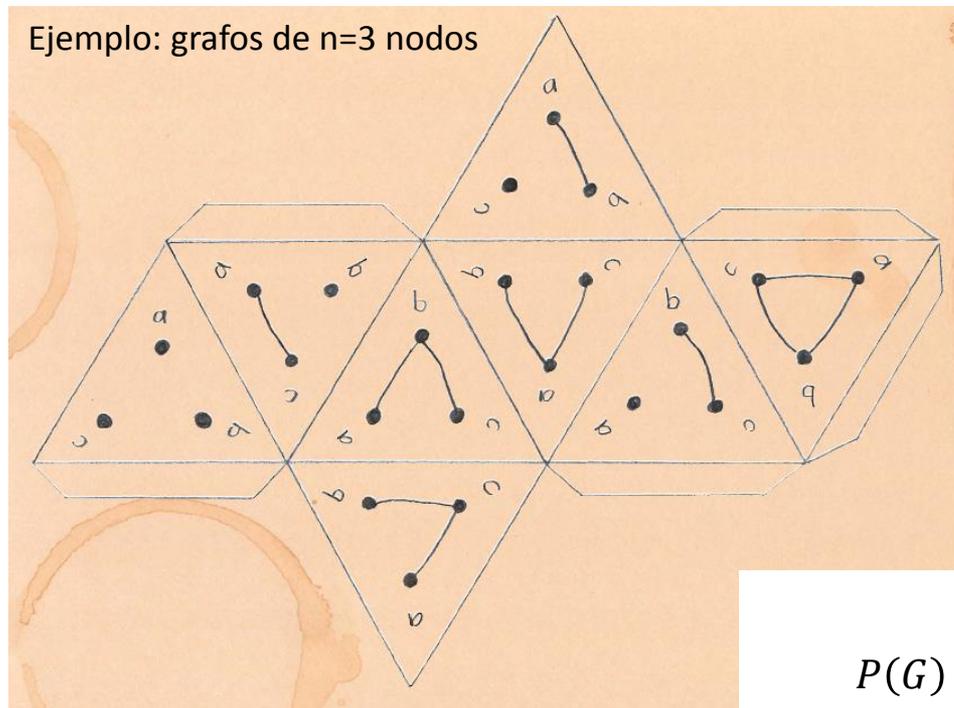


# Redes Aleatorias

# Teoría de redes aleatorias

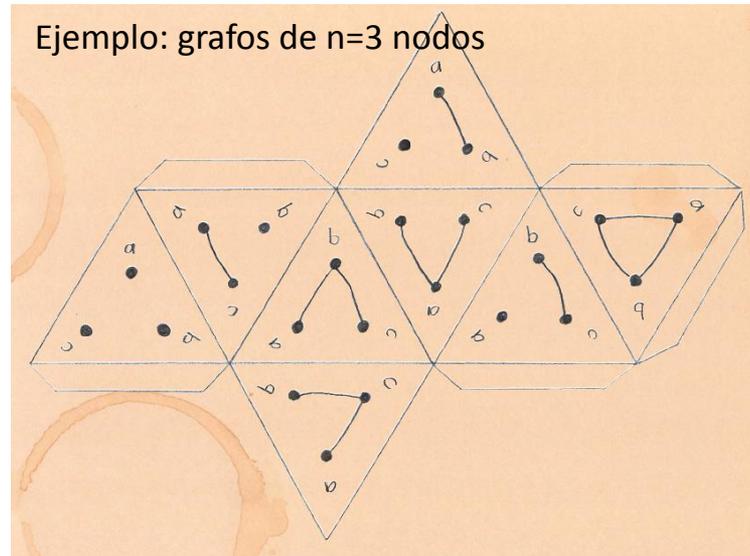
**Grafo aleatorio:** es un grafo obtenido muestreando un **conjunto dado** (*ensamble, familia*) de grafos



$$P(G) = \begin{cases} \frac{1}{8} & \text{si } G \text{ tiene 3 nodos} \\ 0 & \text{en otro caso} \end{cases}$$

# Teoría de redes aleatorias

**Grafo aleatorio:** es un grafo obtenido muestreando un **conjunto dado** (*ensamble, familia*) de grafos



Dado un ensamble de grafos, vamos a querer calcular valores medios de propiedades de interés

$$\langle f \rangle = \sum_G P(G) f(G)$$

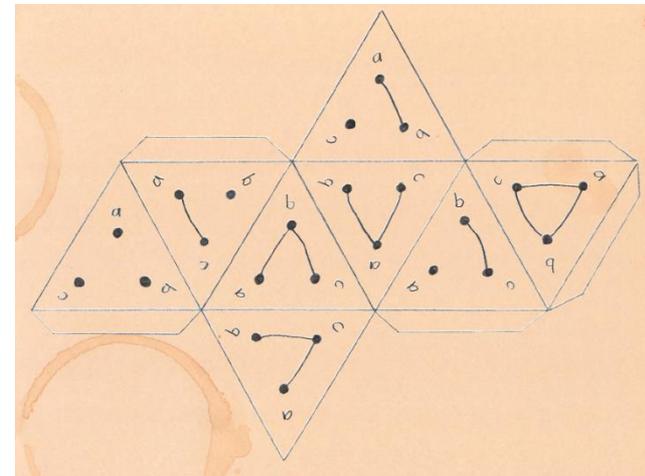
# Teoría de redes aleatorias

**Grafo aleatorio:** es un grafo obtenido muestreando un **conjunto dado** (*ensamble, familia*) de grafos

I. **G(N=n)** ensamble de grafos de **n** nodos

$$P(G) = \begin{cases} \frac{1}{\Omega_n} & \text{si } G \text{ tiene } n \text{ nodos} \\ 0 & \text{en otro caso} \end{cases}$$

$$\Omega_3 = 2^{\binom{3}{2}} = 2^{\frac{3!}{2!}} = 2^3$$



Cuántos hay en total? Cuanto vale  $\Omega_n$ ?

- Si hay  $n$  nodos, hay  $\binom{n}{2} = \frac{n(n-1)}{2}$  posibles pares (enlaces)
- Cada uno puede **estar** o **no-estar**
- Entonces  $\Omega_n = 2^{\binom{n}{2}} \sim e^{\frac{\ln 2}{2} n^2}$

# Teoría de redes aleatorias

**Grafo aleatorio:** es un grafo obtenido muestreando un **conjunto dado** (*ensamble, familia*) de grafos

- I. **G(N)** ensamble de grafos de **n** nodos
- II. **G(N=n, M=m)** conjunto formado por grafos de **n** vértices y **m** enlaces

$$P(G) = \begin{cases} \frac{1}{\Omega_{n,m}} & \text{si } G \text{ tiene } n \text{ nodos y } m \text{ vértices} \\ 0 & \text{si no} \end{cases}$$

$$\Omega_{n,m} = \binom{\binom{n}{2}}{m}$$

nro de maneras de elegir **2** elementos de **n** (i.e. nro maximo de posibles enlaces)

nro de maneras de elegir **m** enlaces del nro maximo de posibles enlaces

# Teoría de redes aleatorias

**Grafo aleatorio:** es un grafo obtenido muestreando un conjunto (*ensamble, familia*) dado de grafos

- I. **G(N)** ensamble de grafos de  $n$  nodos
- II. **G(N=n, M=m)** conjunto formado por grafos de  $n$  vértices y  $m$  enlaces

$$P(G) = \begin{cases} \frac{1}{\Omega_{n,m}} & \text{si } G \text{ tiene } N \text{ nodos y } M \text{ vértices} \\ 0 & \text{si no} \end{cases}$$

Dado un ensamble de grafos, es posible calcular valores medios de propiedades de interés

$$\langle f \rangle = \sum_G P(G) f(G) = \frac{1}{\Omega} \sum_{G \in G(N,M)} f(G)$$

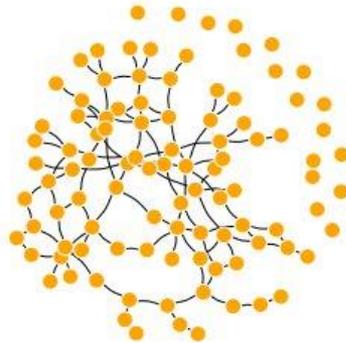
$$\langle k \rangle = \sum_G P(G) \langle k \rangle_G = \sum_G P(G) \frac{2m}{n} = \frac{2m}{n}$$

