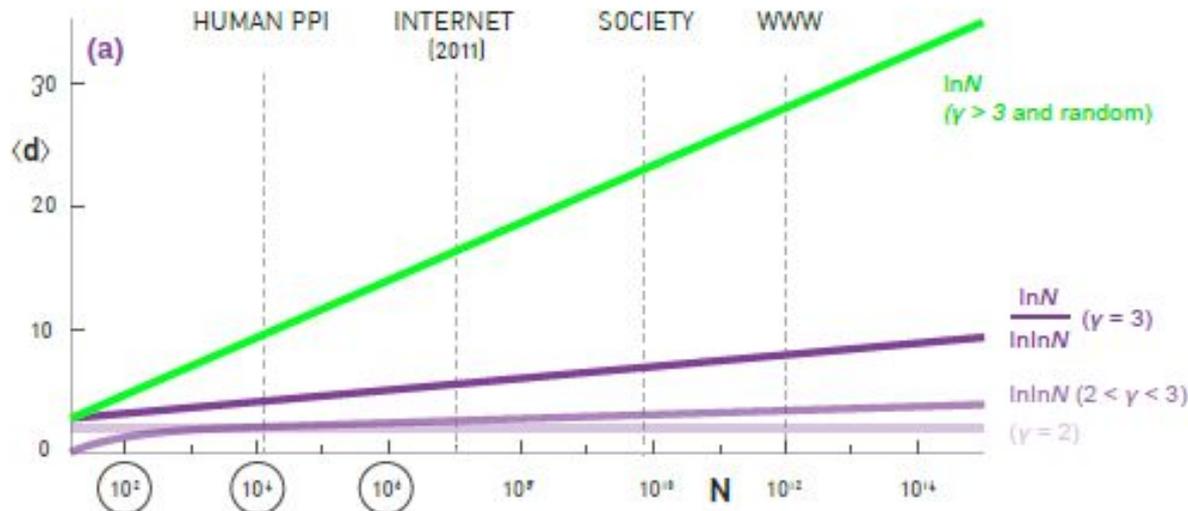


$$k_{max} = k_{min} N^{\frac{1}{\gamma-1}}$$

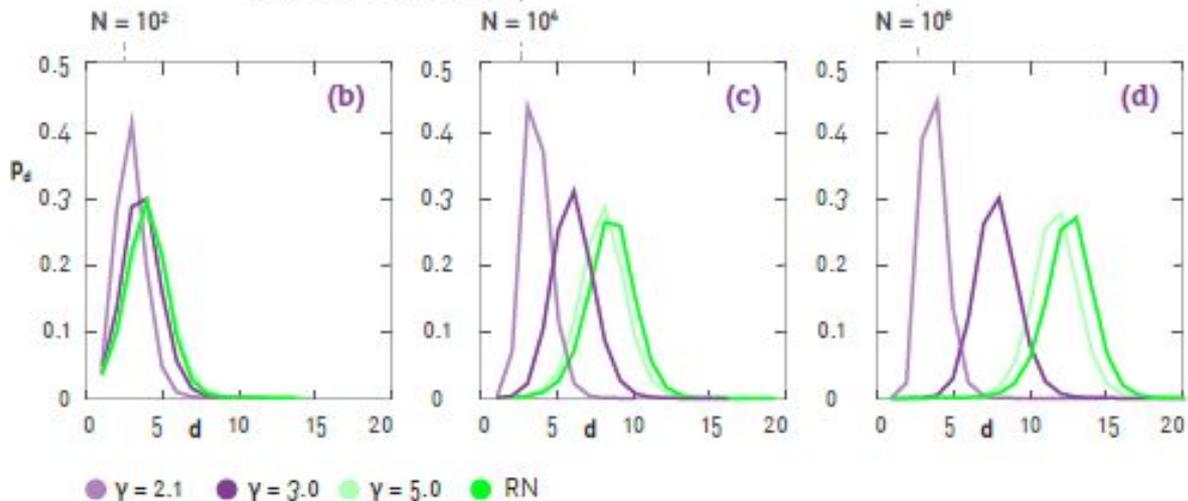
$$\langle k^m \rangle = C \frac{k_{max}^{m-\gamma+1} - k_{min}^{m-\gamma+1}}{m - \gamma + 1}$$

{	$\langle d \rangle \sim$	const.	$\gamma=2$	Tamaño hub $o(N)$ . Estructura tipo <i>spoke</i> . Camino medio independiente del tamaño del sistema
	$\langle d \rangle \sim$	$\ln \ln N$	$2 < \gamma < 3$	Longitud media aumenta mas lento que log (!) <b>Ultra-small-world</b> debido a hubs.
	$\langle d \rangle \sim$	$\frac{\ln N}{\ln \ln N}$	$\gamma=3$	Valor crítico de $\gamma$ para el cual $\langle k^2 \rangle$ deja de diverger.
	$\langle d \rangle \sim$	$\ln N$	$\gamma > 3$	$\langle k^2 \rangle$ es finito. No hay suficientes hubs, ni son tan gdes. Se comporta como <b>small-world</b> random network.

# Distancias en redes libres de escala

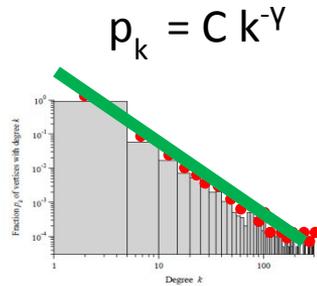


$$\langle d \rangle \sim \begin{cases} \text{const.} & \gamma = 2, \\ \frac{\ln \ln N}{\ln(\gamma - 1)} & 2 < \gamma < 3, \\ \frac{\ln N}{\ln \ln N} & \gamma = 3, \\ \ln N & \gamma > 3. \end{cases}$$



Las diferencias entre diferentes tipos de *escaleo* para las distancias se ponen de manifiesto en redes grandes.

# Leyes de potencia en régimen logN



Es difícil encontrar redes reales con distribución de grado ley de potencia con  $\gamma > 3$

- Se necesita observar nodos con grados que varíen en al menos 2 (o mejor 3) ordenes de magnitud:  $k_{min} \sim 1$ ,  $k_{max} \sim 10^2$

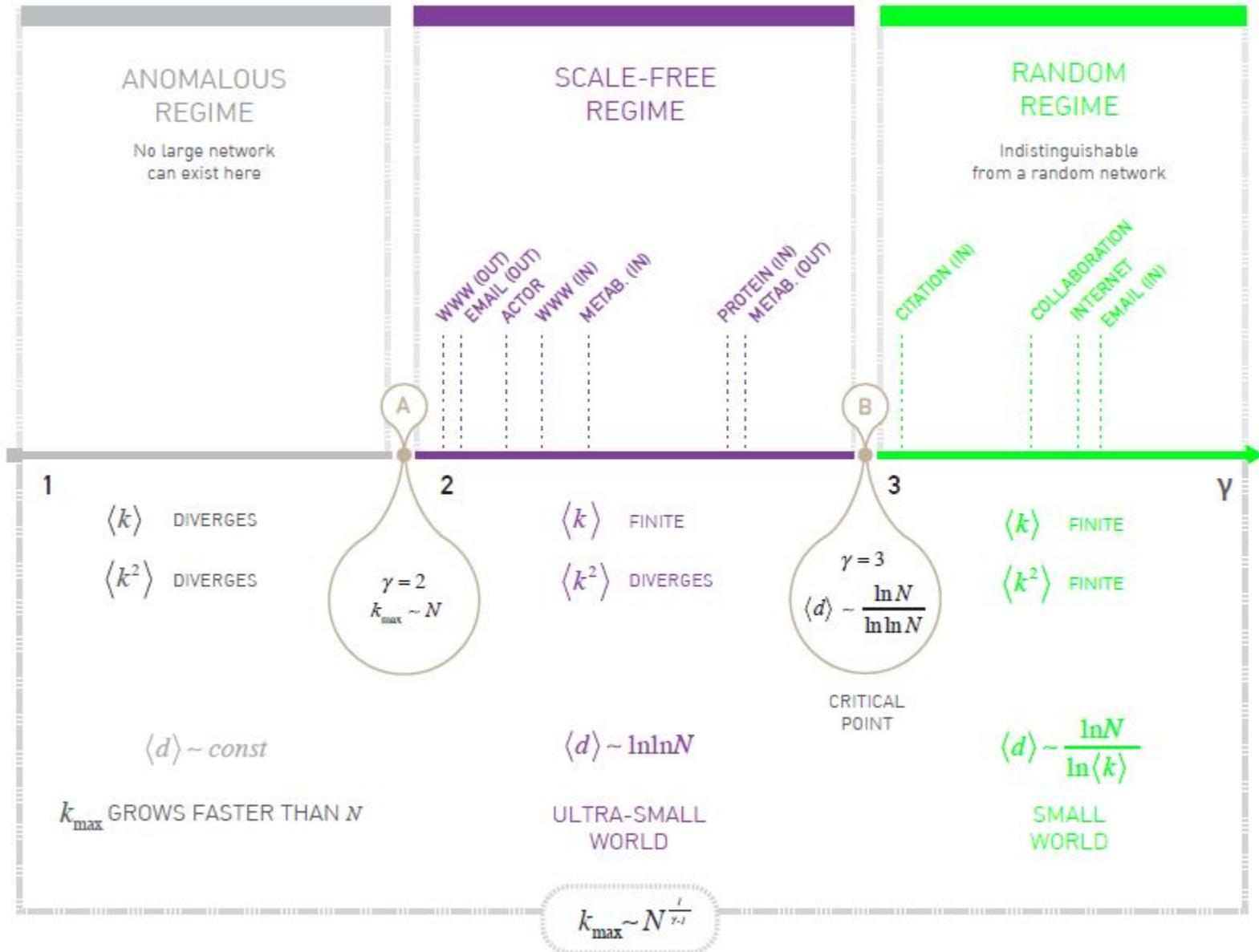
- Pero entonces

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

$$N = \left( \frac{k_{max}}{k_{min}} \right)^{\gamma-1} = 10^{2(\gamma-1)}$$

Para reconocer una red con distrib. de grado de exponente  $\gamma=5$  necesitaría que la misma tuviera al menos  $N \sim 10^8$  nodos (!)

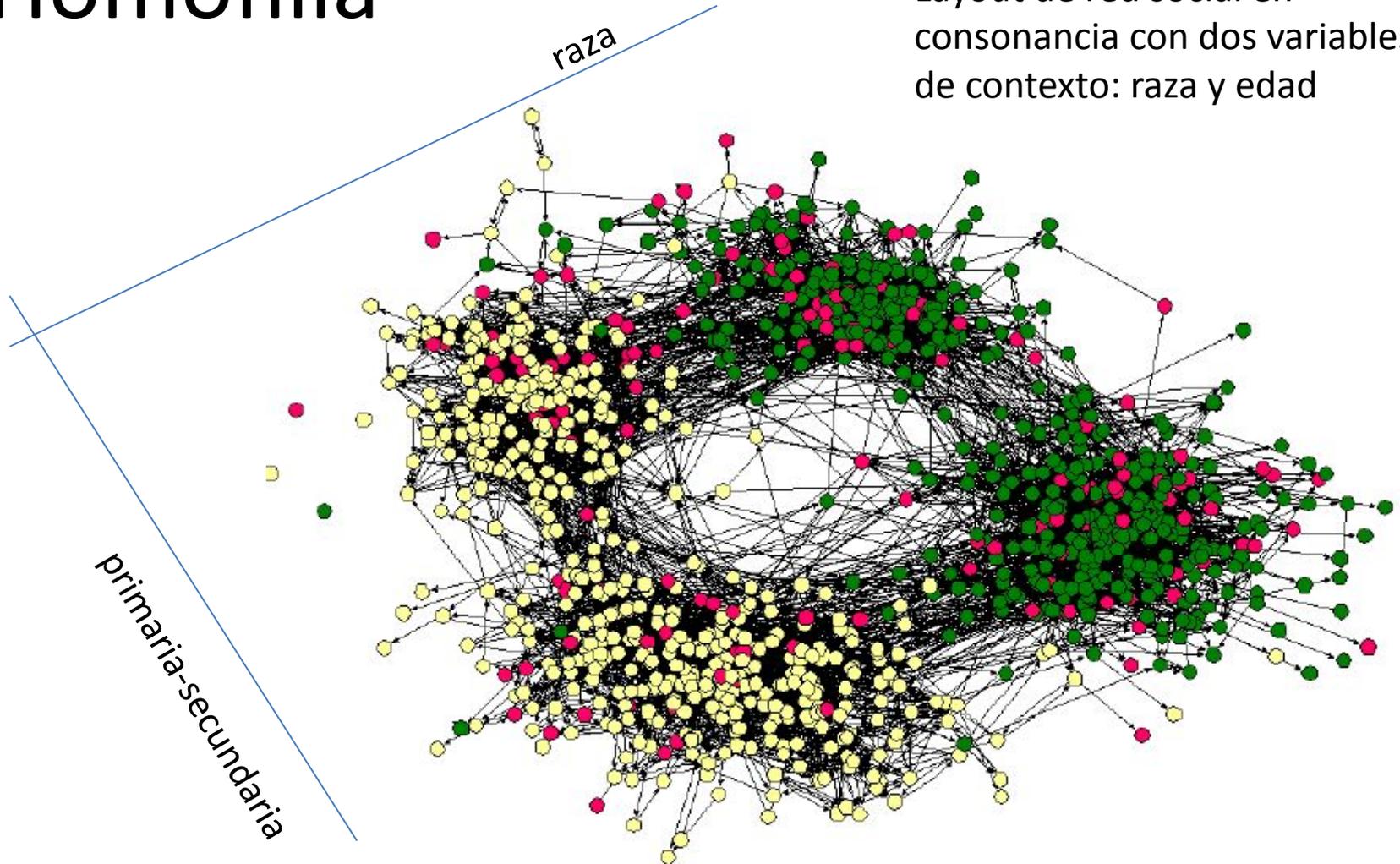
# Exponente crítico



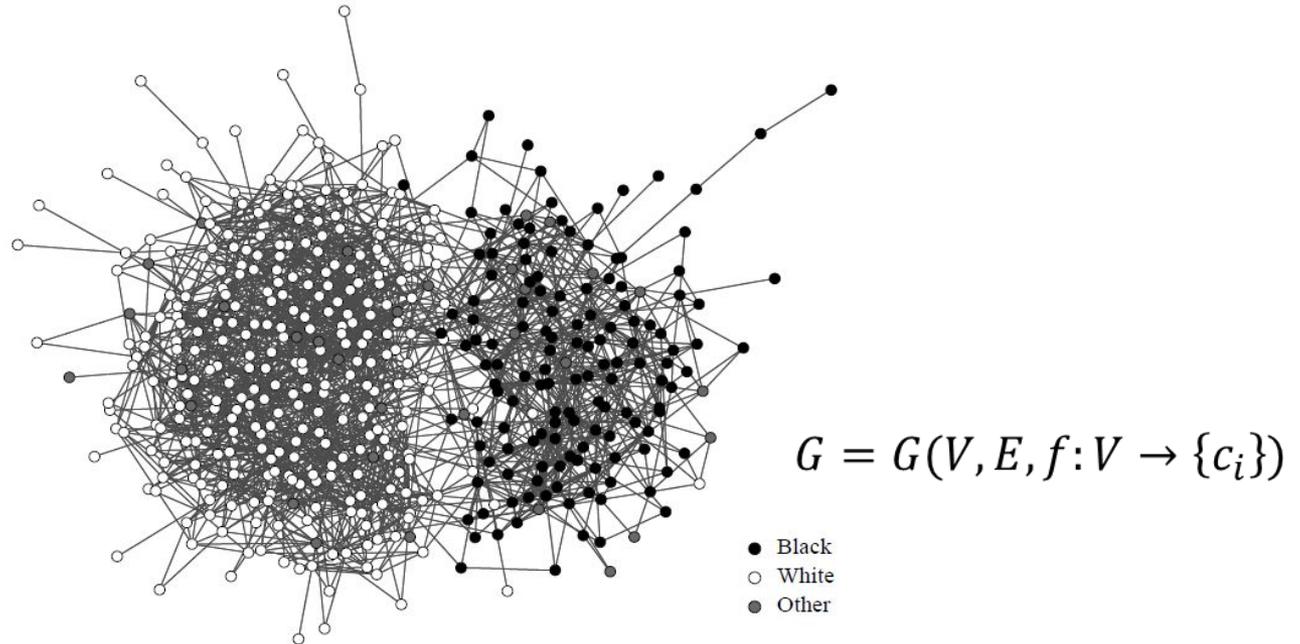
# Estructura y correlaciones en la red

# Homofilia

Layout de red social en consonancia con dos variables de contexto: raza y edad



# Assortative mixing para variables categóricas

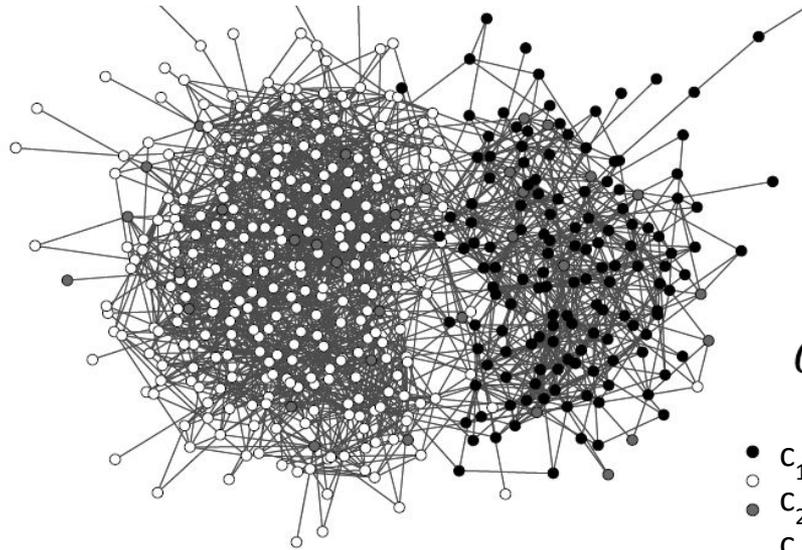


Supongamos que nodos de una red puedan clasificarse de acuerdo a cierta característica en un conjunto discreto y finito de valores (e.g. género, raza, etc)

La red es **asortativa (disortativa)** si una fracción significativa de enlaces se establece entre nodos del **mismo (diferente)** tipo.