# Teoría de redes aleatorias

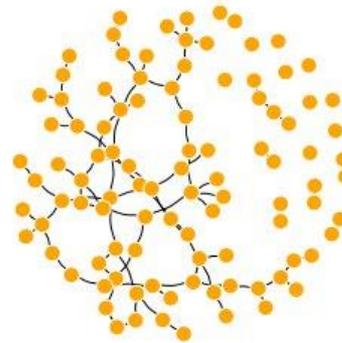
**Grafo aleatorio:** es un grafo obtenido muestreando un conjunto (*ensamble, familia*) dado de grafos

- I.  **$G(N)$**  ensamble de grafos de  $n$  nodos
- II.  **$G(n,m)$**  conjunto formado por grafos de  $n$  vértices y  $m$  enlaces
- III.  **$G(n,p)$**  conjunto formado por grafos de  $n$  vértices tal que la probabilidad de que exista un enlace entre cualquier par de nodos resulta  $p$

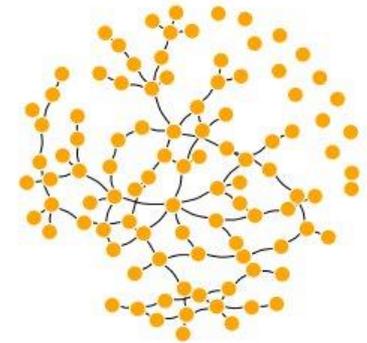
$n=100, p=0.02$



$m=108$



$m=90$



$m=96$

Entre 1950-1960  
Paul Erdos y Alfred  
Renyi desarrollaron  
la teoría de grafos  
aleatorios  
estudiando las  
propiedades de este  
tipo de ensambles

# Teoría de redes aleatorias

**Grafo aleatorio:** es un grafo obtenido muestreando un conjunto (*ensamble, familia*) dado de grafos

- I. **G(n)** ensamble de grafos de  $n$  nodos
- II. **G(n,m)** conjunto formado por grafos de  $n$  vértices y  $m$  enlaces
- III. **G(n,p)** conjunto formado por grafos de  $n$  vértices tal que la probabilidad de que exista un enlace entre cualquier par de nodos resulta  $p$

Prob de observar en este ensamble un **dado** grafo  $G$  con  $n$  vértices y  $m$  enlaces

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

Prob de tener  $m$  enlaces

Prob de no tener los enlaces restantes

# Teoría de redes aleatorias

**Grafo aleatorio:** es un grafo obtenido muestreando un conjunto (*ensamble, familia*) dado de grafos

- I. **G(n)** ensamble de grafos de  $n$  nodos
- II. **G(n,m)** conjunto formado por grafos de  $n$  vértices y  $m$  enlaces
- III. **G(n,p)** conjunto formado por grafos de  $n$  vértices tal que la probabilidad de que exista un enlace entre cualquier par de nodos resulta  $p$

Prob de observar en este ensamble un **dado** grafo  $G$  con  $n$  vértices y  $m$  enlaces

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

nro grafos con  $n$  vértices y  $m$  enlaces

Prob de observar en este ensamble un grafo con  $n$  vértices y  $m$  enlaces

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

La probabilidad de ver un grafo con exactamente  $m$  vértices en este ensamble, sigue una **distribucion Binomial** de probabilidades

# Repaso matematico...

$G(n,p)$

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

$$P(x) = \binom{N}{x} p^x (1-p)^{N-x}$$

Supongamos que en una prueba exito/fracaso la probabilidad de éxito es  $p$ .

La probabilidad de tener  $x$  exitos en  $N$  pruebas independientes :

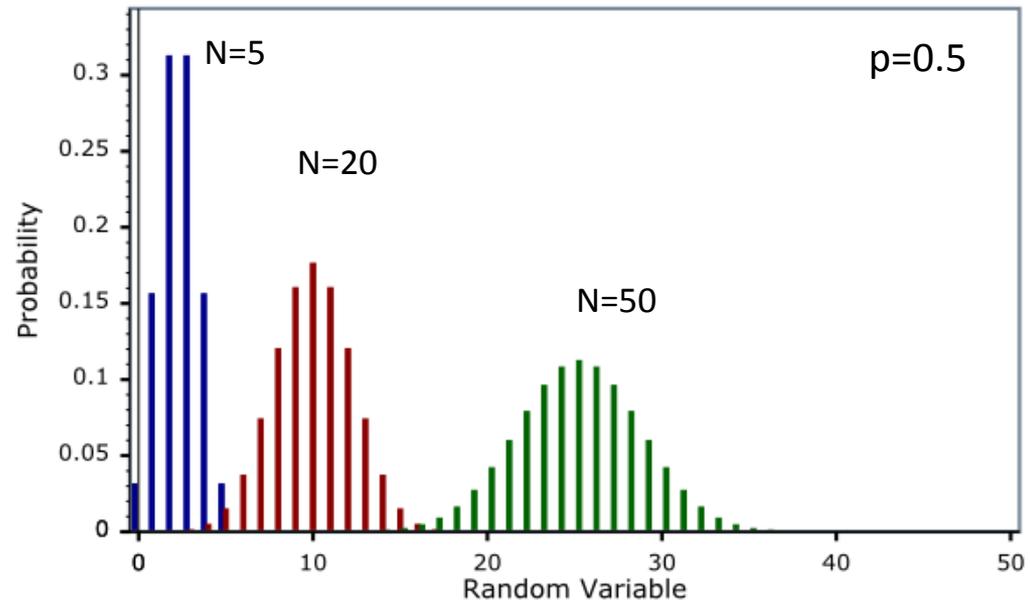
10001000011011100.....10

N pruebas (con  $x$  1's)

$$\langle x \rangle = \sum x P(x) = Np$$

$$\langle x^2 \rangle = p(1-p)N + p^2 N^2$$

$$\sigma = \sqrt{\langle x^2 \rangle - \langle x \rangle^2} = \sqrt{p(1-p)N}$$



# Propiedades de $G(N,p)$

$G(n,p)$

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

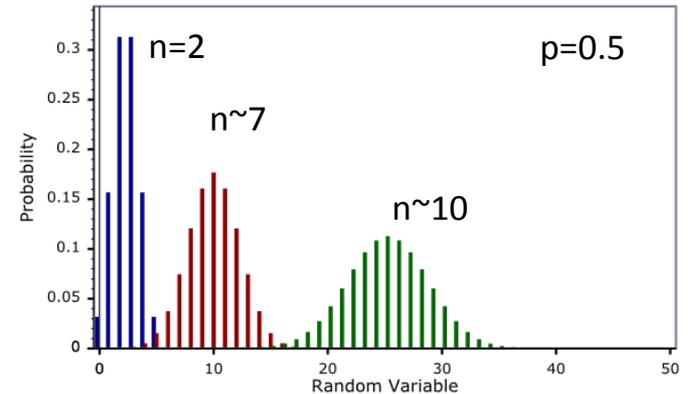
- $$\langle m \rangle = \sum_{m=0}^n \binom{n}{m} P(m) m = \binom{n}{2} p$$

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

$$\sigma = \sqrt{\langle m^2 \rangle - \langle m \rangle^2} = \sqrt{p(1-p) \binom{n}{2}}$$

$$\frac{\sigma}{\langle m \rangle} \sim \frac{1}{\sqrt{\binom{n}{2}}} \sim \frac{1}{n}$$

Aunque el ensamble admite grafos con cualquier número de enlaces la mayor parte de los grafos poseen  $m \sim \langle m \rangle$  si  $n \gg 1$



$$P(x) = \binom{N}{x} p^x (1-p)^{N-x}$$

$$\langle x \rangle = Np$$

$$\langle x^2 \rangle = p(1-p)N + p^2 N^2$$

$$\sigma = \sqrt{\langle x^2 \rangle - \langle x \rangle^2} = \sqrt{p(1-p)N}$$

# Propiedades de $G(N,p)$

$G(n,p)$

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

- $$\langle m \rangle = \sum_{m=0}^{\binom{n}{2}} P(m) m = \binom{n}{2} p$$

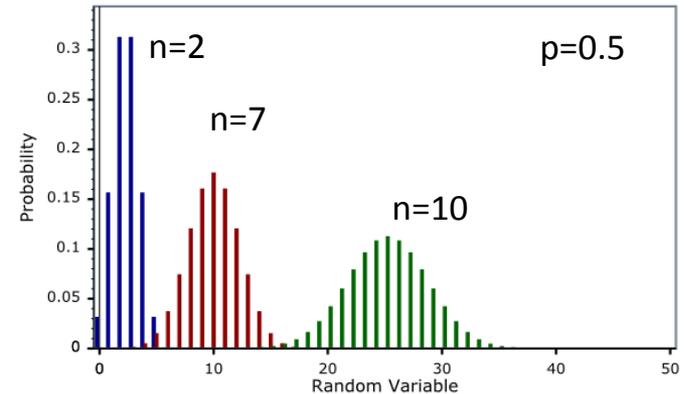
$$= \frac{n(n-1)}{2} p$$

- $$\langle k \rangle = \frac{2\langle m \rangle}{n} = (n-1)p$$

$$\sigma_k = \sqrt{(n-1)(1-p)p}$$

$$\frac{\sigma_k}{\langle k \rangle} \sim \frac{1}{\sqrt{n}}$$

La distribución de grado se angosta para redes gdes.



✓ Razonable: nro total de pares posibles por la probabilidad  $p$  de que se establezca un enlace

✓ Razonable: maximo nro de posibles enlaces de un nodonro por la probabilidad  $p$  de que se establezca un enlace

# Propiedades de $G(N,p)$

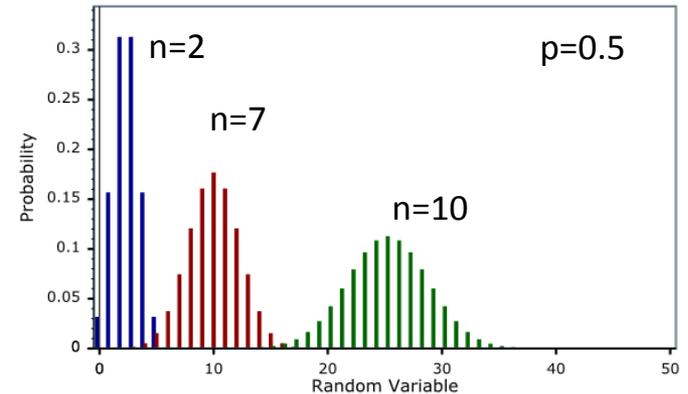
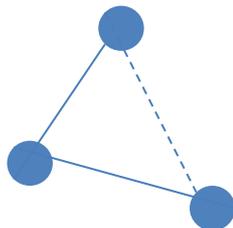
$G(n,p)$

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

- $$\langle m \rangle = \sum_{m=0}^{\binom{n}{2}} P(m) m = \binom{n}{2} p$$

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

- $$\langle k \rangle = \frac{2\langle m \rangle}{n} = (n-1)p$$



- Coeficiente de Clustering:**

$$C_{\Delta} = 3 \frac{\#triangulos}{\#tripletes}$$

$$= 3 \frac{\frac{1}{6} N (N-1)(N-2)p^3}{\frac{1}{2} N (N-1)(N-2)p^2} = p$$

$$C_i = \frac{2 \#vec\ enlazados}{k_i(k_i-1)}$$

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

# Propiedades de $G(N,p)$

$G(n,p)$

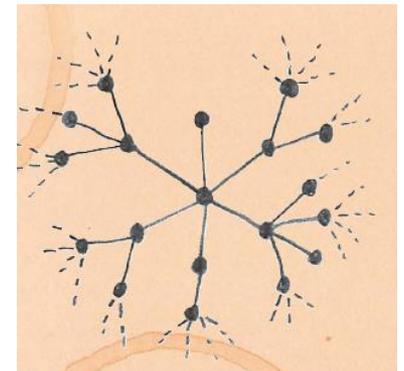
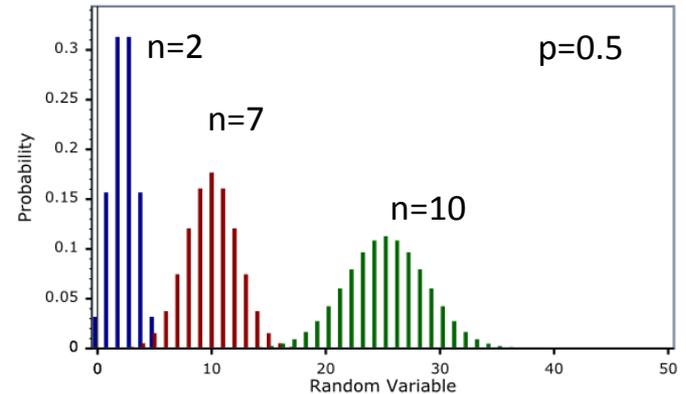
$$P(M = m) = \binom{\binom{n}{2}}{m} p^m (1-p)^{\binom{n}{2}-m}$$

- $\langle m \rangle = \sum_{m=0}^{\binom{n}{2}} P(m) m = \binom{n}{2} p$   
 $= \frac{n(n-1)}{2} p$

- $\langle k \rangle = \frac{2\langle m \rangle}{n} = (n-1)p$

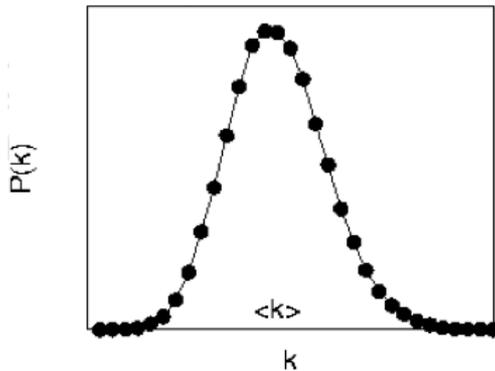
- $C = p$

- Para redes muy gdes,  $n \rightarrow \infty$ , en las que  $\langle k \rangle$  se mantiene finito,  $p \rightarrow 0$
- $C \rightarrow 0$
- Redes presentan pocos *loops*
- Localmente se parecen a arboles



# Distribución de grado

Quiero estimar  $P_k$  probabilidad de que un nodo tomado al azar de una red  $G(N,p)$  tenga  $k$  vecinos



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

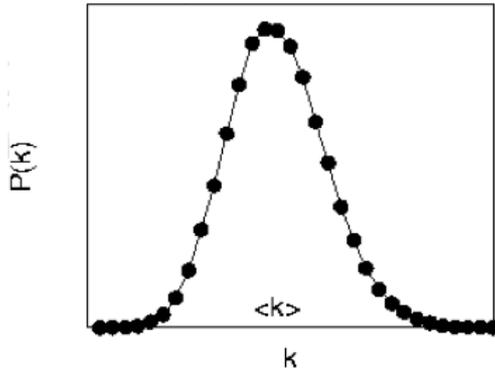


Nro posible de maneras de elegir  $k$  vecinos de los  $n-1$  disponibles

- $P_k$  sigue una distribución binomial
- Si  $n \rightarrow \infty$  y  $p$  se mantiene fijo, tiende a una distribución Gaussiana
- Si  $n \rightarrow \infty$  y  $\langle k \rangle = (n-1)p$  se mantiene finito, entonces  $p \rightarrow 0$ . Vamos a analizar  $P_k$  en éste limite

# Distribución de grado

Quiero estimar  $P_k$  probabilidad de que un nodo tomado al azar de una red  $G(N,p)$  tenga  $k$  vecinos



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

Analizo los diferentes factores cuando

$$\begin{cases} n \rightarrow \infty, p \rightarrow 0 \text{ y} \\ c \equiv \langle k \rangle = (n-1)p \text{ se mantiene finito} \end{cases}$$

$$\underline{(1-p)^{n-1-k}}$$

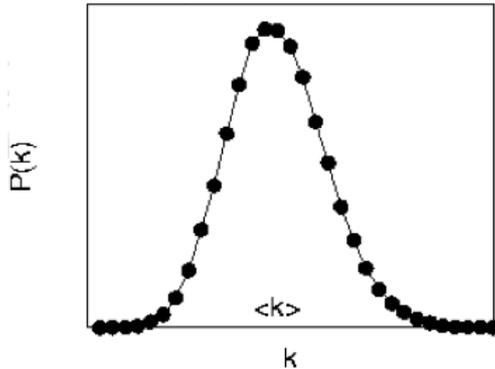
$$\log(1-p)^{n-1-k} = (n-1-k) \log(1-p)$$

$$\log(1-p) \sim -p \sim -(n-1-k)p = -(n-1-k) \frac{c}{n-1} \sim -c$$

$$(1-p)^{n-1-k} \sim e^{-c}$$

# Distribución de grado

Quiero estimar  $P_k$  probabilidad de que un nodo tomado al azar de una red  $G(N,p)$  tenga  $k$  vecinos