# Assortative mixing: características categóricas v.1

Supongamos que existen  $n_c$  clases diferentes para los  $n$  nodos de una red de  $m$  enlaces.  
Sea  $c_i$  la clase del nodo- $i$ . El número de enlaces entre mismo tipo de nodos resulta:



$$G = G(V, E, f: V \rightarrow \{c_i\})$$

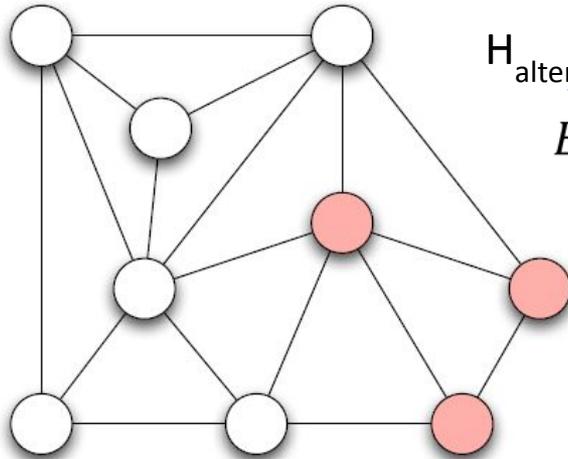
- $c_1$ : blanco
- $c_2$ : negro
- $c_3$ : otro

Red real

$$\sum_{\text{edges } (i,j)} \delta(c_i, c_j) = \frac{1}{2} \sum_{ij} A_{ij} \delta(c_i, c_j) \quad \delta(c_i, c_j) = 1 \text{ si } c_i = c_j$$

Contra que lo comparo??

# Recordemos...Cuantificando homofilia



$H_{\text{alternativa-}}$

$$G = G(V, E, f_v: V \rightarrow \{\text{color1}, \dots, \text{colorN}\})$$

$E \leftrightarrow f_v$

- Existe una tendencia a establecer enlaces entre nodos de igual color?
  - hay *muchos* enlaces entre nodos de igual color?
  - hay *pocos* enlaces entre nodos de color diferente?

Contra qué cosa comparo para saber si hay **pocos** enlaces entre nodos de diferente color?

$H_{\text{null-1}}$ : modelo de *recolorado* Sobre la red original, coloreo **al azar** 6 blancas y 3 salmon ( $f_v$ )

$H_{\text{null-2}}$ : modelo de *la bolsa* tengo 6 bolas blancas y 3 salmon en una bolsa. Saco 18 pares **al azar**

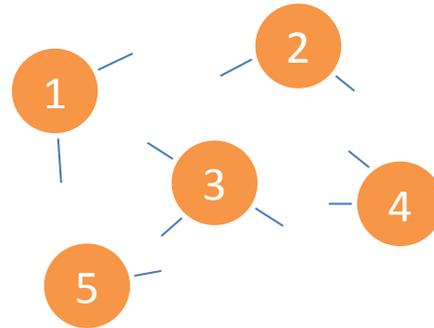
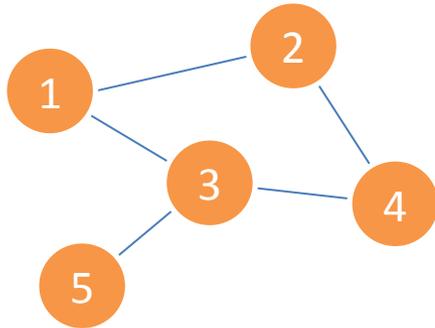
modelo de *recableado* Recableo la red cambiando **al azar** la estructura del grafo ( $E$ )

$H_{\text{null-3}}$ : modelo de *recableado configuracional* Recableo la red cambiando **al azar** la estructura del grafo **respetando el grado** de cada nodo ( $E$ )

# Modelo configuracional

Recableado de la red, **preservando la distribución de grado** original

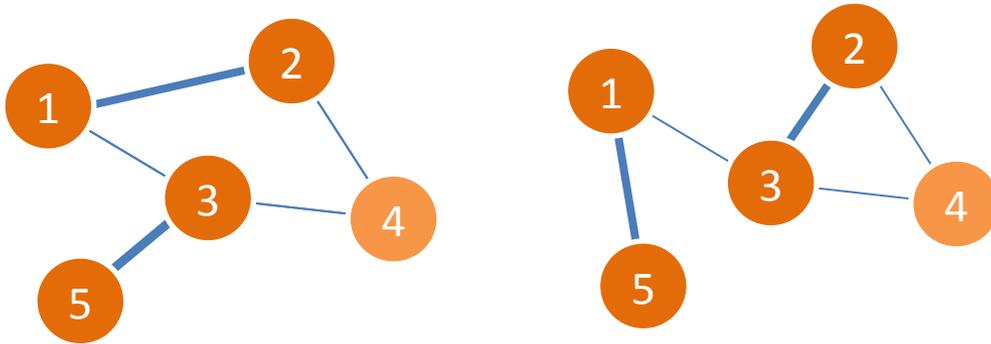
Para cada rayito elijo otro al azar



ojo: Implementado *naivamente* se pueden generar loops o multienlaces

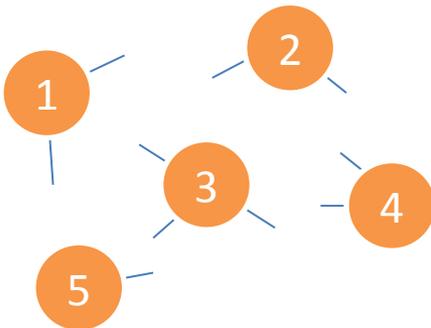
# Modelo configuracional

Recableado de la red, **preservando la distribución de grado original**



- a. elijo 2 pares de nodos enlazados
  - b. intercambio parejas
- Repito a y b muchísimas veces

Este recableado preserva el grado de cada nodo



En la red recableada, la probabilidad de que exista un enlace entre el nodo  $i$  y el  $j$

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

nro total de *rayos*

fracción de enlaces que *conducen* a  $j$

# Assortative mixing: características categóricas

Supongamos que existen  $n_c$  clases diferentes para los  $n$  nodos de una red de  $m$  enlaces. Sea  $c_i$  la clase del nodo- $i$ . El número de enlaces entre mismo tipo de nodos resulta:

Red real

$$\sum_{\text{edges } (i,j)} \delta(c_i, c_j) = \frac{1}{2} \sum_{ij} A_{ij} \delta(c_i, c_j)$$

$$\delta(c_i, c_j) = 1 \text{ si } c_i = c_j$$

Red aleatoria (recableado)

$$\frac{1}{2} \sum_{ij} \frac{k_i k_j}{2m} \delta(c_i, c_j).$$

Consideramos la diferencia:  $\frac{1}{2} \sum_{ij} A_{ij} \delta(c_i, c_j) - \frac{1}{2} \sum_{ij} \frac{k_i k_j}{2m} \delta(c_i, c_j)$

modularidad

$$Q = \frac{1}{2m} \sum_{ij} \left( A_{ij} - \frac{k_i k_j}{2m} \right) \delta(c_i, c_j)$$

- $Q > 0$  si hay más enlaces entre vertices del mismo tipo que los esperados por azar
- $Q < 0$  si hay menos enlaces entre vertices del mismo tipo que los esperados por azar

## Assortative mixing: **características categóricas**

$$Q = \frac{1}{2m} \sum_{ij} \left( A_{ij} - \frac{k_i k_j}{2m} \right) \delta(c_i, c_j)$$

En general  $Q < 1$  aun para casos perfectamente asortativos:  $A_{ij} = \delta(c_i, c_j)$

Para este caso resulta  $\sum_{ij} A_{ij} \delta(c_i, c_j) = \sum_{ij} A_{ij} = 2m$

$$Q_{\max} = \frac{1}{2m} \left( 2m - \sum_{ij} \frac{k_i k_j}{2m} \delta(c_i, c_j) \right)$$

Y podemos definir una modularidad normalizada

$$\frac{Q}{Q_{\max}} = \frac{\sum_{ij} (A_{ij} - k_i k_j / 2m) \delta(c_i, c_j)}{2m - \sum_{ij} (k_i k_j / 2m) \delta(c_i, c_j)}$$

# Assortative mixing: características categóricas

Otra manera de computar asortatividad/modularidad:

$$Q = \frac{1}{2m} \sum_{ij} \left( A_{ij} - \frac{k_i k_j}{2m} \right) \overbrace{\sum_r \delta(c_i, r) \delta(c_j, r)}^{\delta(c_i, c_j)}$$

$$= \sum_r \left[ \frac{1}{2m} \sum_{ij} A_{ij} \delta(c_i, r) \delta(c_j, r) - \frac{1}{2m} \sum_i k_i \delta(c_i, r) \frac{1}{2m} \sum_j k_j \delta(c_j, r) \right]$$

$$e_{rs} = \frac{1}{2m} \sum_{ij} A_{ij} \delta(c_i, r) \delta(c_j, s)$$

fracción de enlaces entre nodos de la clases  $r$  y  $s$  (o sea, probabilidad de encontrar un nodo de clase  $r$  y otro de clase  $s$  en los extremos de un enlace).

$$a_r = \frac{1}{2m} \sum_{ij} A_{ij} \delta(c_i, r)$$

fracción de enlaces adyacente a nodos del tipo  $r$  (o sea, probabilidad de encontrar un nodo de tipo  $r$  en un extremo de un enlace)

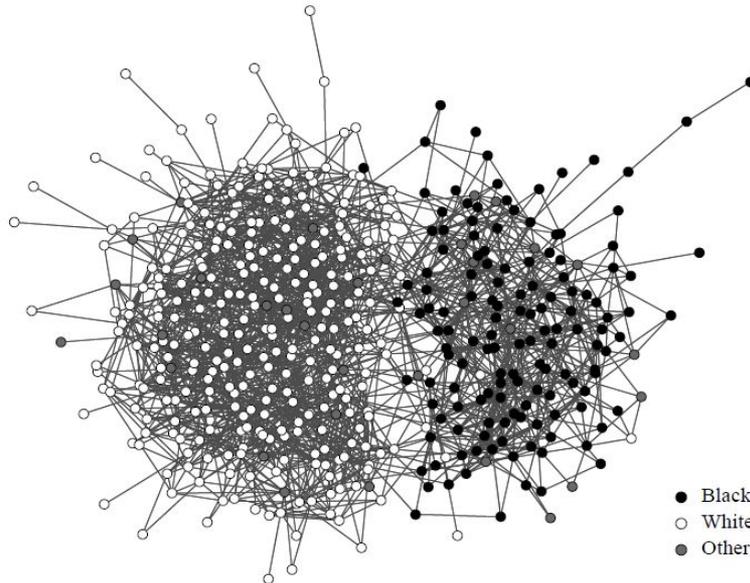
Entonces...

$$Q = \sum_r (e_{rr} - a_r^2)$$

Notar que en una red sin correlaciones

$$e_{rs} = a_r a_s \rightarrow Q = 0$$

# Assortative mixing: características categóricas



Recordar:

$$G = G(V, E, f: V \rightarrow \{c_i\})$$

En cualquiera de los dos casos estoy computando si veo mas enlaces entre mismo tipo de nodos que en el modelo configuracional.

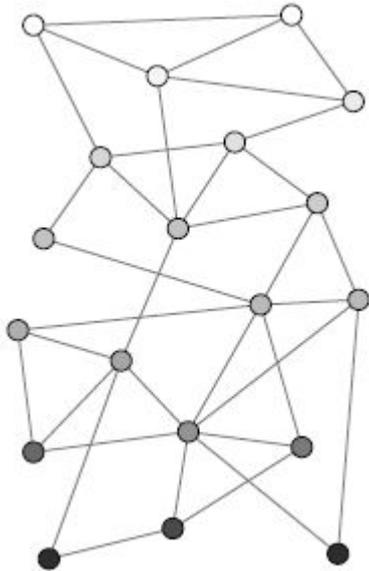
$$Q = \frac{1}{2m} \sum_{ij} \left( A_{ij} - \frac{k_i k_j}{2m} \right) \delta(c_i, c_j)$$

fracción enlaces adyacentes a  
nodos de categoría  $r$

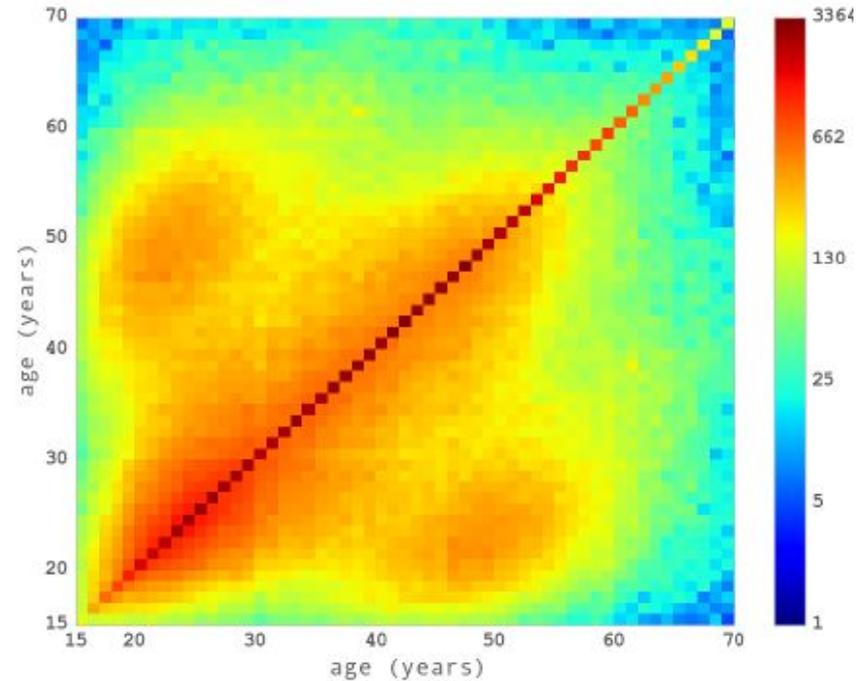
$$Q = \sum_r (e_{rr} - a_r^2)$$

fracción enlaces entre  
nodos de categoría  $r$

# Assortative mixing: características escalares



$$x(v): V \rightarrow \mathbb{R}$$

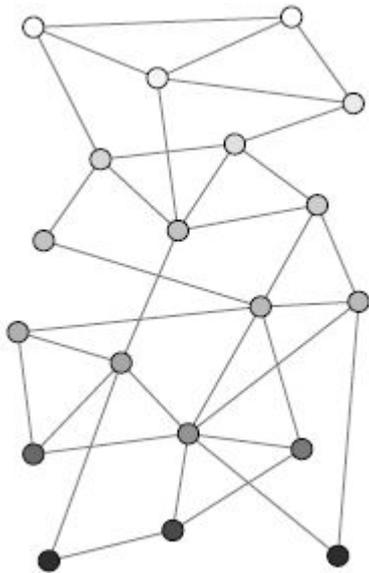


- Red de comunicaciones (llamadas tecnología móvil y SMS)
- Información geo-temporal de llamadas (3 meses, 70e6 usuarios – México)
- Información de la edad de 500000 usuarios
- Interés: Usando la **red y la información parcial** de la edad de algunos vértices, **inferir** la edad de usuarios de la red.
- Hipótesis de trabajo: principio de **homofilia** en la variable **edad**

# Assortative mixing: características escalares

Definimos la media de  $X$  en el extremo de un enlace.