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

Analizo los diferentes factores cuando

$$\begin{cases} n \rightarrow \infty, p \rightarrow 0 \text{ y} \\ c \equiv \langle k \rangle = (n-1)p \text{ se mantiene finito} \end{cases}$$

$$(1-p)^{n-1-k} \sim e^{-c}$$

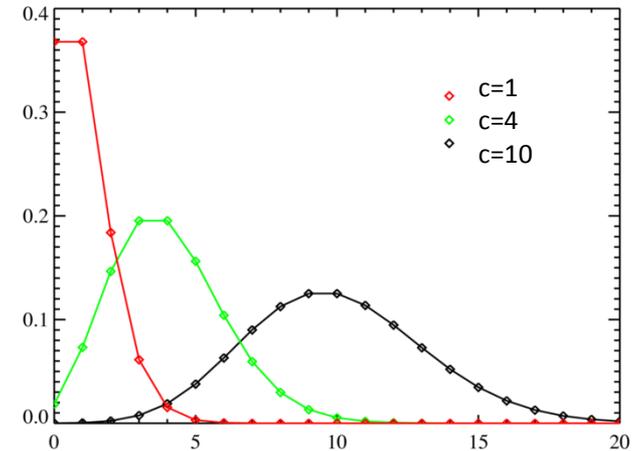
$$\binom{n-1}{k} = \frac{(n-1)!}{(n-1-k)! k!} = \frac{(n-1)(n-2) \dots (n-1-k+1)}{k!} \sim \frac{(n-1)^k}{k!}$$

$$P_k \sim \frac{(n-1)^k}{k!} p^k e^{-c} = \frac{c^k}{k!} e^{-c}$$

Distribución de Poisson

# Repaso matemático

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



$$\langle k \rangle = \sum_{k=0}^{\infty} k P_k = e^{-c} \sum_{k=0}^{\infty} k \frac{c^k}{k!} = c e^{-c} \sum_{k=1}^{\infty} \frac{c^{k-1}}{(k-1)!} = c e^{-c} \sum_{j=0}^{\infty} \frac{c^j}{j!} = c$$

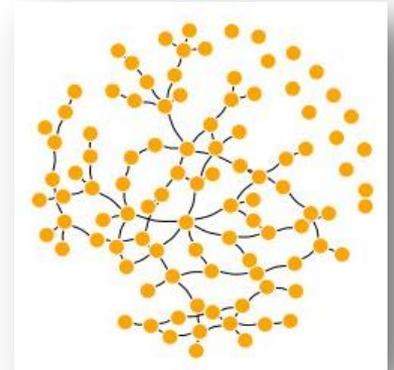
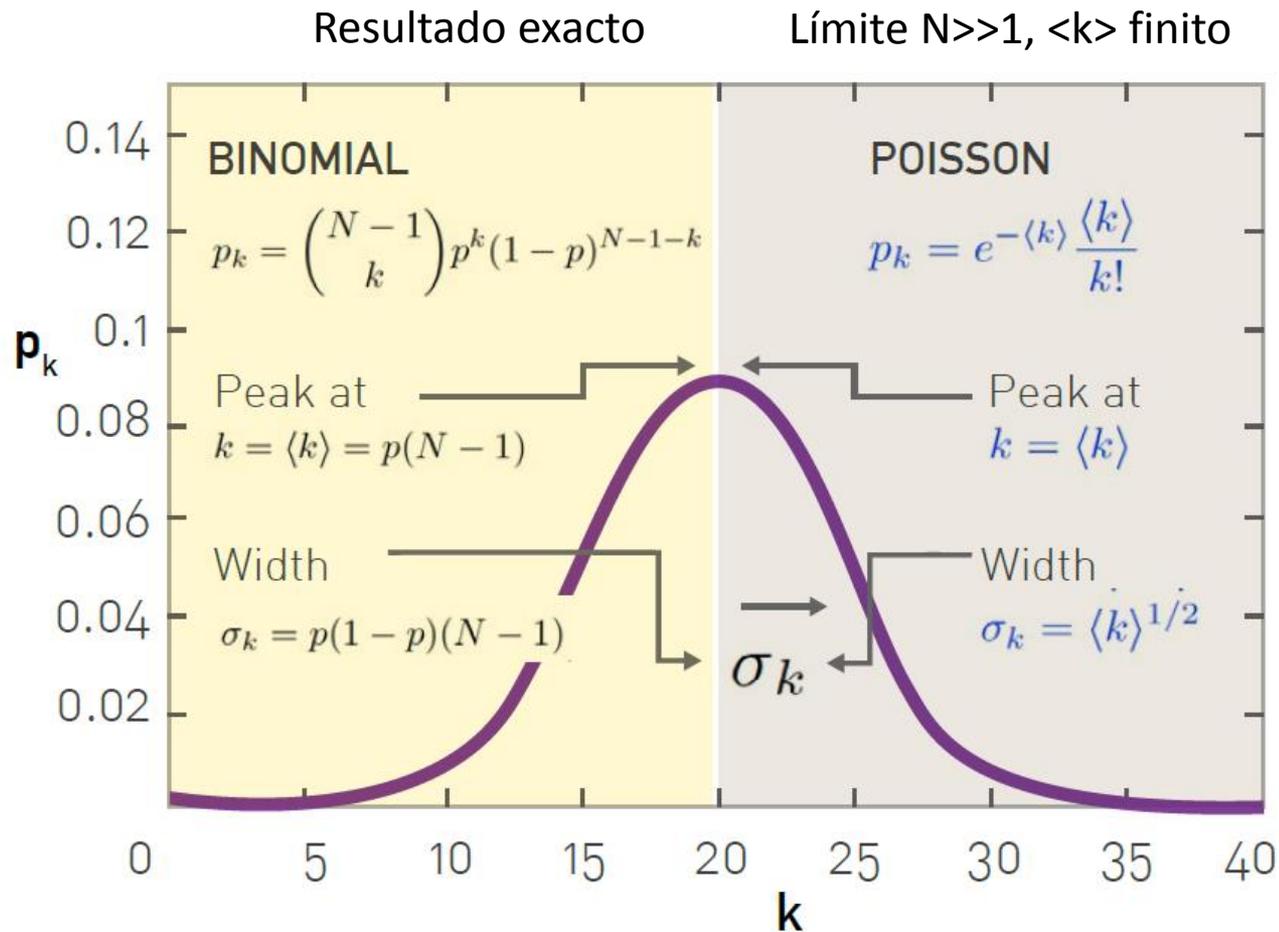
$$\langle k^2 \rangle = \sum_{k=0}^{\infty} k^2 P_k = \langle k \rangle^2 + \langle k \rangle$$

$$\sigma^2 = \langle k^2 \rangle - \langle k \rangle^2 = \langle k \rangle$$

La varianza es igual a la media (!)

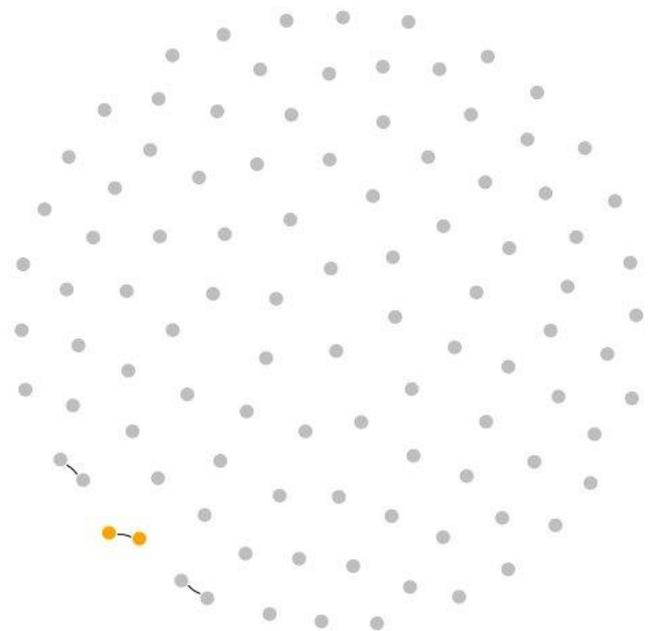
$$\frac{\sigma}{\langle k \rangle} = \frac{1}{\sqrt{\langle k \rangle}}$$