Recorro enlaces (ambos sentidos) asignando a  $c/u$  el valor de sus vertices :



un vértice de valor  $x_i$   
aparece en  $k_i$  enlaces

$$\mu = \frac{\sum_{ij} A_{ij} x_i}{\sum_{ij} A_{ij}} = \frac{\sum_i k_i x_i}{\sum_i k_i} = \frac{1}{2m} \sum_i k_i x_i.$$

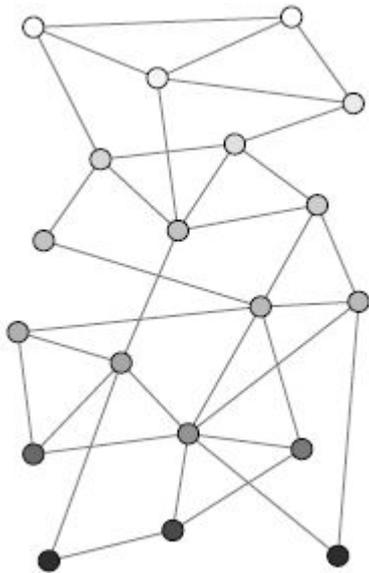
$$\overline{cov(x_{end_1}, x_{end_2})} = \frac{\sum_{ij} A_{ij} (x_i - \mu)(x_j - \mu)}{\sum_{ij} A_{ij}}$$

$$x(v): V \rightarrow \mathbb{R}$$

# Assortative mixing: características escalares

Definimos la media de  $X$  en el extremo de un enlace.

Recorro enlaces (ambos sentidos) asignando a  $c/u$  el valor de sus vertices :



un vértice de valor  $x_i$   
aparece en  $k_i$  enlaces

$$\mu = \frac{\sum_{ij} A_{ij} x_i}{\sum_{ij} A_{ij}} = \frac{\sum_i k_i x_i}{\sum_i k_i} = \frac{1}{2m} \sum_i k_i x_i.$$

$$\overline{cov(x_{end_1}, x_{end_2})} = \frac{\sum_{ij} A_{ij} (x_i - \mu)(x_j - \mu)}{\sum_{ij} A_{ij}}$$

$$= \frac{1}{2m} \sum_{ij} A_{ij} (x_i x_j - \mu x_i - \mu x_j + \mu^2)$$

$$= \frac{1}{2m} \sum_{ij} A_{ij} x_i x_j - \mu^2$$

$$= \frac{1}{2m} \sum_{ij} A_{ij} x_i x_j - \frac{1}{(2m)^2} \sum_{ij} k_i k_j x_i x_j$$

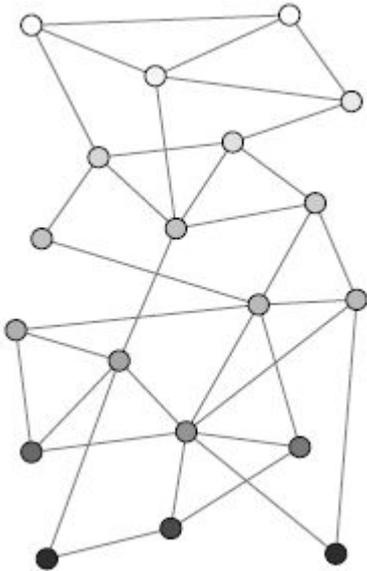
$$= \frac{1}{2m} \sum_{ij} \left( A_{ij} - \frac{k_i k_j}{2m} \right) x_i x_j,$$

$$x(v): V \rightarrow \mathbb{R}$$

Notar que:

$$Q = \frac{1}{2m} \sum_{ij} \left( A_{ij} - \frac{k_i k_j}{2m} \right) \delta(c_i, c_j)$$

# Assortative mixing: características escalares



$$\text{cov}(x_{\text{end}_1}, x_{\text{end}_2}) = \frac{1}{2m} \sum_{ij} \left( A_{ij} - \frac{k_i k_j}{2m} \right) x_i x_j,$$

El valor máximo de  $\text{cov}(x_i, x_j)$  es cuando todo enlace conecta nodos de igual valor

$$\sum_{ij} A_{ij} x_i x_j = \sum_{ij} A_{ij} x_i^2 = \sum_i k_i x_i^2 = \sum_i k_i x_i x_j \delta_{ij}$$

Y finalmente, la versión normalizada resulta

$$r = \frac{\sum_{ij} (A_{ij} - k_i k_j / 2m) x_i x_j}{\sum_{ij} (k_i \delta_{ij} - k_i k_j / 2m) x_i x_j}$$

$$x(v): V \rightarrow \mathbb{R}$$

Es posible extender este razonamiento a campos **vectoriales** definidos sobre la red:

$$\mathbf{x}(v): V \rightarrow \mathbb{R}^n$$



# Assortative mixing por grado

## Correlación de grados

### Caso especial:

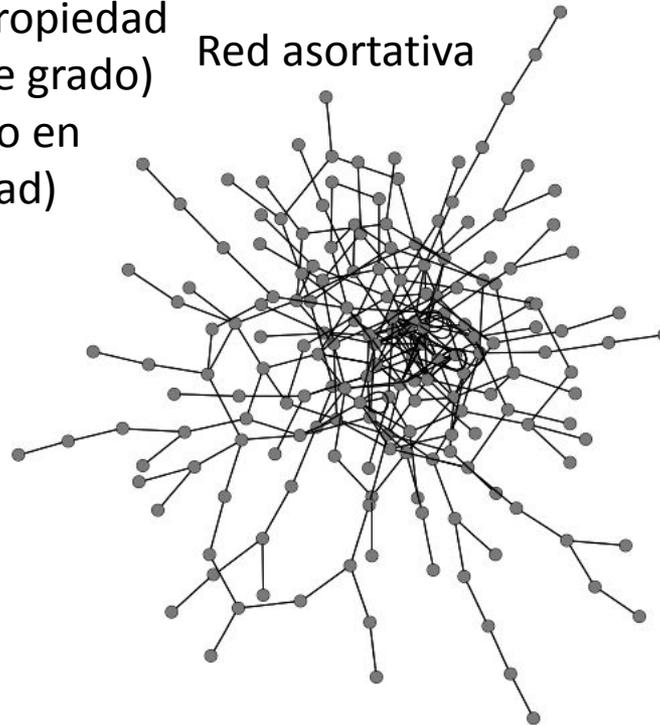
Asortatividad de una cantidad escalar:  $x(v) = k(v)$   
el **campo de conectividad** sobre la red  $x(v): V \rightarrow \mathbb{N}$

$$r = \frac{\sum_{ij} (A_{ij} - k_i k_j / 2m) x_i x_j}{\sum_{ij} (k_i \delta_{ij} - k_i k_j / 2m) x_i x_j}$$

$$r = \frac{\sum_{ij} (A_{ij} - k_i k_j / 2m) k_i k_j}{\sum_{ij} (k_i \delta_{ij} - k_i k_j / 2m) k_i k_j}$$

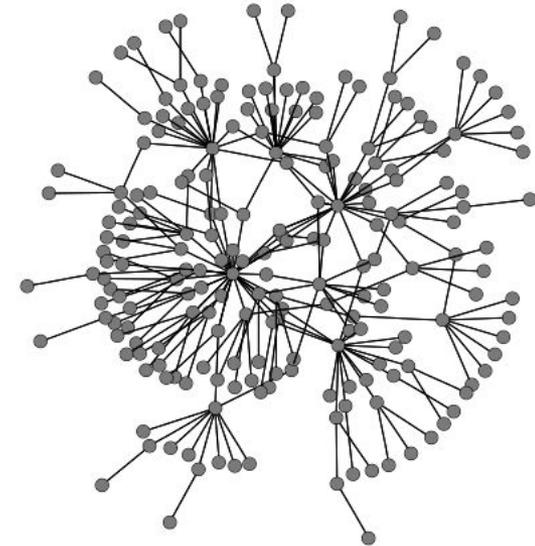
Estamos analizando una propiedad estructural (distribución de grado) que es función de 1-cuerpo en relación a otra (conectividad) función de 2-cuerpos

Red asortativa



Estructura núcleo-periferia

Red disortativa



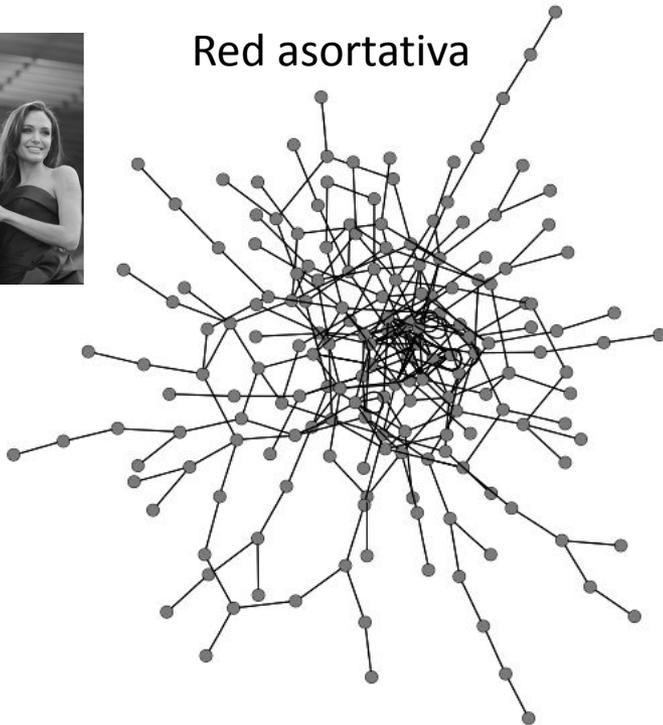
Estructura tipo estrella

# Assortative mixing por grado

$$r = \frac{\sum_{ij}(A_{ij} - k_i k_j / 2m) k_i k_j}{\sum_{ij}(k_i \delta_{ij} - k_i k_j / 2m) k_i k_j}$$

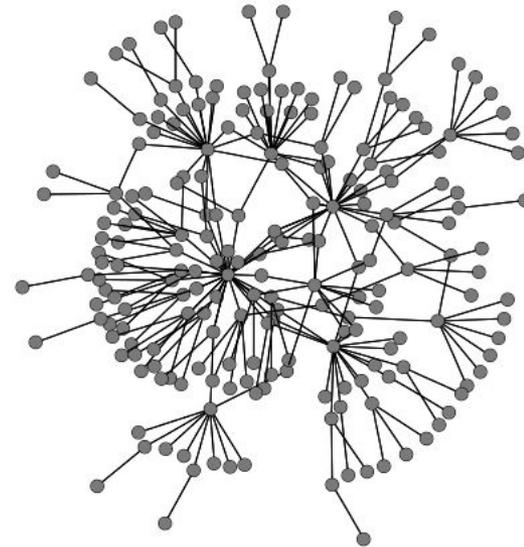


Red asortativa

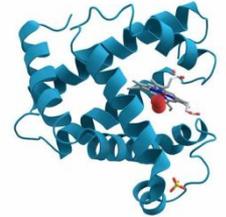


Estructura núcleo-periferia

Red disortativa



Estructura tipo estrella



# Topología de la red de proteínas

1458 proteínas  
1746 interacciones

Hubs tienden a conectarse con nodos de bajo grado...es *esperable* esto?

En una **red aleatoria** la probabilidad de que nodos  $i, j$  interactuen resulta

$$p_{k_i k_j} = \frac{k_i k_j}{2m}$$

$$p_{2,1} = 0.0006$$

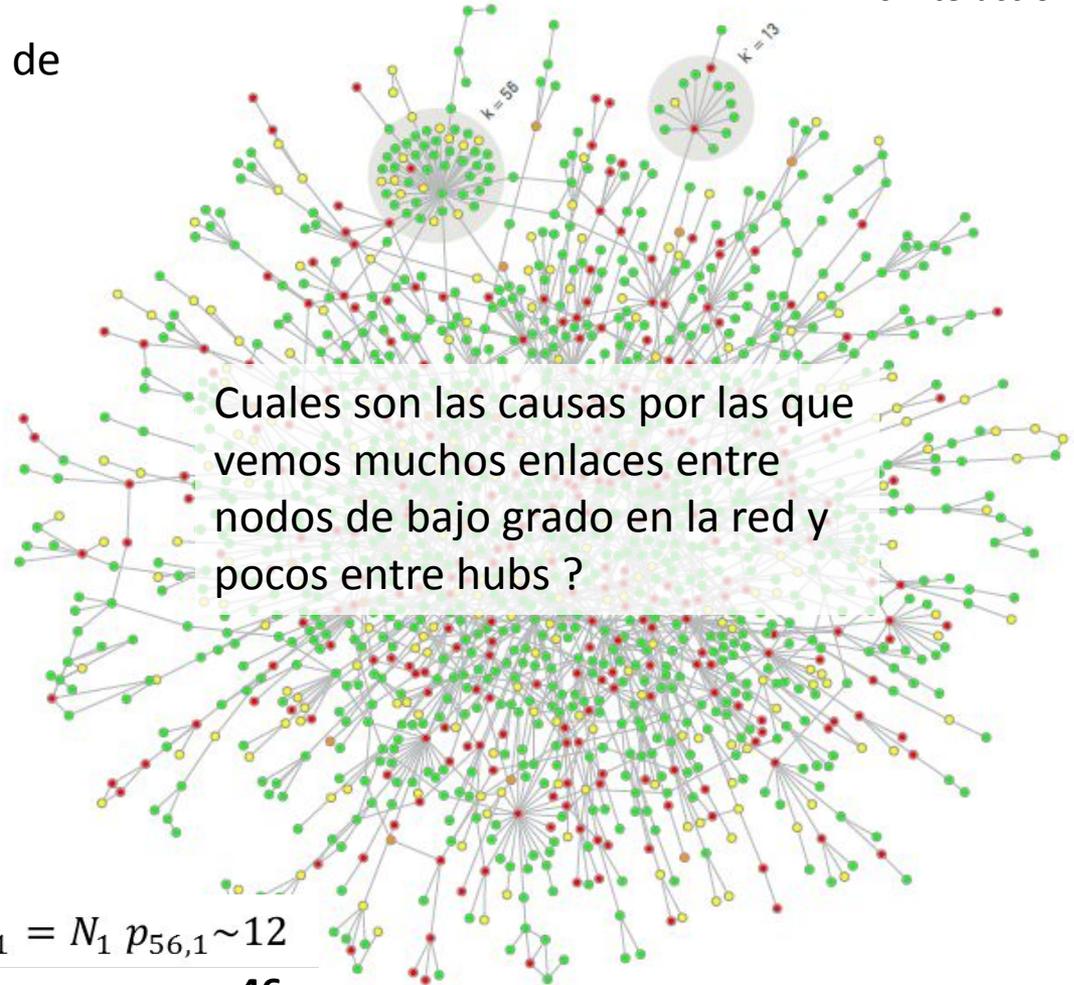
$$p_{56,13} = 0.2$$

...sin embargo no vemos un link entre  $k_{56}$  y  $k_{13}$  en la red real

$$p_{56,1} = 0.016$$

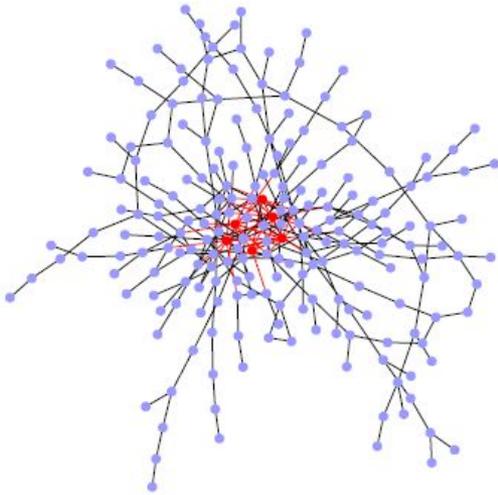
$$\text{\#enlaces esperados del hub a nodos } k_1 = N_1 p_{56,1} \sim 12$$