# G(N,p): Distribución de grado



# Conectividad de $G(N,p)$

$\langle k \rangle: 0.1 - M: 3 \text{ CG}: 2$

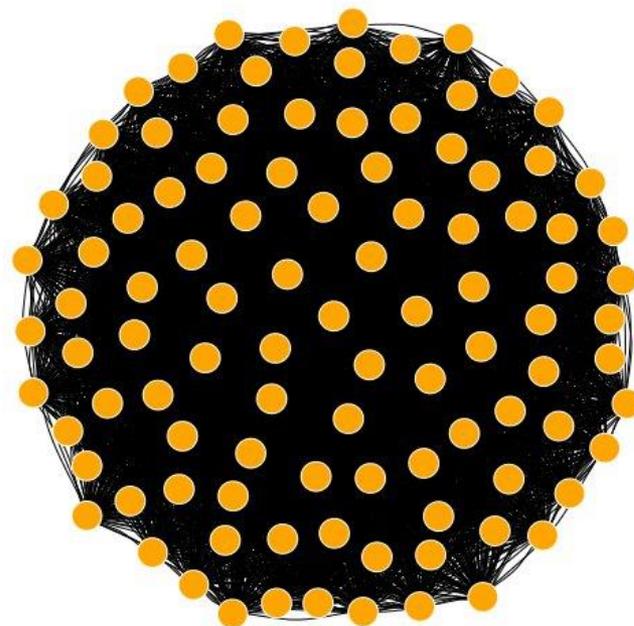


$p \sim 0$

Como se produce esta transición?



$\langle k \rangle: 98 - M: 4898 \text{ CG}: 100$



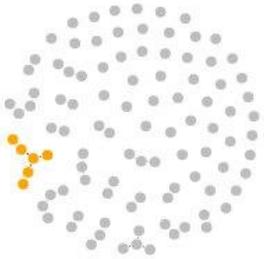
$p \sim 1$

- Grafo desconexo
- Tamaño componente gigante es una cantidad **intensiva**:  $o(1)$

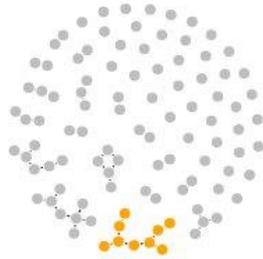
- Grafo completamente conexo
- Tamaño componente gigante es una cantidad **extensiva**:  $o(N)$

# La transición

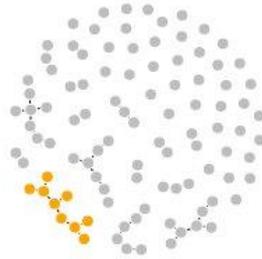
$\langle k \rangle: 0.4$  - M:33 CG:6



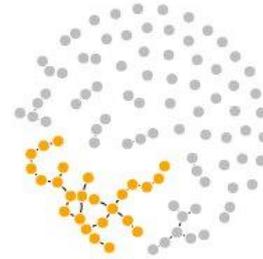
$\langle k \rangle: 0.8$  - M:43 CG:9



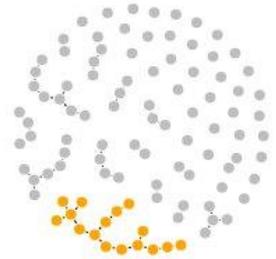
$\langle k \rangle: 0.9$  - M:45 CG:9



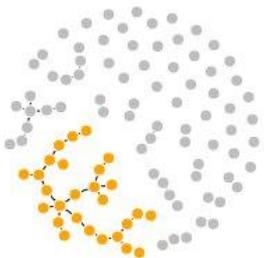
$\langle k \rangle: 0.96$  - M:51 CG:25



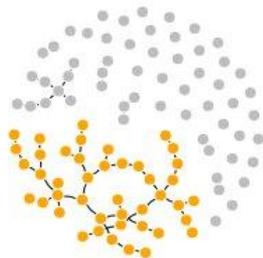
$\langle k \rangle: 0.98$  - M:50 CG:16



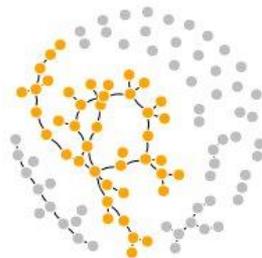
$\langle k \rangle: 1$  - M:52 CG:27



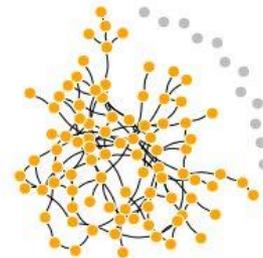
$\langle k \rangle: 1.1$  - M:58 CG:43



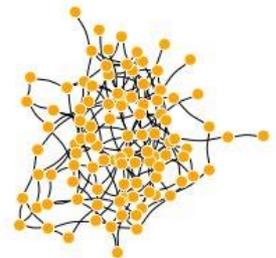
$\langle k \rangle: 1.2$  - M:74 CG:43



$\langle k \rangle: 2$  - M:127 CG:88



$\langle k \rangle: 4$  - M:185 CG:100



- Como es el cambio de régimen  $|CG|$  intensivo/extensivo?
- Qué sucede con lo que no es componente gigante...?

# Tamaño de la Componente Gigante

Sea  $u$  la fracción de la red que **no** pertenece a la CG ( $u=1$  si no hay componente gigante)

Un nodo- $i$  **no** pertenece a la CG si **no** está conectado a ningún vértice que pertenezca.

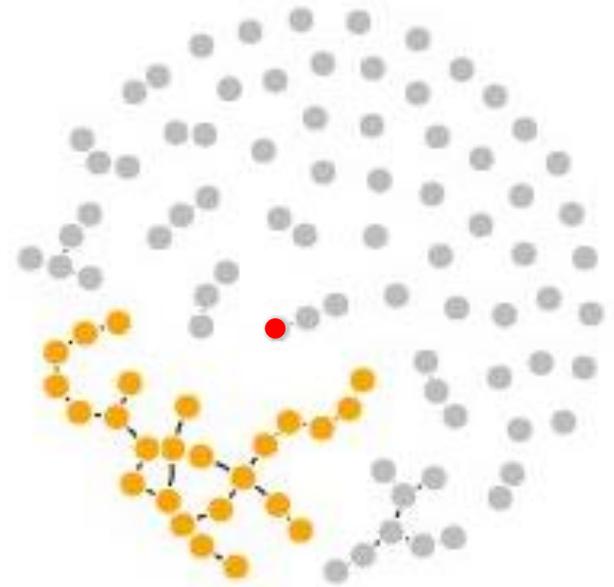
- Es decir, para todo nodo- $j$ :
  - a)  $i$  no está conectado a  $j$ . Esto ocurre con probabilidad :  $(1-p)$
  - b)  $i$  está conectado a  $j$ , pero  $j$  no pertenece a la CG. Esto ocurre con probabilidad:  $p*u$

Por (a) y (b), la probabilidad de que nodo- $i$  **no** pertenezca al CG, vía el nodo- $j$  resulta :

$$(1-p) + p u$$

Entonces, la probabilidad de que nodo- $i$  **no** pertenezca al CG resulta :  $(1-p + p u)^{n-1}$

$$\log(1-x) \sim -x$$



$$\begin{aligned} u &= [1 - p(1 - u)]^{n-1} \\ &= \left[1 - \frac{c}{n-1}(1 - u)\right]^{n-1} \end{aligned}$$

$$\begin{aligned} \log u &= (n-1) \log \left[1 - \frac{c}{n-1}(1 - u)\right] \\ &\sim -c(1 - u) \end{aligned}$$

# Tamaño de la Componente Gigante

Sea  $u$  la fracción de la red que **no** pertenece a la CG ( $u=1$  si no hay componente gigante)

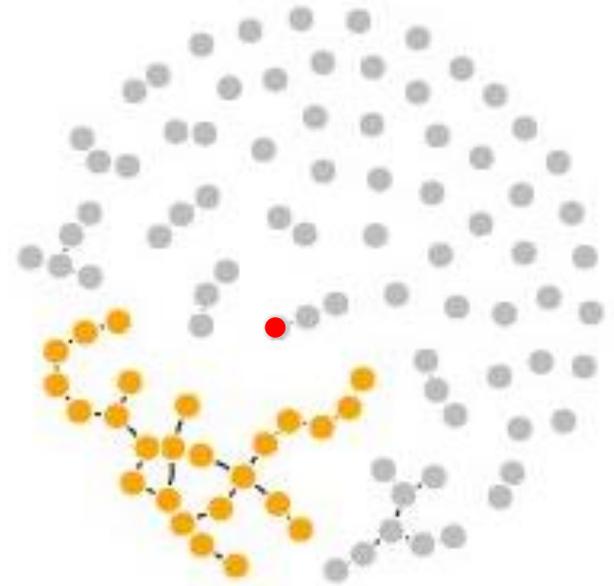
Un nodo- $i$  **no** pertenece a la CG si **no** está conectado a ningún vértice que pertenezca.