$$\text{\#enlaces observados}_{56,1} = 46$$

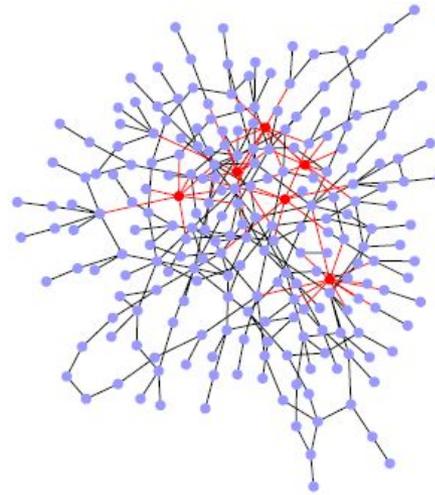


# Correlación de grado

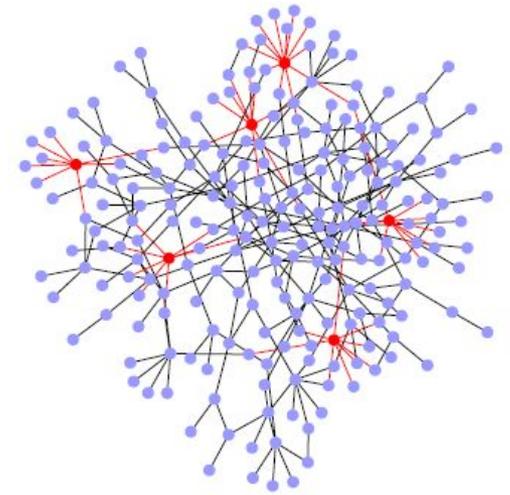
El carácter de asortatividad de grado repercute en la estructura de la red:



**Asortativa**  
Hubs conectados con hubs



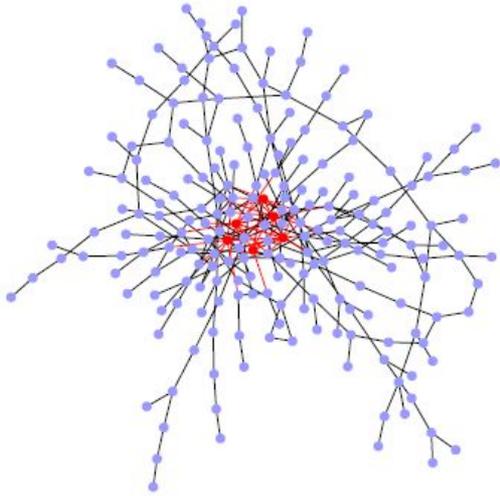
**Neutra**  
Nodos conectados como se  
espera del caso aleatorio



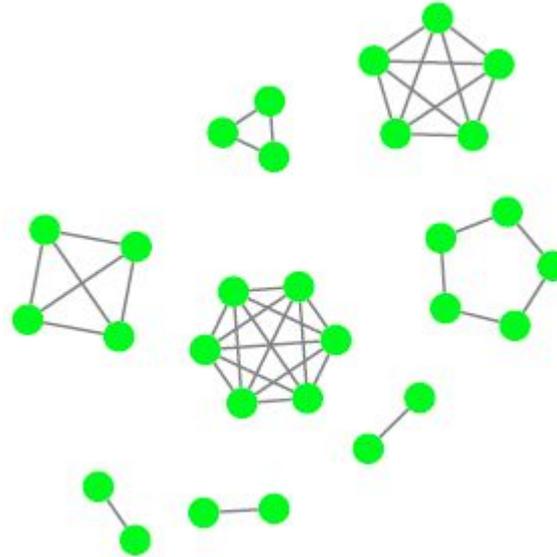
**Disortativa**  
Hubs evitan hubs

Maslov and Sneppen, Science 2001  
Pastor Satorras and Vespignani, PRL 2001  
Newman, PRL 2002

# Correlación de grado



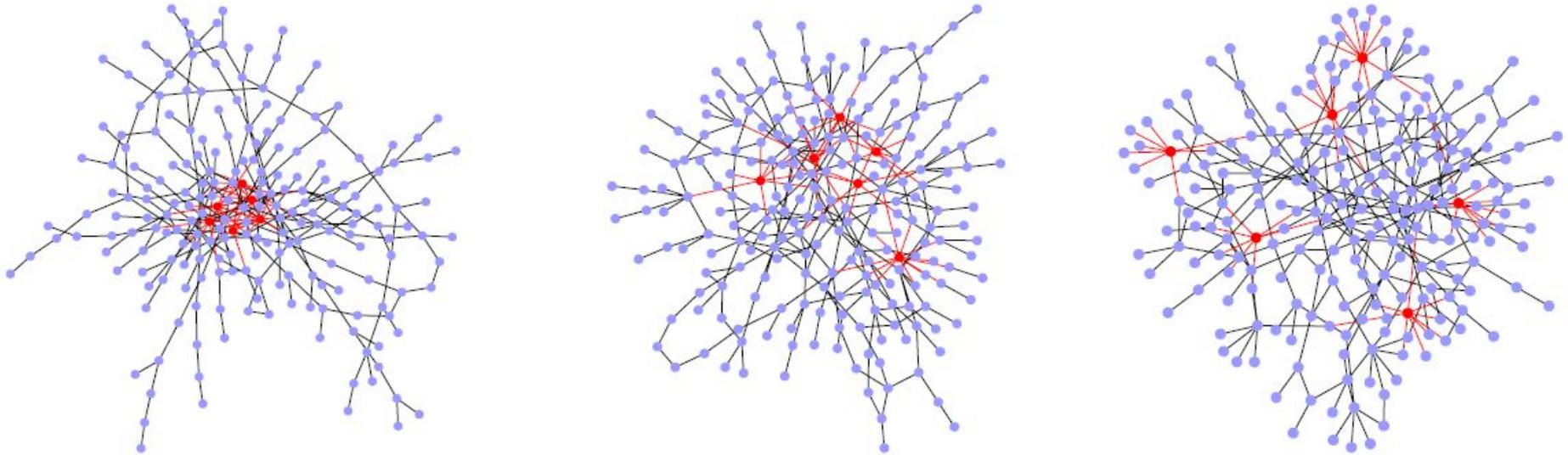
Asortativa  
Hubs conectados con hubs



Red perfectamente assortativa

Maslov and Sneppen, Science 2001  
Pastor Satorras and Vespignani, PRL 2001  
Newman, PRL 2002

# Correlación de grado



Dos maneras de cuantificar

$\text{cor}(k_i, k_j)$

$$r = \frac{\sum_{ij} (A_{ij} - k_i k_j / 2m) k_i k_j}{\sum_{ij} (k_i \delta_{ij} - k_i k_j / 2m) k_i k_j}$$

grado medio de vecinos →

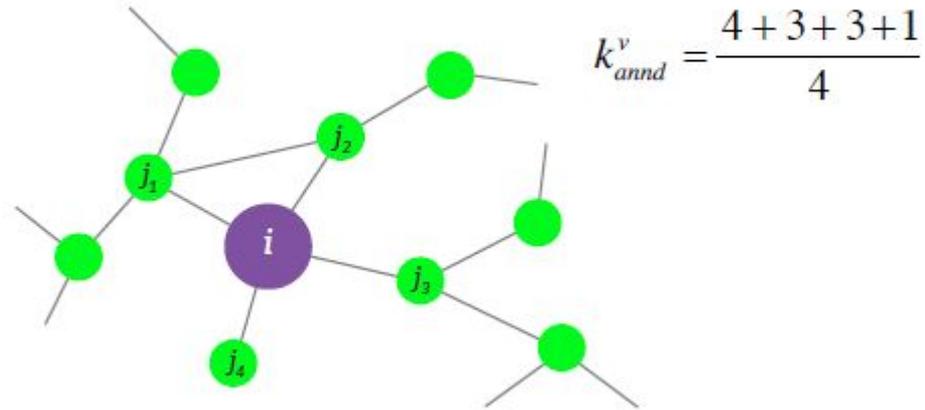
Pastor Satorras and Vespignani, PRL 2001

# Grado medio de vecinos (Average Nearest Neighbor Degree)

$k_{\text{annd}}(\mathbf{k})$ : grado medio de primeros vecinos de nodos con grado  $k$

Para el nodo- $i$ :

$$k_{nn,i} = \frac{1}{k_i} \sum_{j=1}^N A_{ij} k_j$$



En general, para nodos de grado  $k$ :

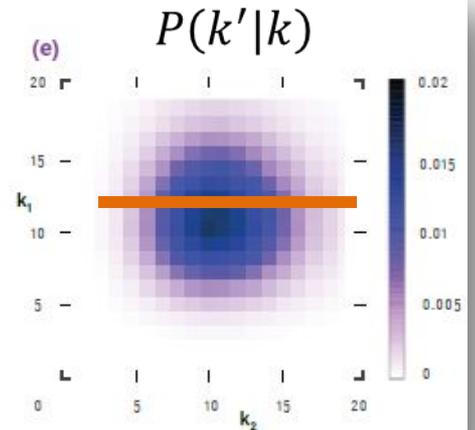
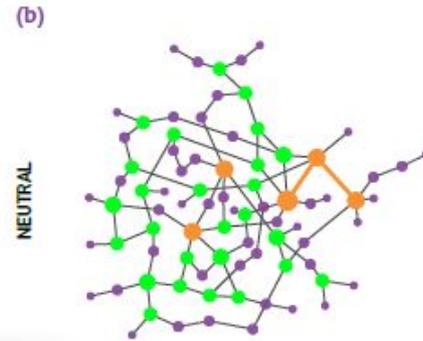
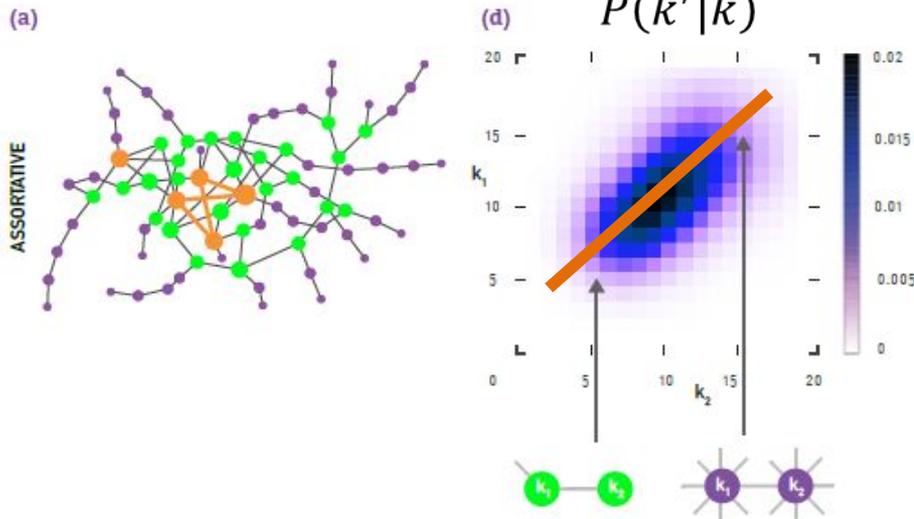
valor medio de  $k'$  dado  $k$

$$k_{nn}(k) = \sum_{k'} k' P(k'|k)$$

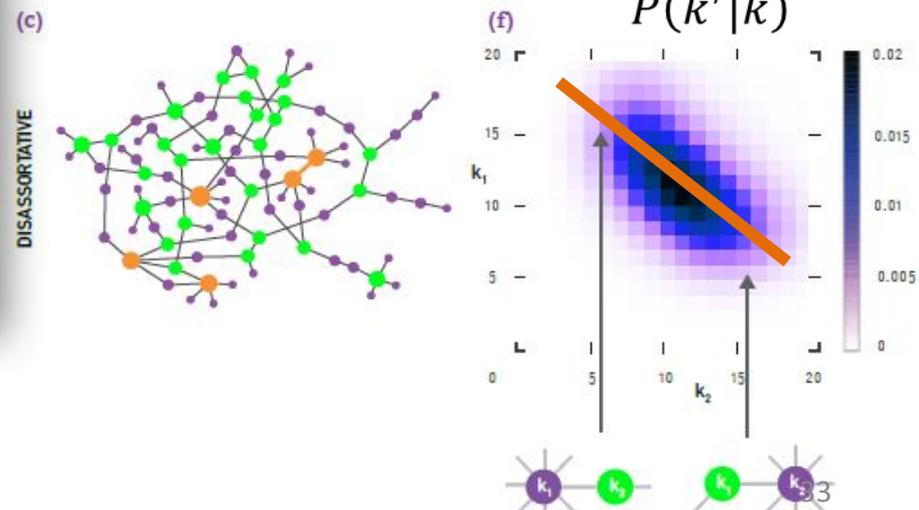
prob condicional de encontrar un nodo de grado  $k'$   
en un enlace que tiene un extremo de grado  $k$

# Asortatividad-Correlación de grado

$$k_{nn}(k) = \sum_{k'} k' P(k'|k)$$



Redes con misma distribución de grado (Poisson).  
En **naranja** los 5 nodos de mayor grado.

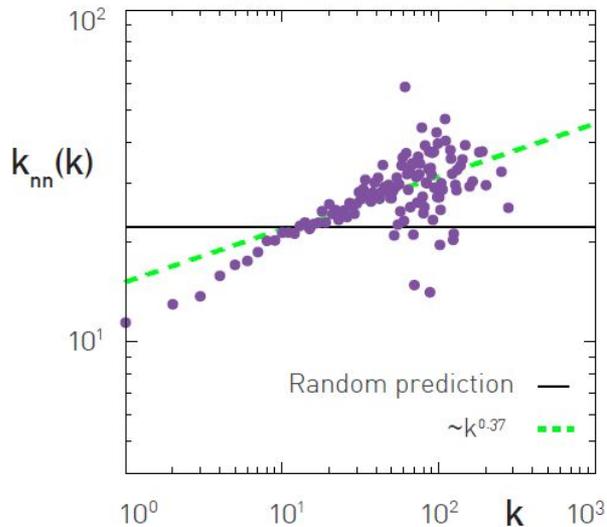


Las rectas representan fiteos a  $k_{nn}(k)$   
como función de  $k$  y hablan de redes de  
distinto caracter asortativo

# Exponente de correlación

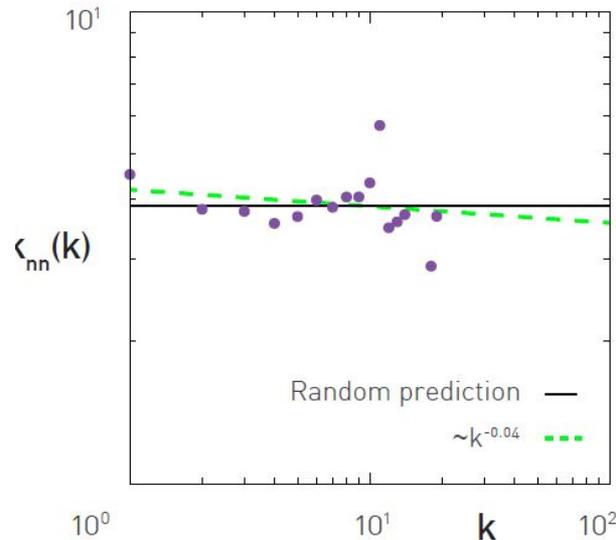
$$k_{nn}(k) = \sum_{k'} k' P(k'|k)$$

SCIENTIFIC COLLABORATION



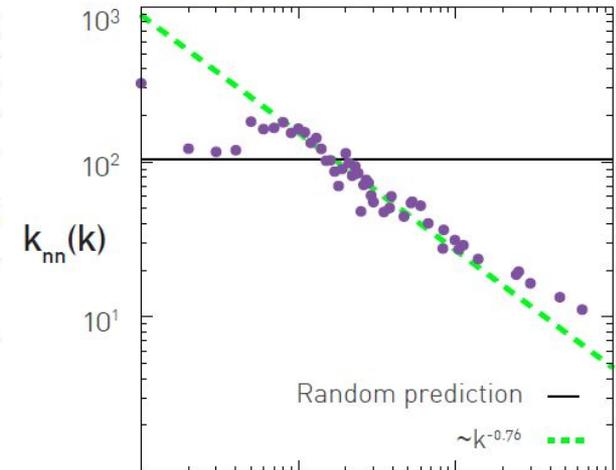
$$\mu = 0.37 \pm 0.11$$

POWER GRID



$$\mu = 0.04 \pm 0.05$$

METABOLIC NETWORK



$$\mu = -0.76 \pm 0.04$$

A partir de fiteos a redes reales  
Barabasi introduce la relación  
(notar **escala log** en las figs)

$$k_{nn}(k) = ak^\mu$$

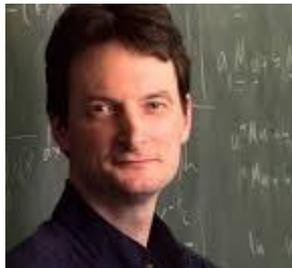
# Dos medidas de asortatividad

Coeficiente de correlación entre  
grados de vertices de un enlace

$$r = \frac{\sum_{ij}(A_{ij} - k_i k_j / 2m) k_i k_j}{\sum_{ij}(k_i \delta_{ij} - k_i k_j / 2m) k_i k_j}$$

Compatible con  $k_{nn}(k) \sim r k$

$$-1 \leq r \leq 1$$



Dos medidas (dos  
hipotesis) de  
asortatividad

Fiteo de relación power-law  
del Grado medio de vecinos  
de nodos de grado  $k$

$$k_{nn}(k) = \sum_{k'} k' P(k' | k)$$

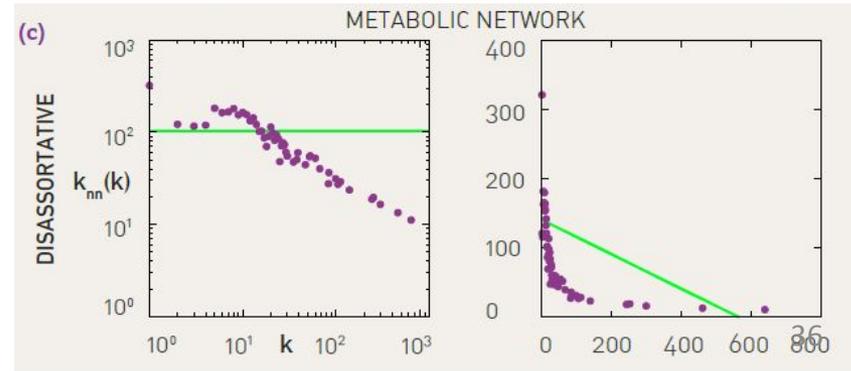
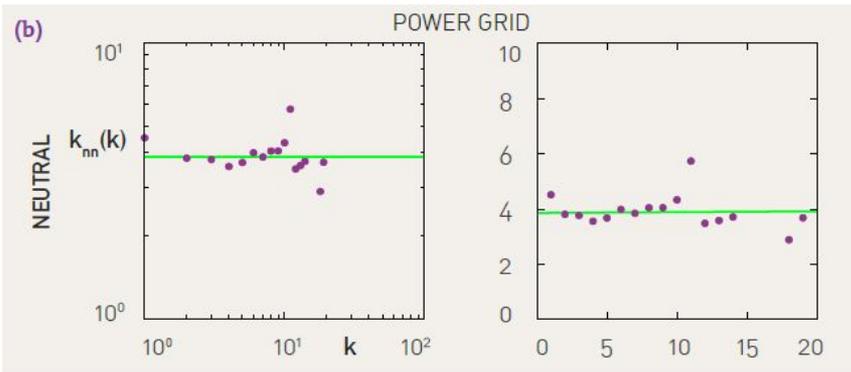
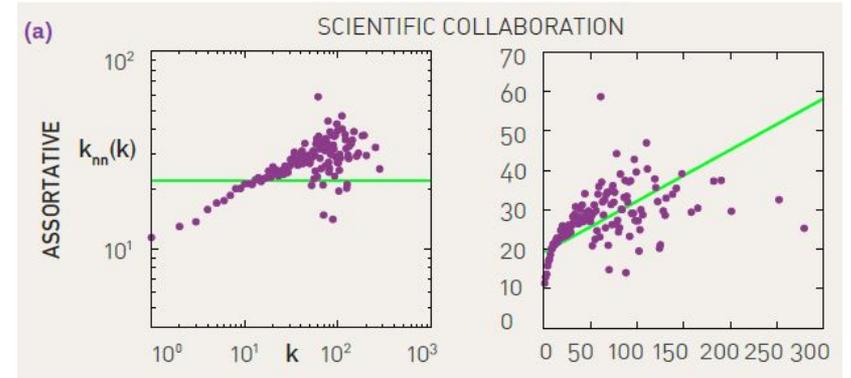
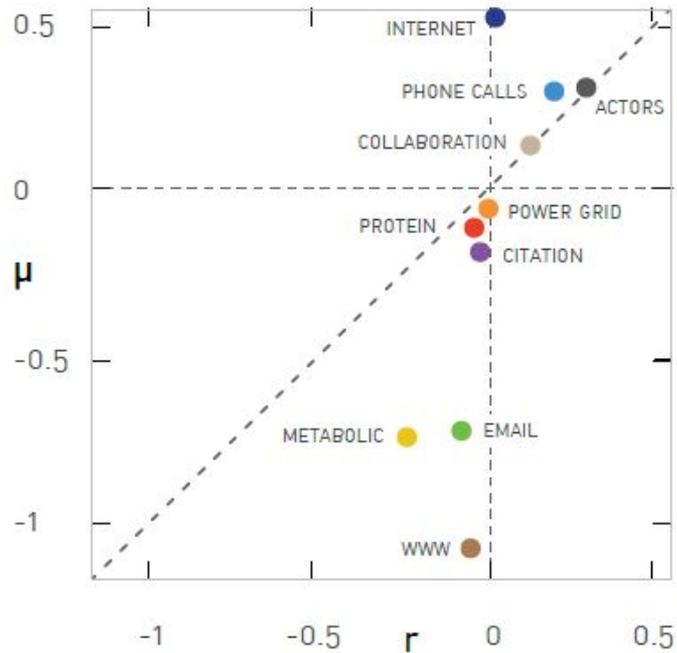
$$k_{nn}(k) = a k^\mu$$



# Dos medidas (modelos) de asortatividad

$$r = \frac{\sum_{ij}(A_{ij} - k_i k_j / 2m) k_i k_j}{\sum_{ij}(k_i \delta_{ij} - k_i k_j / 2m) k_i k_j} \quad k_{nn}(k) \sim r k$$

$$k_{nn}(k) = \sum_{k'} k' P(k'|k) \quad k_{nn}(k) = a k^\mu$$



- Estuvimos hablando de correlaciones en la red: cableado vs campos definidos en nodos

$$G = G(V, E, f: V \rightarrow \{c_i\})$$

$$Q = \frac{1}{2m} \sum_{ij} \left( A_{ij} - \frac{k_i k_j}{2m} \right) \delta(c_i, c_j)$$

$$Q = \sum_r (e_{rr} - a_r^2)$$

$$G = G(V, E, x: V \rightarrow \mathbb{R})$$

$$Q = \frac{1}{2m} \sum_{ij} \left( A_{ij} - \frac{k_i k_j}{2m} \right) x_i x_j$$

$$G = G(V, E, x: V \rightarrow k(v))$$

$$\mathbf{r} = \frac{\sum_{ij} (A_{ij} - k_i k_j / 2m) k_i k_j}{\sum_{ij} (k_i \delta_{ij} - k_i k_j / 2m) k_i k_j}$$

$$k_{nn}(k) = \sum_{k'} k' P(k'|k)$$



# Corte estructural

En redes **simples con distribuciones de grado de cola pesada** la presencia de hubs puede introducir un sesgo hacia la disortatividad...

Cuántos enlaces cabría esperar en una **red neutral** entre dos *hubs* de la red?

Recordemos...

fracción de enlaces que conectan nodos de grado  $r$  con nodos de grado  $s$

$$e_{rs} = \frac{1}{2m} \sum_{ij} A_{ij} \delta(k_i, r) \delta(k_j, s)$$

En una red neutral

$$e_{rs} = e_r e_s$$

fracción de enlaces adyacentes a un nodo de grado  $k$

$$\begin{aligned} e_k &= \frac{1}{2m} \sum_{ij} A_{ij} \delta(k_i, k) = \frac{1}{2m} \sum_i k_i \delta(k_i, k) \\ &= \frac{k p(k) N}{2m} = \frac{k p(k)}{\langle k \rangle} \end{aligned}$$

$$e_{kk'} = \frac{k p_k k' p_{k'}}{\langle k \rangle^2}$$

# Corte estructural

En redes **simples con distribuciones de grado de cola pesada** la presencia de hubs puede introducir un sesgo hacia la disortatividad...

Cuántos enlaces cabría esperar en una **red neutral** entre dos *hubs* de la red?

- Fracción de enlaces ( $k=55, k'=46$ ):

$$e_{kk'} = a_k a_{k'} = \frac{k p_k k' p_{k'}}{\langle k \rangle^2}$$

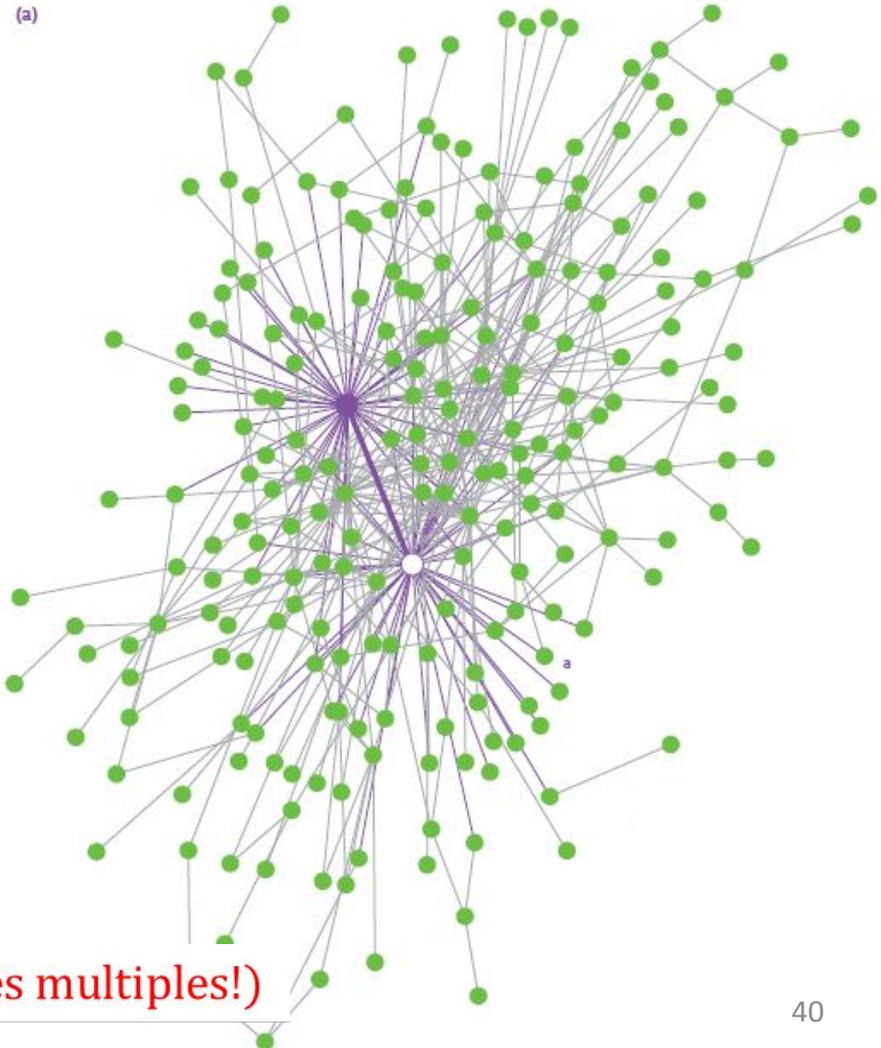
- Nro esperado de enlaces ( $k=55, k'=46$ ):

$$N_{kk'} = N \langle k \rangle e_{kk'} = N \frac{k p_k k' p_{k'}}{\langle k \rangle}$$

$$2m = N \langle k \rangle$$

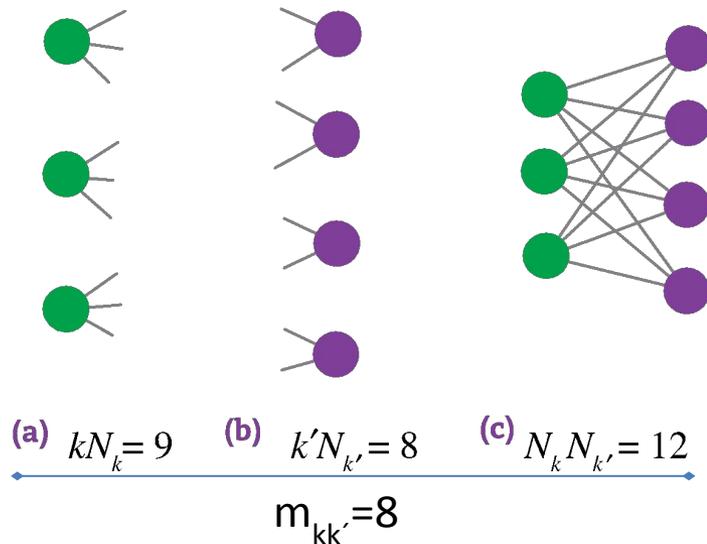
$$= 300 \frac{55 \cdot 46}{3} \sim 2.8 > 1 \text{ (enlaces multiples!)}$$