- Es decir, para todo nodo- $j$ :
  - a)  $i$  no está conectado a  $j$ . Esto ocurre con probabilidad :  $(1-p)$
  - b)  $i$  está conectado a  $j$ , pero  $j$  no pertenece a la CG. Esto ocurre con probabilidad:  $p*u$

Por (a) y (b), la probabilidad de que nodo- $i$  **no** pertenezca al CG, vía el nodo- $j$  resulta :

$$(1-p) + p u$$

Entonces, la probabilidad de que nodo- $i$  **no** pertenezca al CG resulta :  $(1 - p + p u)^{n-1}$



$$\begin{aligned} u &= [1 - p(1 - u)]^{n-1} \\ &= \left[1 - \frac{c}{n-1} (1 - u)\right]^{n-1} \end{aligned}$$

$$\underline{u \sim e^{-c(1-u)}}$$

# Tamaño de la Componente Gigante

$$c = \langle k \rangle = p(n - 1)$$

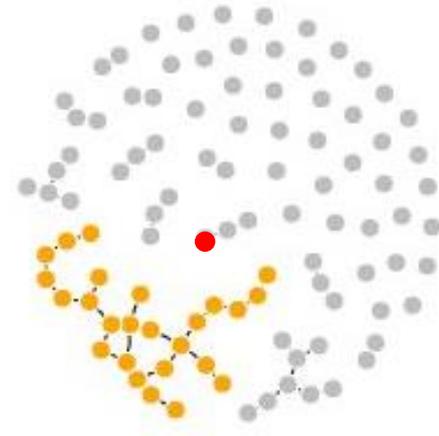
$$u \sim e^{-c(1-u)}$$

$S = 1 - u$  , fracción de nodos en la CG

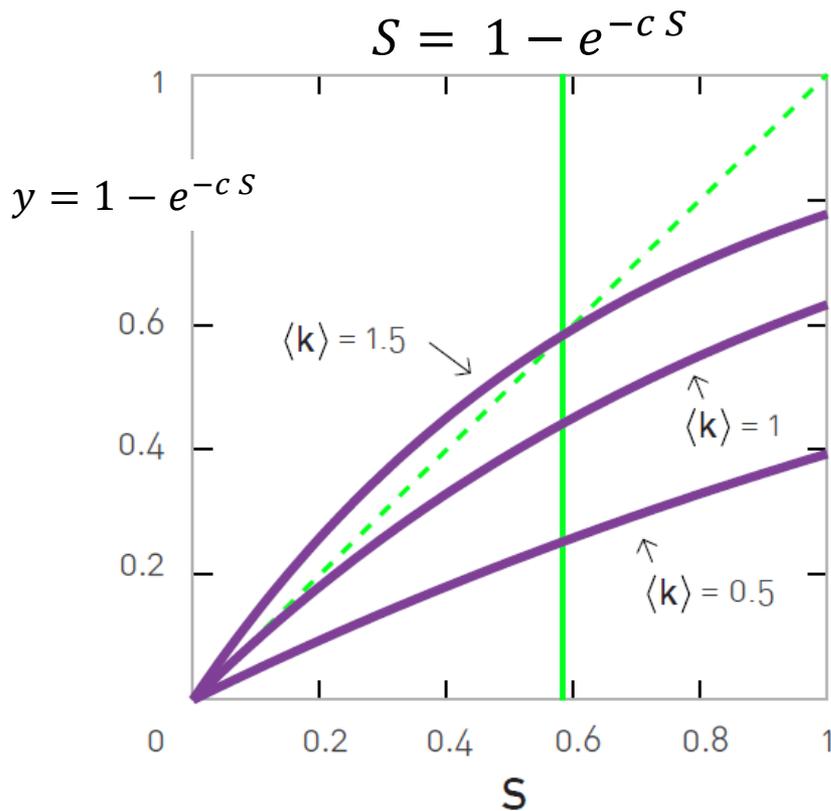
$$1 - S \sim e^{-cS}$$

$$\underline{S = 1 - e^{-cS}}$$

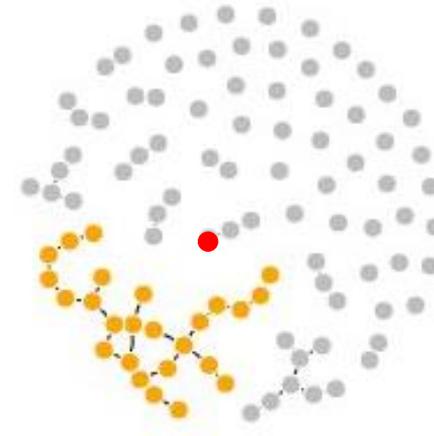
(ER 1959)



# Tamaño de la Componente Gigante



$$c = \langle k \rangle = p(n - 1)$$



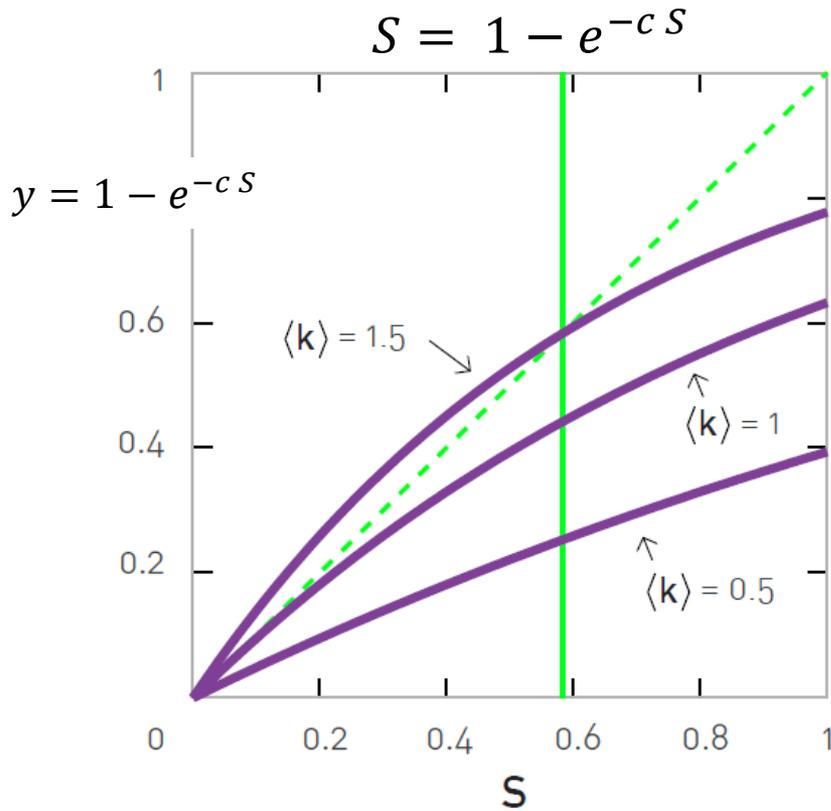
- $y(S)$  es monotonamente creciente y tiene curvatura negativa
- Hay un **cambio cualitativo** en las soluciones  $S^* = y(S^*)$  para los casos en que la **pendiente en el origen** toma valor **uno**