Red scale-free  $N=300, m=450, \gamma=2.2$   
Disortativa:  $r=-0.19$



# Corte estructural

**Máximo** número de enlaces entre dos grupos de nodos  $m_{kk'}$ :



tres numeros que limitan

$$m_{kk'} = \min(kN_k, k'N_{k'}, N_k N_{k'})$$

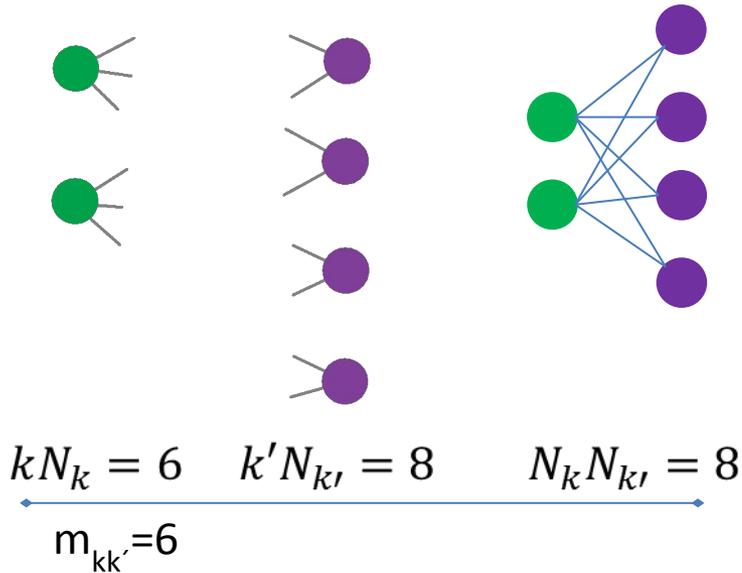
nro de enlaces adyacentes  
a grupo verde

nro max de enlaces  
posible entre ambos  
grupos

nro de enlaces adyacentes  
a grupo violeta

# Corte estructural

**Máximo** número de enlaces entre dos grupos de nodos  $m_{kk'}$ :  $m_{2,3}$



tres numeros que limitan

$$m_{kk'} = \min(kN_k, k'N_{k'}, N_k N_{k'})$$

nro de enlaces adyacentes  
a grupo verde

nro max de enlaces  
posible entre ambos  
grupos

nro de enlaces adyacentes  
a grupo violeta

# Corte estructural

**Máximo** número de enlaces entre dos grupos de nodos  $m_{kk'}$ :  $m_{2,3}$



tres numeros que limitan

$$m_{kk'} = \min(kN_k, k'N_{k'}, N_k N_{k'})$$

nro de enlaces adyacentes  
a grupo verde

nro max de enlaces  
posible entre ambos  
grupos

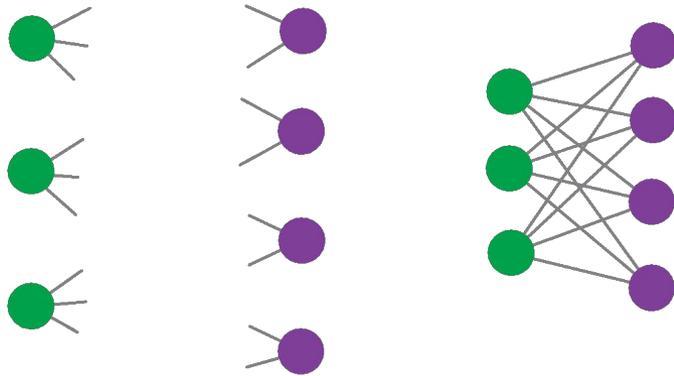
nro de enlaces adyacentes  
a grupo violeta

$$kN_k = 3 \quad k'N_{k'} = 2 \quad N_k N_{k'} = 1$$

$m_{kk'} = 1$

# Corte estructural

**Máximo** número de enlaces entre dos grupos de nodos  $m_{kk'}$ :



tres numeros que limitan

$$m_{kk'} = \min(kN_{k'}, k'N_k, N_k N_{k'})$$

nro de enlaces adyacentes  
a grupo verde

nro de enlaces adyacentes  
a grupo violeta

nro max de enlaces  
posible entre ambos  
grupos

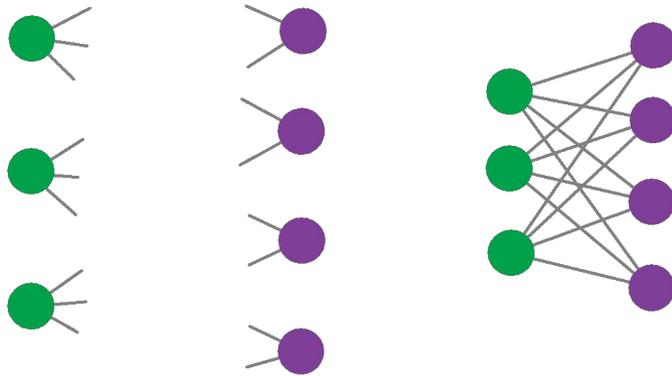
Definimos:

$$r_{kk'} = \frac{N_{kk'}}{m_{kk'}}$$

Si  $r_{kk'} \leq 1$  no actua ningun corte  
estructural para lo que vemos en la red

# Corte estructural

**Máximo** número de enlaces entre dos grupos de nodos  $m_{kk'}$ :



tres numeros que limitan

$$m_{kk'} = \min(kN_k, k'N_{k'}, N_k N_{k'})$$

nro de enlaces adyacentes  
a grupo verde

nro max de enlaces  
posible entre ambos  
grupos

nro de enlaces adyacentes  
a grupo violeta

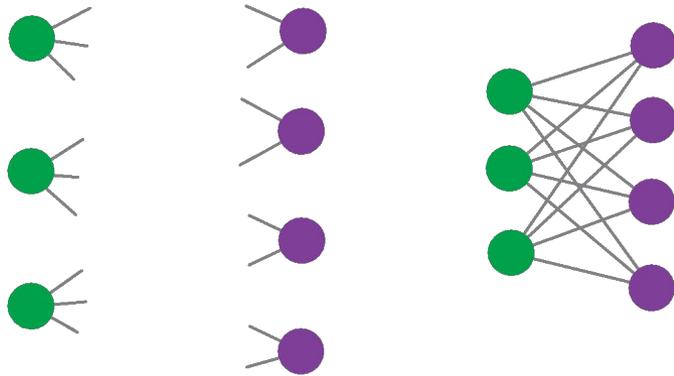
Definimos:

$$r_{kk'} = \frac{N_{kk'}}{m_{kk'}} = \frac{N\langle k \rangle e_{kk'}}{m_{kk'}} = \frac{\langle k \rangle e_{kk'}}{\min\{kp(k), k'p(k'), \underline{Np(k)p(k')}\}} \leq 1$$

Vamos a ver que para hubs ( $k, k'$  gdes) **la tercera condicion** es la limitante

# Corte estructural

**Máximo** número de enlaces entre dos grupos de nodos  $m_{kk'}$ :



**$k, k'$  gdes:**

si  $k > Np_{k'}$

$kp(k) > Np(k')p(k)$

$k' > Np_k$

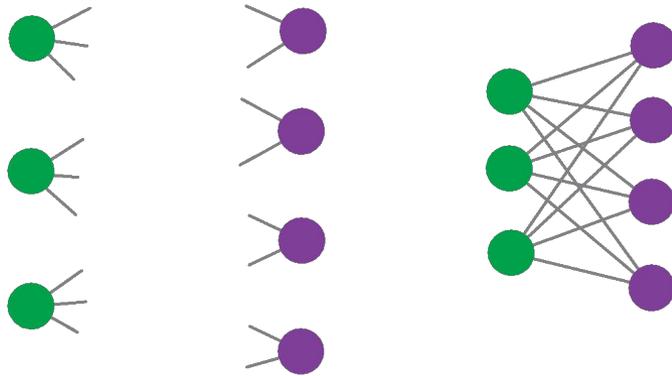
$k'p(k') > Np(k)p(k')$

Definimos:

$$r_{kk'} = \frac{N_{kk'}}{m_{kk'}} = \frac{N\langle k \rangle e_{kk'}}{m_{kk'}} = \frac{\langle k \rangle e_{kk'}}{\min\{kp(k), k'p(k'), Np(k)p(k')\}} \leq 1$$

# Corte estructural

**Máximo** número de enlaces entre dos grupos de nodos  $m_{kk'}$ :



$$\text{si } k > Np_{k'}$$

$$kp(k) > Np(k')p(k)$$

$$k' > Np_k$$

$$k'p(k') > Np(k)p(k')$$

Definimos:

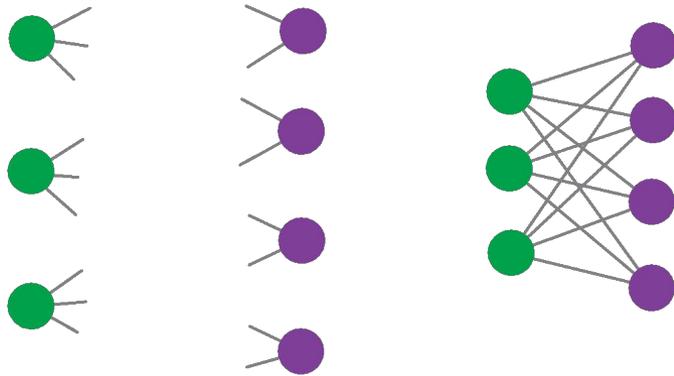
$$r_{kk'} = \frac{N_{kk'}}{m_{kk'}} = \frac{N\langle k \rangle e_{kk'}}{m_{kk'}} = \frac{\langle k \rangle e_{kk'}}{Np(k)p(k')} = \frac{\langle k \rangle \frac{kk'p(k)p(k')}{\langle k \rangle^2}}{Np(k)p(k')} = \frac{kk'}{\langle k \rangle N}$$

para una red neutral

$$e_{kk'} = a_k a_{k'} = \frac{kp_k k'p_{k'}}{\langle k \rangle^2}$$

# Corte estructural

**Máximo** número de enlaces entre dos grupos de nodos  $m_{kk'}$ :



condicion límite

$$r_{kk'} = \frac{kk'}{\langle k \rangle N} = 1$$

corte estructural ( $k \sim k' \sim k_s$ )

$$k_s \sim \sqrt{\langle k \rangle N}$$

Definimos:

$$r_{kk'} = \frac{N_{kk'}}{m_{kk'}} = \frac{N \langle k \rangle e_{kk'}}{m_{kk'}} = \frac{\langle k \rangle e_{kk'}}{N p(k) p(k')} = \frac{\langle k \rangle \frac{kk' p(k) p(k')}{\langle k \rangle^2}}{N p(k) p(k')} = \frac{kk'}{\langle k \rangle N}$$

$$r_{kk'} = \frac{kk'}{\langle k \rangle N}$$

para una red neutral

$$e_{kk'} = a_k a_{k'} = \frac{k p_k k' p_{k'}}{\langle k \rangle^2}$$

# Consecuencias del corte estructural

Supongamos una red con distribución de grado tipo ley de potencia (cola pesada)  
 Por efecto de tamaño finito hay dos escalas:

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

$$k_s \sim \sqrt{\langle k \rangle N^{\frac{1}{2}}}$$

impone un umbral tal que nodos con  $k > k_s$   
 presentarán  $E_{kk'} > I$  en redes neutras

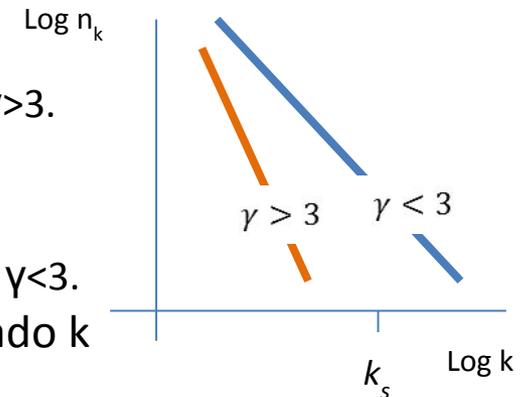
Existen dos regimenes para un grafo simple

**Grafo sin corte estructural:** redes aleatorias y libres de escala con  $\gamma > 3$ .

El exponente de  $k_{max}$  es menor que  $\frac{1}{2}$  y  $k_s > k_{max}$

**Grafo con disasortatividad estructural:** redes libres de escala con  $\gamma < 3$ .

El exponente de  $k_{max}$  es mayor que  $\frac{1}{2}$  y  $k_s < k_{max}$ . Nodos con grado  $k \in [k_s, k_{max}]$  tendrán, si el grafo es simple, forzosamente menos enlaces que en el caso neutro.



# (Dis)asortatividad estructural o inherente?

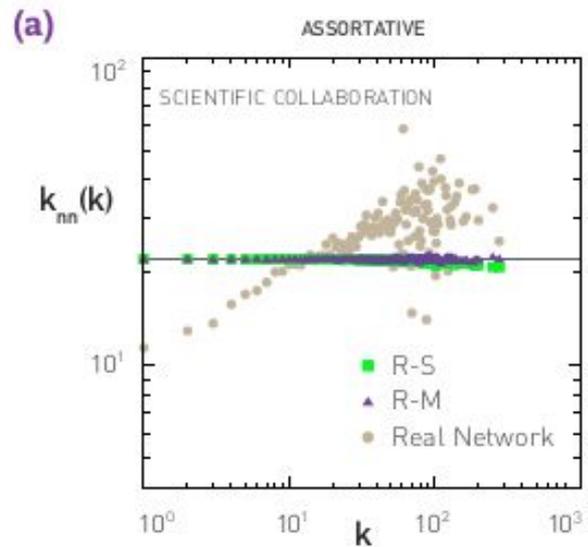
Para ver si la **asortatividad o disasortatividad** de una red real es **estructural** o se debe a un **proceso de la formación** de la red...comparación con modelo nulo

# (Dis)asortatividad estructural o inherente?

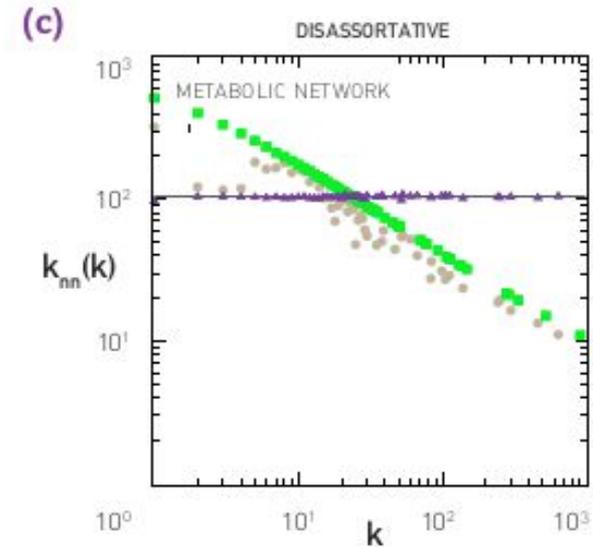
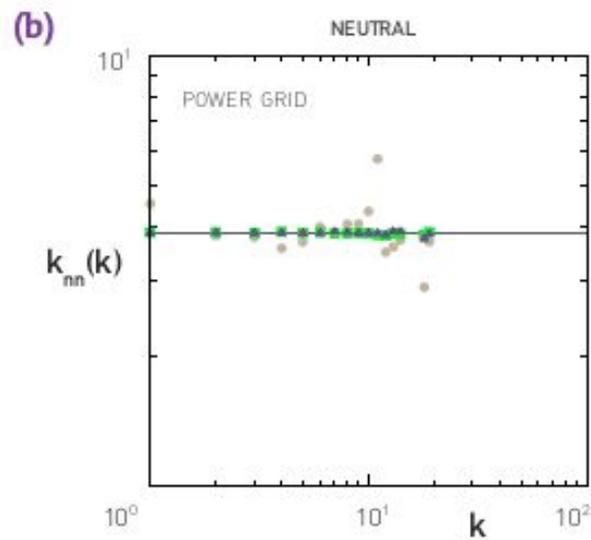
Para ver si la **asortatividad o disasortatividad** de una red real es **estructural** o se debe a un **proceso de la formación** de la red...comparación con modelo nulo

R-S: modelo configuracional, restringido a grafos simples

R-M: modelo configuracional que permite multi-enlaces (no es afectado por corte estructural)

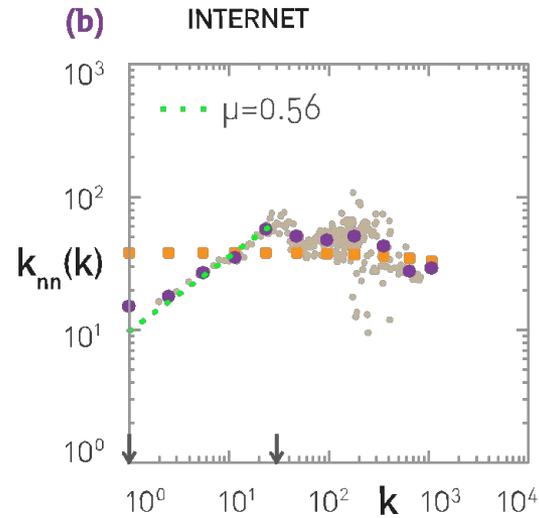
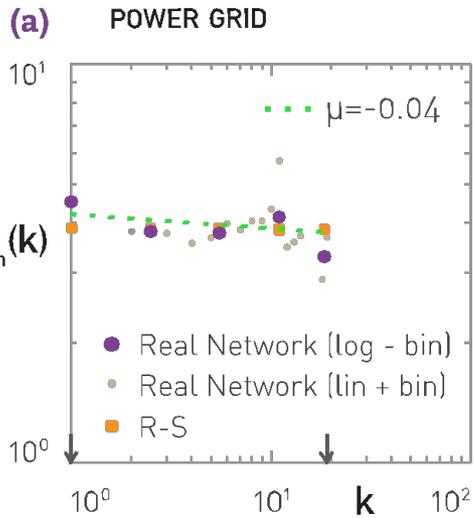


Patrón de correlaciones entre grados **no se explican** por propiedades de la distribución de grado...hay algo más



Patrón de correlaciones entre grados **se explican** enteramente por propiedades de la distribución de grado

# Mas redes reales (tecnológicas)



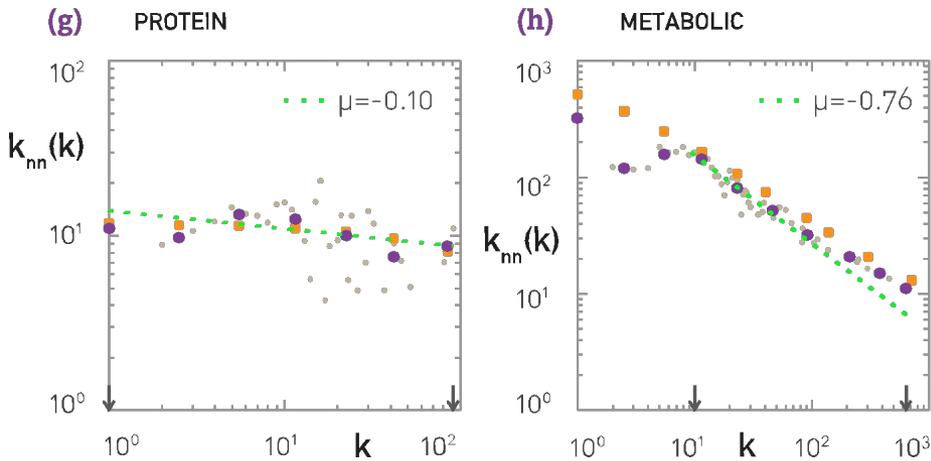
- **Power Grid**

For the power grid  $k_{nn}(k)$  is flat and indistinguishable from its randomized version, indicating a lack of degree correlations (Figure 7.10a). Hence the power grid is neutral.