$$\left. \frac{dy}{dS} \right|_{S=0} = \left. c e^{-cS} \right|_{S=0} = c$$

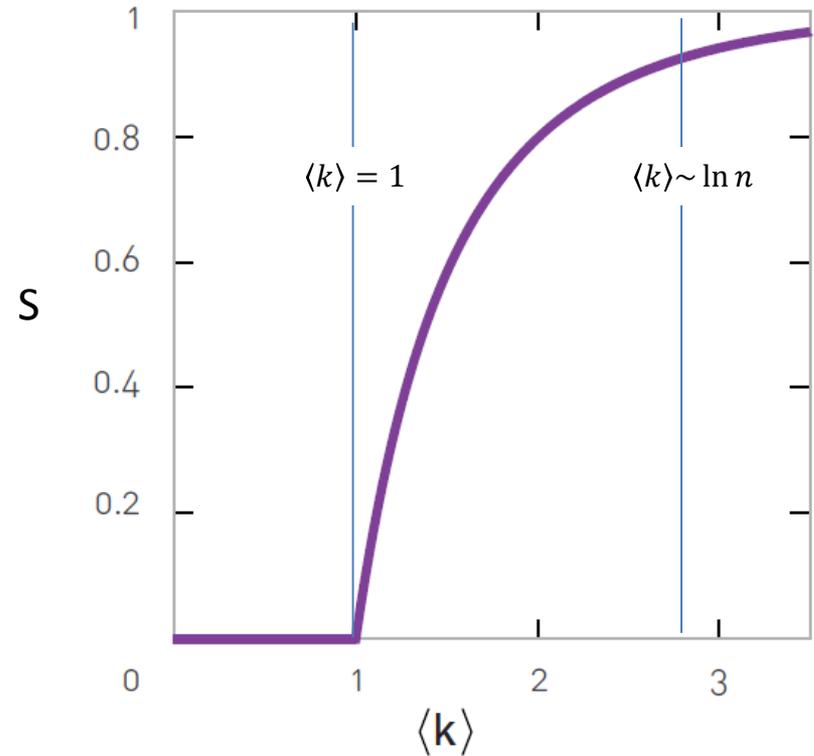
$$c < 1 \quad S^* = 0$$

$$c \geq 1 \quad S^* > 0 \text{ además de la trivial}$$

# Tamaño de la Componente Gigante



Es posible distinguir 4 regimenes



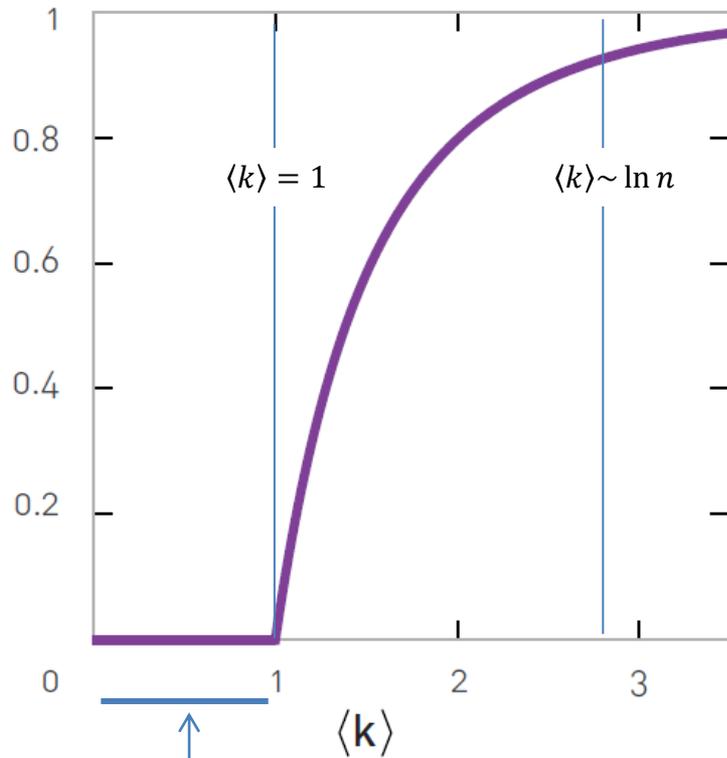
- $y(S)$  es monotonamente creciente y tiene curvatura negativa
- Hay un **cambio cualitativo** en las soluciones  $S^* = y(S^*)$  para los casos en que la **pendiente en el origen** toma valor **uno**

$$\left. \frac{dy}{dS} \right|_{S=0} = \left. c e^{-cS} \right|_{S=0} = c$$

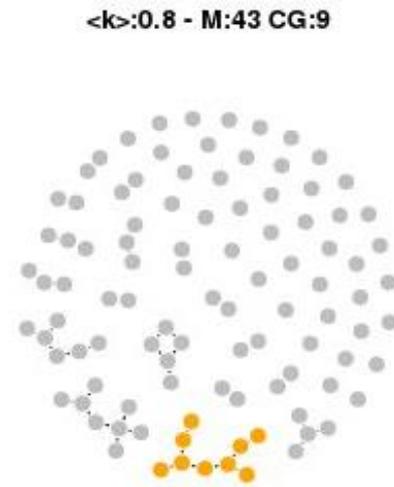
$$c < 1 \quad S^* = 0$$

$$c \geq 1 \quad S^* > 0$$

# Conectividad $G(N,p)$ . Régimen sub-crítico



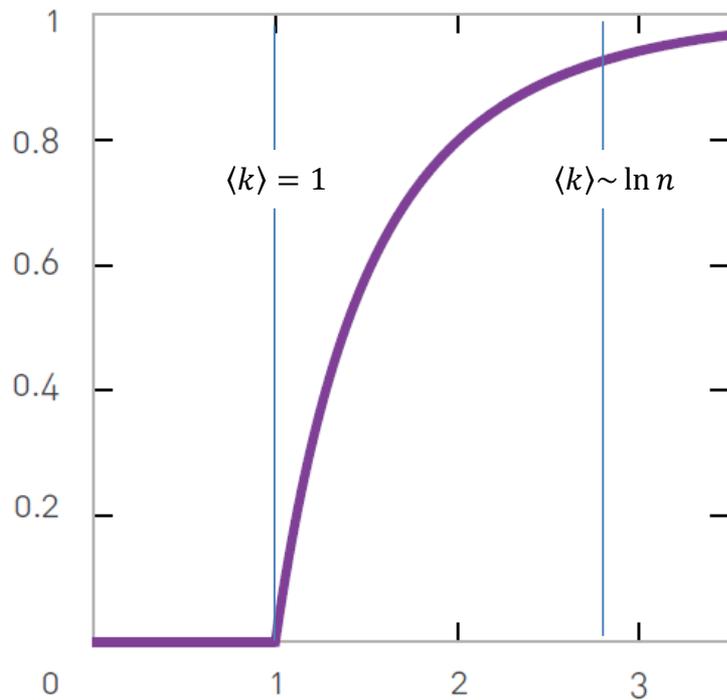
régimen  
subcrítico



- no hay componente gigante
- componente más gde de tamaño  $\sim \ln n$
- componentes aislados con distribución de tamaños exponencial

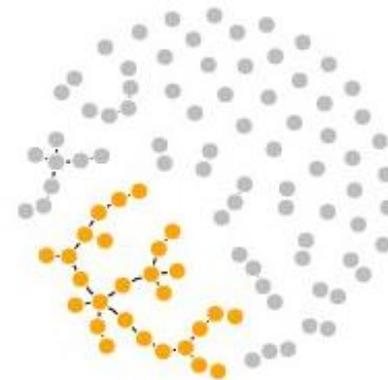
$$p(\text{size}) \sim s^{-3/2} e^{-\frac{(c-1)s + (s-1) \ln c}{c}}$$

# Conectividad $G(N,p)$ . Régimen crítico



↑  
régimen crítico

$\langle k \rangle : 1$  - M:52 CG:27



- componente gigante  $n_{CG} \sim n^{2/3}$
- componente gigante incipiente  $S_{CG} \sim n^{-1/3}$

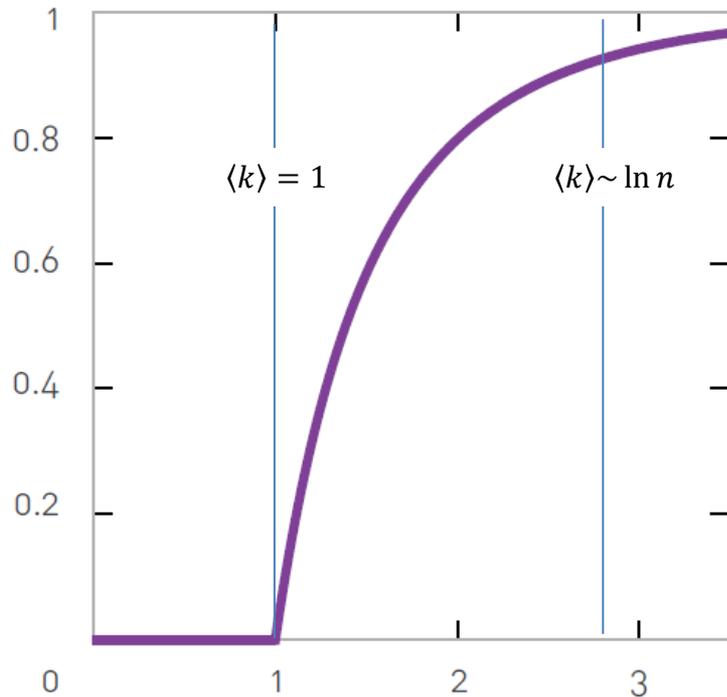
igual...comparar con regimen subcritico:

$$n=1000: \ln n \sim 7 \text{ vs } n^{2/3} \sim 95$$

$$n=10^6 : \ln n \sim 14 \text{ vs } n^{2/3} \sim 10000$$

- componentes no-gigantes sin loops
- distribución de tamaños de exponentes sigue ley de potencias  $p(\text{size}) \sim s^{-3/2}$