- **Internet**

For small degrees ( $k \leq 30$ )  $k_{nn}(k)$  shows a clear assortative trend, an effect that levels off for high degrees (Figure 7.10b). The degree correlations vanish in the randomized version of the Internet map. Hence the Internet is assortative, but structural cutoffs eliminate the effect for high  $k$ .

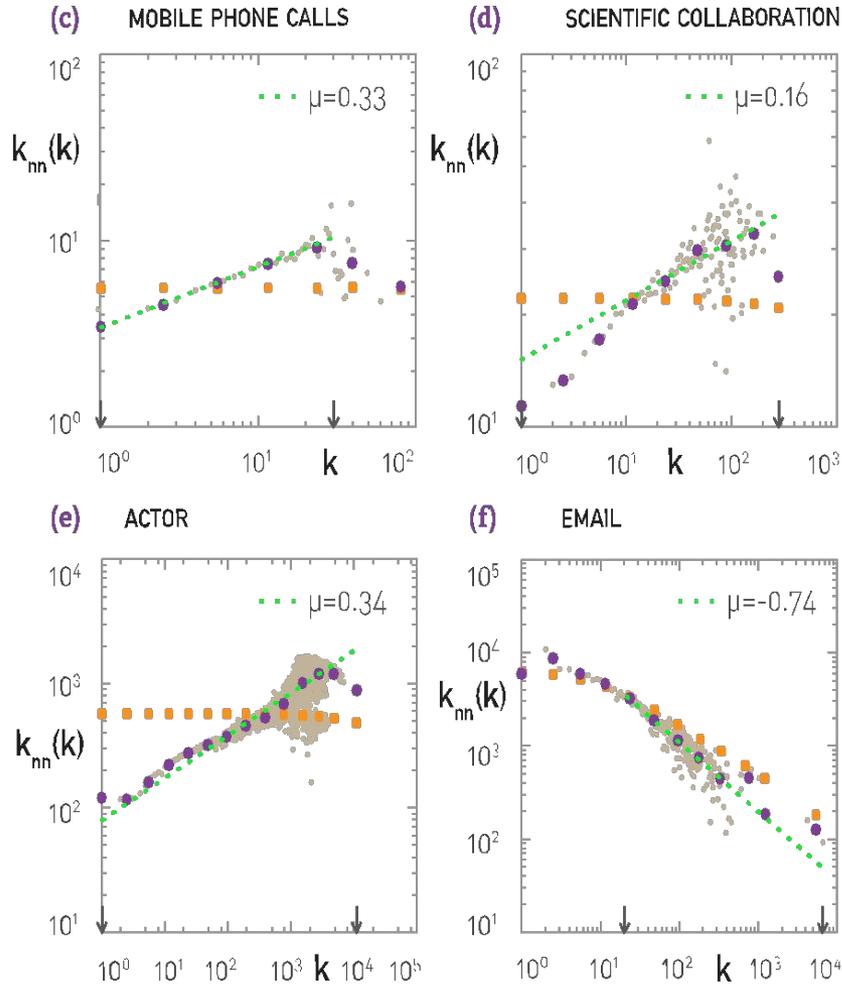
# Redes reales (biológicas)



- **Biological Networks**

The protein interaction and the metabolic network both have a negative  $\mu$ , suggesting that these networks are disassortative. Yet, the scaling of  $k_{min}^{R-S}(k)$  is indistinguishable from  $k_{nn}(k)$ , indicating that we are observing structural disassortativity, rooted in the scale-free nature of these networks (Figure 7.10 g,h).

# Redes reales (sociales)



- **Social Networks**

The three networks capturing social interactions, the mobile phone network, the science collaboration network and the actor network, all have an increasing  $k_{nn}(k)$ , indicating that they are assortative (Figures 7.10c-e). Hence in these networks hubs tend to link to other hubs and low-degree nodes tend to link to low-degree nodes. The fact that the observed  $k_{nn}(k)$  differs from the  $k_{nn}^{R-S}(k)$ , indicates that the assortative nature of social networks is not due to their scale-free the degree distribution.

- **Email Network**

While the email network is often seen as a social network, its  $k_{nn}(k)$  decreases with  $k$ , documenting a clear disassortative behavior (Figure 7.10f). The randomized  $k_{nn}^{R-S}(k)$  also decays, indicating that we are observing structural disassortativity, a consequence of the network's scale-free nature.

# Asortatividad en redes

	Network	Type	$n$	$m$	$c$	$S$	$\ell$	$\alpha$	$C$	$C_{ws}$	$r$	Ref(s).
Social	Film actors	Undirected	449913	25516482	113.43	0.980	3.48	2.3	0.20	0.78	0.208	16,323
	Company directors	Undirected	7673	55392	14.44	0.876	4.60	-	0.59	0.88	0.276	88,253
	Math coauthorship	Undirected	253339	496489	3.92	0.822	7.57	-	0.15	0.34	0.120	89,146
	Physics coauthorship	Undirected	52909	245300	9.27	0.838	6.19	-	0.45	0.56	0.363	234,236
	Biology coauthorship	Undirected	1520251	11803064	15.53	0.918	4.92	-	0.088	0.60	0.127	234,236
	Telephone call graph	Undirected	47000000	80000000	3.16				2.1			9,10
	Email messages	Directed	59812	86300	1.44	0.952	4.95	1.5/2.0		0.16		103
	Email address books	Directed	16881	57029	3.38	0.590	5.22	-	0.17	0.13	0.092	248
	Student dating	Undirected	573	477	1.66	0.503	16.01	-	0.005	0.001	-0.029	34
	Sexual contacts	Undirected	2810						3.2			197,198
Information	WWW nd. edu	Directed	269504	1497135	5.55	1.000	11.27	2.1/2.4	0.11	0.29	-0.067	13,28
	WWW AltaVista	Directed	203549046	1466000000	7.20	0.914	16.18	2.1/2.7				56
	Citation network	Directed	783339	6716198	8.57			3.0/-				280
	Roget's Thesaurus	Directed	1022	5103	4.99	0.977	4.87	-	0.13	0.15	0.157	184
	Word co-occurrence	Undirected	460902	16100000	66.96	1.000			2.7		0.44	97,116
Technological	Internet	Undirected	10697	31992	5.98	1.000	3.31	2.5	0.035	0.39	-0.189	66,111
	Power grid	Undirected	4941	6594	2.67	1.000	18.99	-	0.10	0.080	-0.003	323
	Train routes	Undirected	587	19603	66.79	1.000	2.16	-		0.69	-0.033	294
	Software packages	Directed	1439	1723	1.20	0.998	2.42	1.6/1.4	0.070	0.082	-0.016	239
	Software classes	Directed	1376	2213	1.61	1.000	5.40	-	0.033	0.012	-0.119	315
	Electronic circuits	Undirected	24097	53248	4.34	1.000	11.05	3.0	0.010	0.030	-0.154	115
	Peer-to-peer network	Undirected	880	1296	1.47	0.805	4.28	2.1	0.012	0.011	-0.366	6,282
Biological	Metabolic network	Undirected	765	3686	9.64	0.996	2.56	2.2	0.090	0.67	-0.240	166
	Protein interactions	Undirected	2115	2240	2.12	0.689	6.80	2.4	0.072	0.071	-0.156	164
	Marine food web	Directed	134	598	4.46	1.000	2.05	-	0.16	0.23	-0.263	160
	Freshwater food web	Directed	92	997	10.84	1.000	1.90	-	0.20	0.087	-0.326	209
	Neural network	Directed	307	2359	7.68	0.967	3.97	-	0.18	0.28	-0.226	323,328

## Observaciones

- en general las redes resultan disortativas
- salvo las de tipo social, que suelen resultar asortativas.

## Causas :

- Efectos de tener hubs en grafos simples (disasortatividad estructural)
- predominio de grupos pequeños con nodos de bajo grado conectados entre sí dentro del grupo

# Estructura a gran escala I

De acuerdo a la distribución de grado las redes pueden clasificarse en

## Redes acotadas exponencialmente:

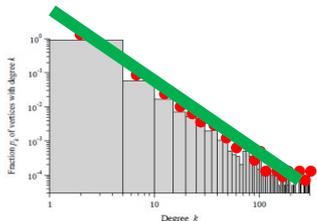
- distribución decae para altos valores de  $k$  exponencialmente o más rapido
- Ausencia de grandes fluctuaciones en el grado [  $\langle k^2 \rangle < \langle k \rangle^2$  ]

## Redes de cola pesada:

- redes cuya distribución decae como ley de potencia para altos valores de  $k$
- Pueden presentarse grandes fluctuaciones en el grado [  $\langle k^2 \rangle \gg \langle k \rangle^2$  ]
- Típicamente presentan outliers (hubs)

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

$$\log p_k = -\gamma \log k + c$$



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

# Estructura a gran escala II

- Asortatividad / disortatividad

variable categorica  $Q = \frac{1}{2m} \sum_{ij} \left( A_{ij} - \frac{k_i k_j}{2m} \right) \delta(c_i, c_j)$

campo escalar  $\text{cov}(x_i, x_j) = \frac{1}{2m} \sum_{ij} \left( A_{ij} - \frac{k_i k_j}{2m} \right) x_i x_j$

grado  $\mathbf{r} = \frac{\sum_{ij} (A_{ij} - k_i k_j / 2m) k_i k_j}{\sum_{ij} (k_i \delta_{ij} - k_i k_j / 2m) k_i k_j}$

$$k_{\text{nn}}(k) = ak^\mu$$

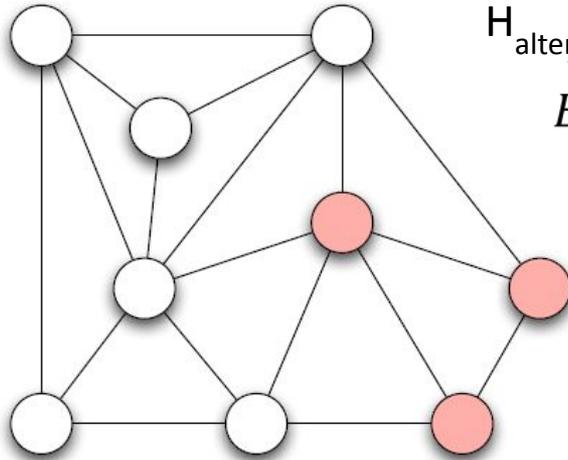
corte estructural

$$k_s \sim \sqrt{\langle k \rangle N}$$



# Recordemos...

## Cuantificando homofilia



$H_{\text{alternativa-}}$

$$G = G(V, E, f_v: V \rightarrow \{\text{color1}, \dots, \text{colorN}\})$$

$E \leftrightarrow f_v$

- Existe una tendencia a establecer enlaces entre nodos de igual color?
  - hay *muchos* enlaces entre nodos de igual color?
  - hay *pocos* enlaces entre nodos de color diferente?

Contra qué cosa comparo para saber si hay **pocos** enlaces entre nodos de diferente color?

$H_{\text{null-1}}$ : modelo de *la bolsa* tengo 6 bolas blancas y 3 salmon en una bolsa. Saco 18 pares **al azar**

$H_{\text{null-2}}$ : modelo de *recolorado* Sobre la red original, coloreo **al azar** 6 blancas y 3 salmon ( $f_v$ )

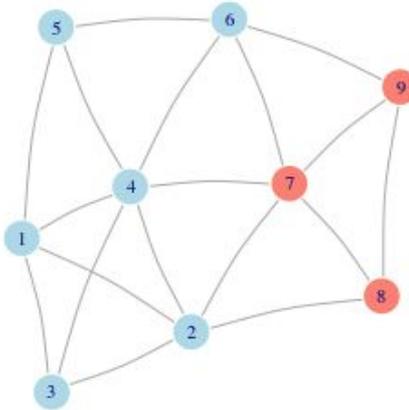
$H_{\text{null-3}}$ : modelo de *recableado* Recableo la red cambiando **al azar** la estructura del grafo ( $E$ )

$H_{\text{null-4}}$ : modelo de *recableado configuracional* Recableo la red cambiando **al azar** la estructura del grafo **respetando el grado** de cada nodo ( $E$ )

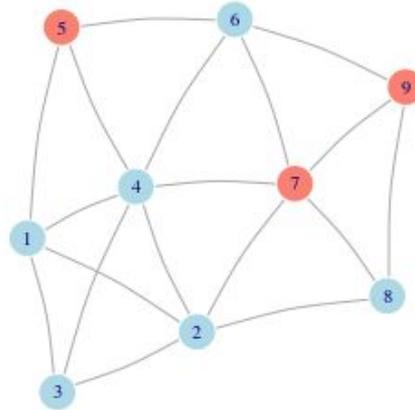
# Recordemos...

Cuantificando homofilia

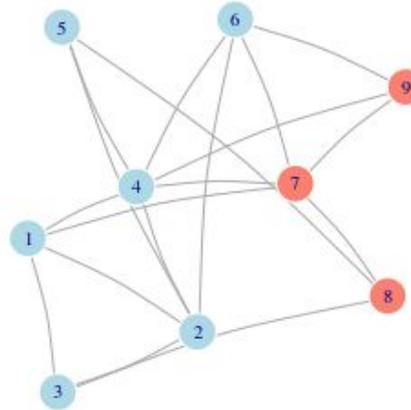
real



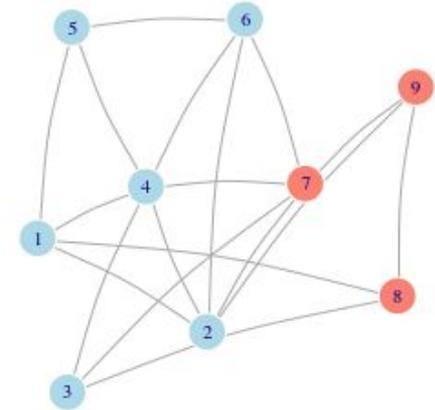
color shuffling



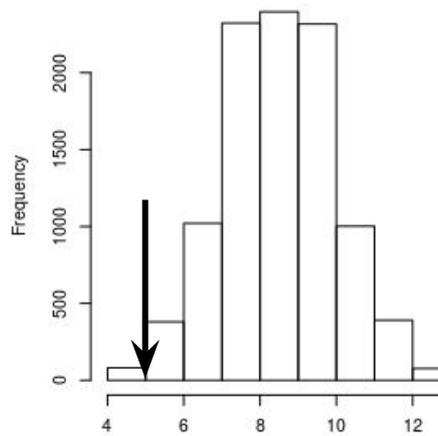
rewired



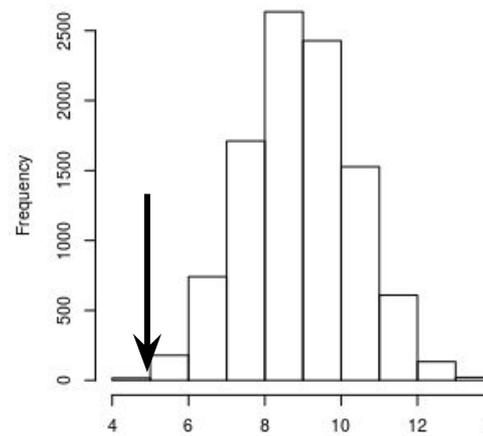
degree preserving



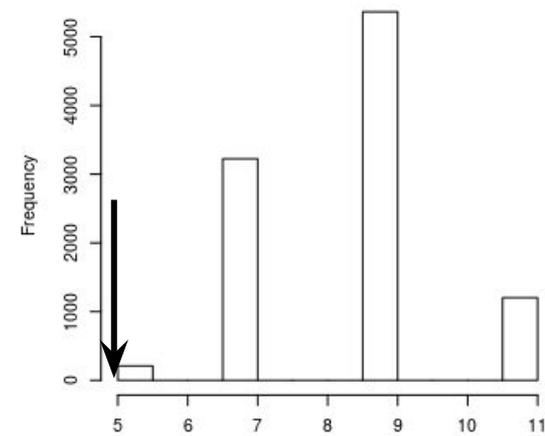
mean= 9



mean= 9.4

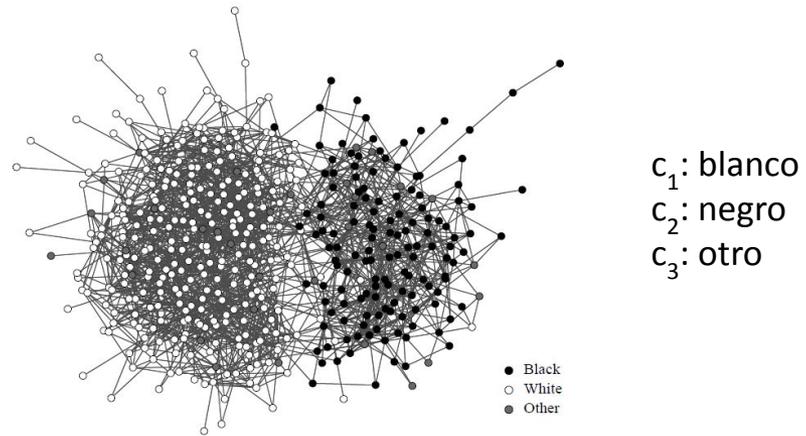


mean= 8.5



# Assortative mixing: características categóricas

Supongamos que existen  $n_c$  clases diferentes para los  $n$  nodos de una red de  $m$  enlaces.  
 Sea  $c_i$  la clase del nodo- $i$ . El número de enlaces entre mismo tipo de nodos resulta:



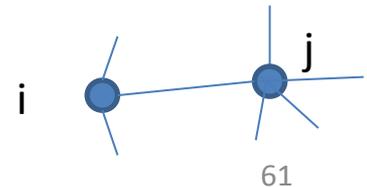
## Red real

$$\sum_{\text{edges } (i,j)} \delta(c_i, c_j) = \frac{1}{2} \sum_{ij} A_{ij} \delta(c_i, c_j) \quad \delta(c_i, c_j) = 1 \text{ si } c_i = c_j$$

## Red aleatoria (recableado)

$$\frac{1}{2} \sum_{ij} k_i \frac{k_j}{2m} \delta(c_i, c_j)$$

chances de que nodo- $j$  esté al otro extremo de un enlace del nodo- $i$



# Assortative mixing: características categóricas

Otra manera de computar asortatividad/modularidad, en términos de:

- la fracción de enlaces entre nodos de la clases  $r$  y  $s$  (o sea, probabilidad de encontrar un nodo de clase  $r$  y otro de clase  $s$  en los extremos de un enlace).

$$e_{rs} = \frac{1}{2m} \sum_{ij} A_{ij} \delta(c_i, r) \delta(c_j, s)$$

- La fracción de enlaces adyacente a nodos del tipo  $r$  (o sea, probabilidad de encontrar un nodo de tipo  $r$  en un extremo de un enlace)

$$\begin{aligned} a_r &= \sum_s e_{rs} = \frac{1}{2m} \sum_s \sum_{ij} A_{ij} \delta(c_i, r) \delta(c_j, s) \\ &= \frac{1}{2m} \sum_{ij} A_{ij} \delta(c_i, r) \sum_s \delta(c_j, s) \end{aligned}$$

$$a_r = \frac{1}{2m} \sum_{ij} A_{ij} \delta(c_i, r)$$

En una red sin correlaciones

$$e_{rs} = a_r a_s$$

# Assortative mixing: características categóricas

$$e_{rs} = \frac{1}{2m} \sum_{ij} A_{ij} \delta(c_i, r) \delta(c_j, s)$$

$$a_r = \frac{1}{2m} \sum_i k_i \delta(c_i, r)$$

En una red sin correlaciones  $e_{rs} = a_r a_s$

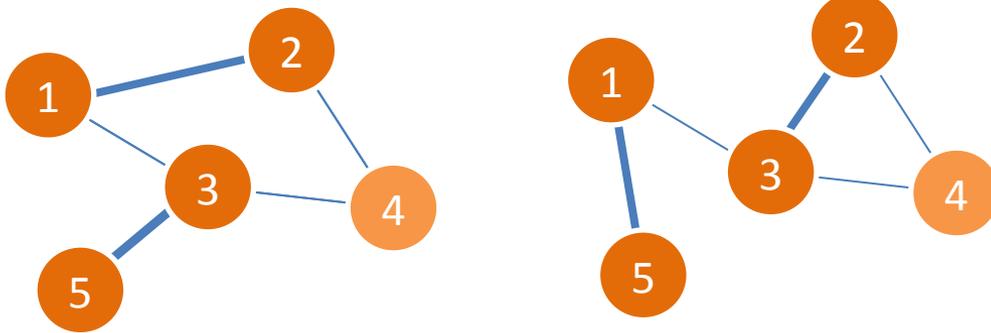
Desviaciones de este valor indican grado de correlación entre conexas y la caracterización externa de los nodos

$$Q = \frac{1}{2m} \sum_{ij} \left( A_{ij} - \frac{k_i k_j}{2m} \right) \overbrace{\sum_r \delta(c_i, r) \delta(c_j, r)}^{\delta(c_i, c_j)}$$
$$= \sum_r \left[ \frac{1}{2m} \sum_{ij} A_{ij} \delta(c_i, r) \delta(c_j, r) - \frac{1}{2m} \sum_i k_i \delta(c_i, r) \frac{1}{2m} \sum_j k_j \delta(c_j, r) \right]$$

$$Q = \sum_r (e_{rr} - a_r^2)$$

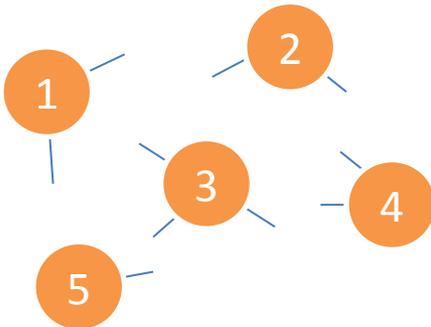
# Modelo configuracional

Recableado de la red, **preservando la distribución de grado** original



- elijo 2 pares de nodos enlazados
  - intercambio parejas
- Repito a y b muchísimas veces

Este recableado preserva el grado de cada nodo



En la red recableada, la probabilidad de que exista un enlace entre el nodo  $i$  y el  $j$

$$A_{ij} = k_i \frac{k_j}{2m}$$

# Asortatividad-Correlación de grado

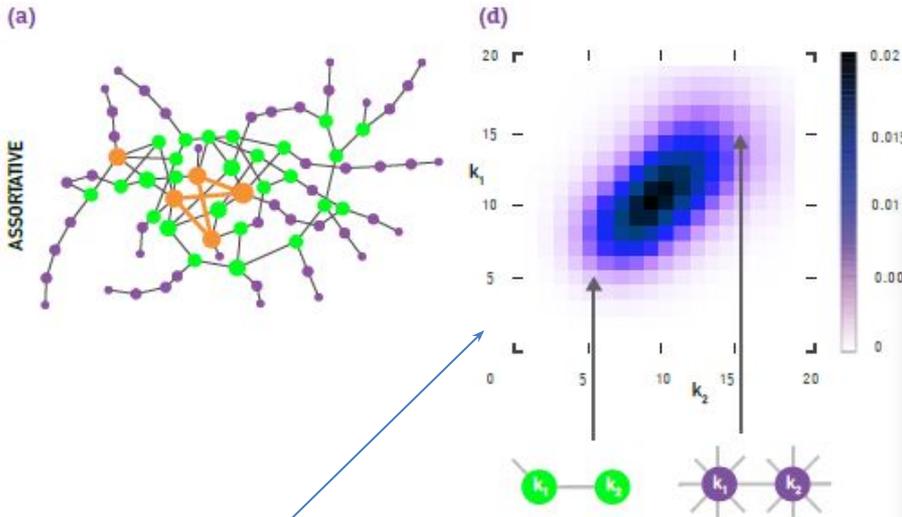
fracción de enlaces que conectan nodos de grado  $r$  con nodos de grado  $s$

$$e_{rs} = \frac{1}{2m} \sum_{ij} A_{ij} \delta(k_i, r) \delta(k_j, s)$$

fracción de enlaces adyacentes a un nodo de grado  $k$

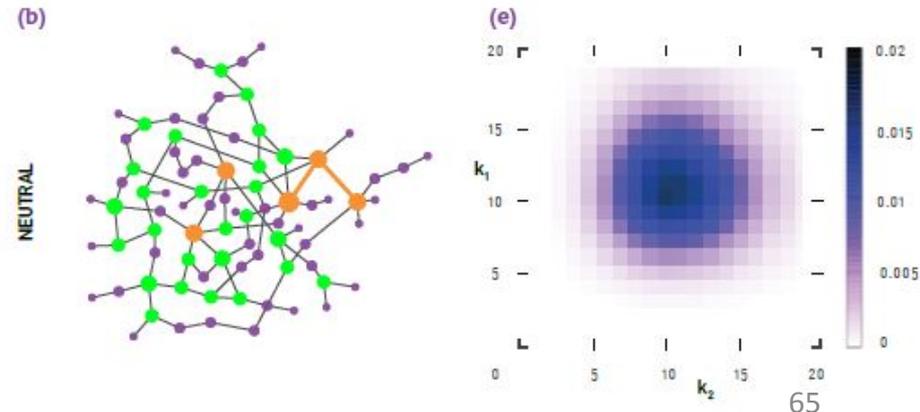
$$e_k = \frac{1}{2m} \sum_{ij} A_{ij} \delta(k_i, k) = \frac{1}{2m} \sum_i k_i \delta(k_i, k)$$

$$= \frac{k p(k) N}{2m} = \frac{k p(k)}{\langle k \rangle}$$



$e_{ij}$ : fracción de enlaces que conectan nodos de grado  $k_1=i$  y  $k_2=j$

$$e_{rs} \stackrel{?}{=} e_r e_s$$

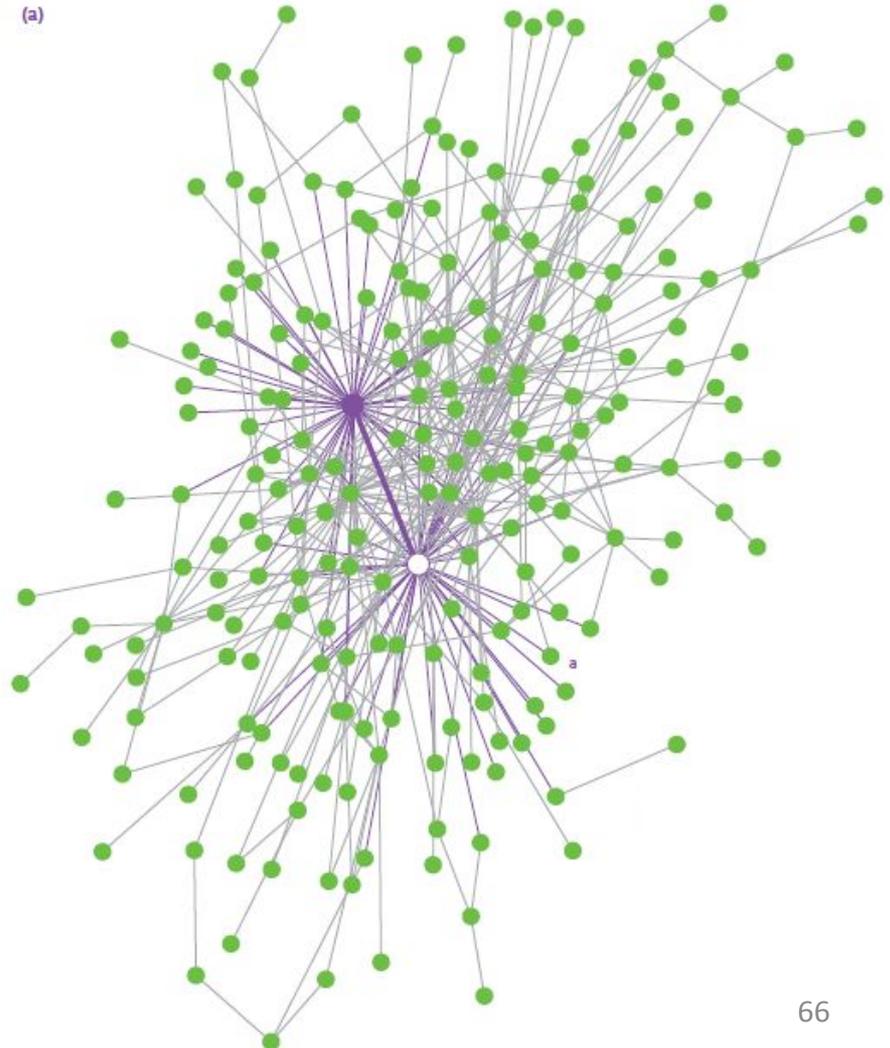


# Corte estructural

En redes **simples con distribuciones de grado de cola pesada** la presencia de hubs puede introducir un sesgo hacia la disortatividad...

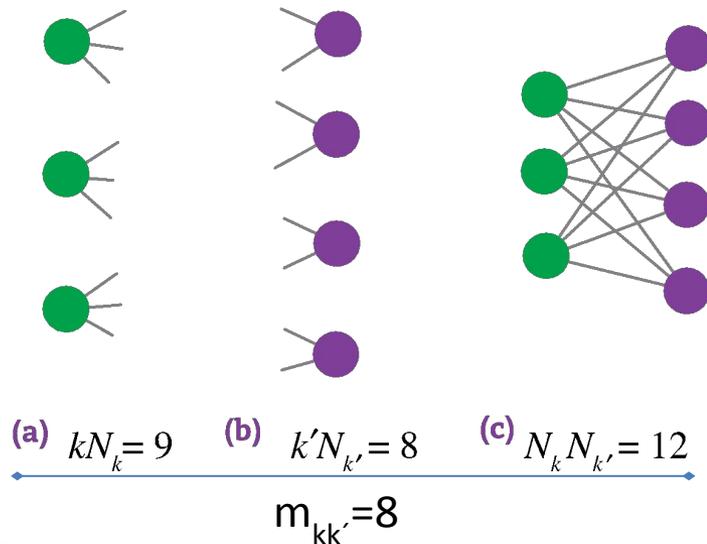
Cuántos enlaces cabría esperar en una **red neutral** entre dos *hubs* de la red?

Red scale-free  $N=300$ ,  $m=450$ ,  $\gamma=2.2$   
Disortativa:  $r=-0.19$



# Corte estructural

**Máximo** número de enlaces entre dos grupos de nodos  $m_{kk'}$ :



tres numeros que limitan

$$m_{kk'} = \min(kN_k, k'N_{k'}, N_k N_{k'})$$

nro de enlaces adyacentes  
a grupo verde

nro max de enlaces  
posible entre ambos  
grupos

nro de enlaces adyacentes  
a grupo violeta

Definimos:

$$r_{kk'} = \frac{N_{kk'}}{m_{kk'}}$$

Si  $r_{kk'} \leq 1$  no actua ningun corte  
estructural para lo que vemos en la red