# Conectividad $G(N,p)$ . Régimen supercrítico

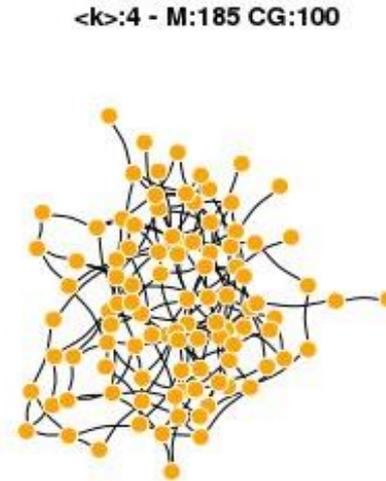
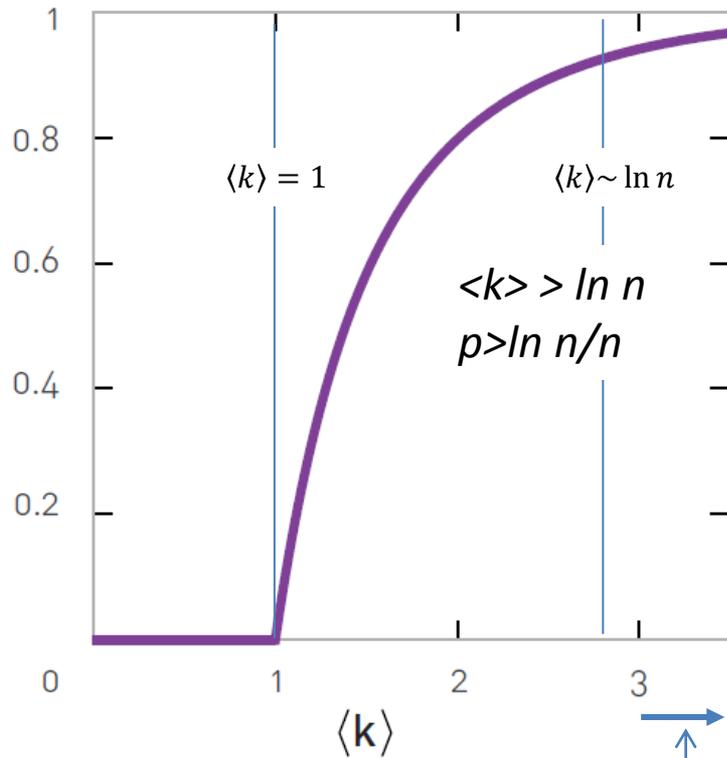


- componente gigante único  $n_{CG} \sim (p - p_c)n$
- componente gigante tiene loops
- otros componentes aislados con distribución de tamaños exponencial

$$p(\text{size}) \sim s^{-3/2} e^{-\frac{(c-1)s + (s-1) \ln c}{c}}$$

régimen  
supercrítico

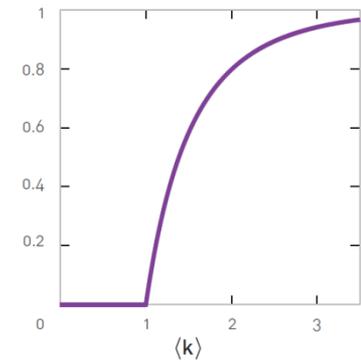
# Conectividad $G(N,p)$ . Régimen conectado



régimen  
conectado

- la probabilidad de que un nodo **no** pertenezca a la CG se anula exponencialmente
- una única componente en el grafo

# Las demás se quedan piolas



Veamos que cuando existe, sólo existe una **única** componente gigante

- Supongamos que existen 2 CG de tamaños:  $S_1 \cdot n$  y  $S_2 \cdot n$
- Nro de posibles pares de vértices  $(i, j)$  con  $i$  en  $CG_1$  y  $j$  en  $CG_2$ :  $S_1 S_2 n^2$
- Cada uno se conecta con prob:  $p$   
**no** se conecta con prob:  $(1-p)$
- Para que haya 2 CGs la probabilidad,  $q$ , de que algunos de esos pares **no** se conecte debe ser nula

$$q = (1 - p)^{S_1 S_2 n^2} = \left(1 - \frac{c}{n-1}\right)^{S_1 S_2 n^2} \sim e^{-ncS_1 S_2}$$

$$\begin{aligned} \log q &= S_1 S_2 \lim_{n \rightarrow \infty} n^2 \left(1 - \frac{c}{n-1}\right) \\ &= cS_1 S_2 \left[-n + \frac{1}{2}(c-1)\right] \sim -ncS_1 S_2 \end{aligned}$$

La probabilidad de que **no** exista vínculo entre  $S_1$  y  $S_2$  decae exponencialmente

# Redes Aleatorias

Familias de redes:  $G(N,M)$  ,  $G(N,p)$

grado medio  $\langle k \rangle = p(n - 1)$

nro enlaces  $\langle m \rangle = p \frac{n(n - 1)}{2}$

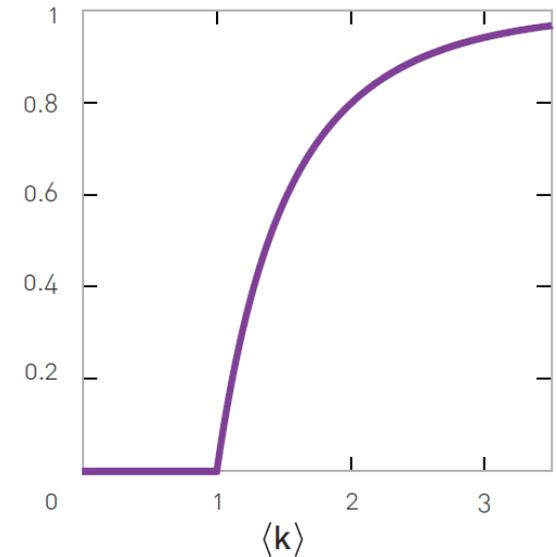
distribucion de grado

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

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

coef. clustering  $C = \frac{c}{n}$

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



$$n_{CG} = \begin{cases} \ln n & c < 1 \\ n^{2/3} & 1 \leq c < \ln n \\ (p-p_c)n & c > \ln n \end{cases}$$

# Redes Aleatorias

Familias de redes:  $G(N,M)$  ,  $G(N,p)$

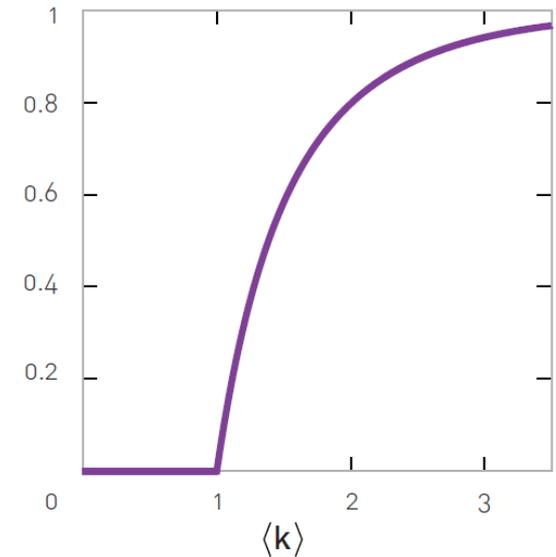
grado medio  $\langle k \rangle = p(n - 1)$

nro enlaces  $\langle m \rangle = p \frac{n(n - 1)}{2}$

distribucion de grado  $\left\{ \begin{array}{l} P_k = \binom{n-1}{k} p^k (1-p)^{n-1-k} \\ P_k = \frac{c^k}{k!} e^{-c} \end{array} \right.$

coef. clustering  $C = \frac{c}{n}$

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



$$n_{CG} = \begin{cases} \ln n & c < 1 \\ n^{2/3} & 1 \leq c < \ln n \\ (p-p_c)n & c > \ln n \end{cases}$$

