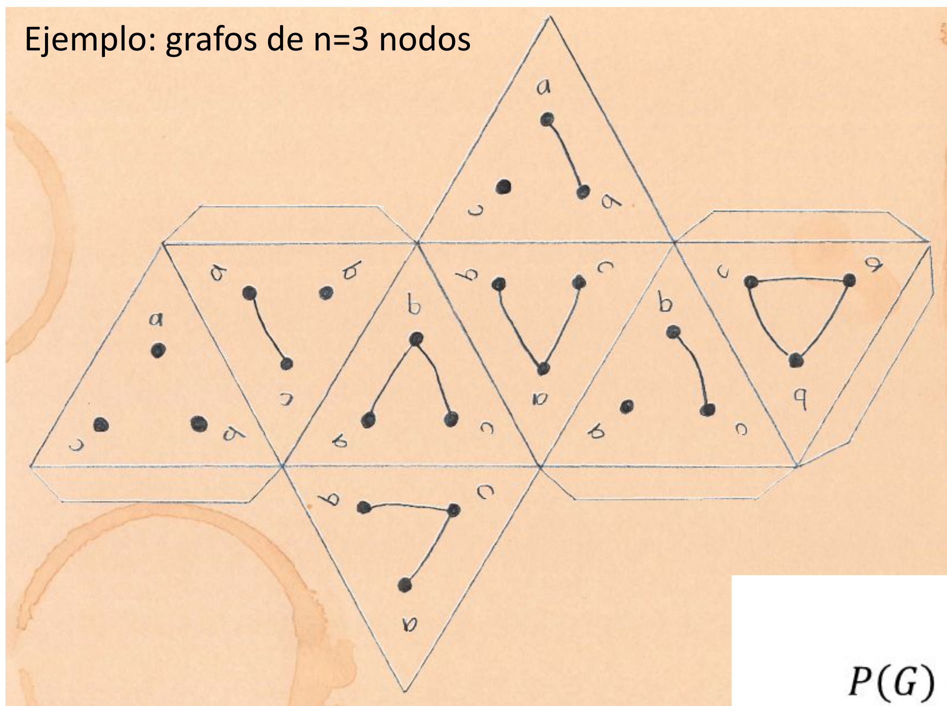
Redes Aleatorias

v2.

Teoría de redes aleatorias

Grafos aleatorios: grafos obtenidos muestreando, de acuerdo a una dada **distribución de probabilidad**, un **conjunto dado** (*ensamble, familia*) de grafos

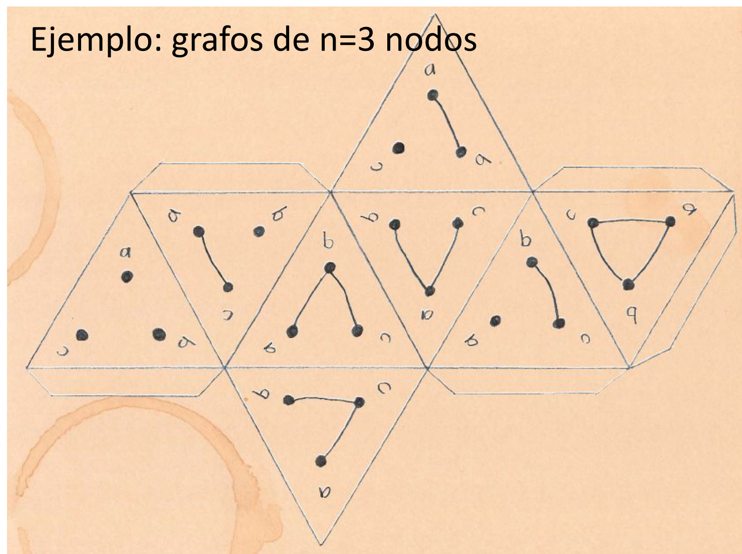


$$P(G) = \begin{cases} \frac{1}{8} & \text{si G tiene 3 nodos} \\ 0 & \text{en otro caso} \end{cases}$$

Adaptado de
Peter Dodds
Applications of Random Networks
Complex Networks
CSYS/MATH 303, Spring, 2011
www.uvm.edu/pdodds/teaching/courses/coconuts

Teoría de redes aleatorias

Grafos aleatorios: grafos obtenidos muestreando, de acuerdo a una dada **distribución de probabilidad**, un **conjunto dado** (*ensamble, familia*) de grafos



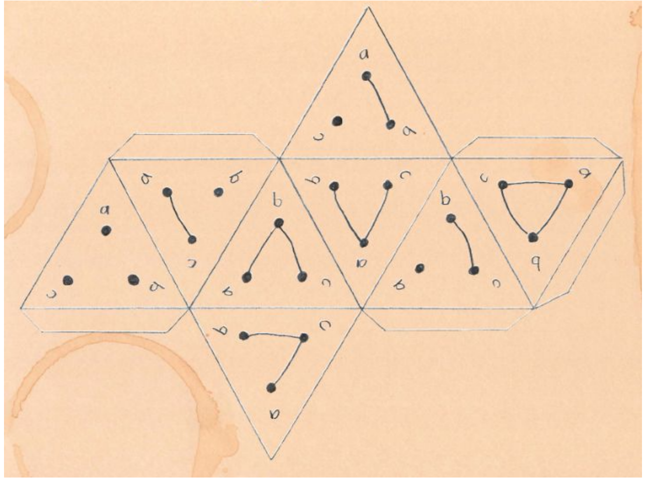
Dado un ensamble de grafos, vamos a querer calcular valores medios de propiedades de interés

$$\langle f \rangle = \sum_{G} P(G)\, f(G)$$

Teoría de redes aleatorias

Grafo aleatorio: es un grafo obtenido muestreando un *ensamble* de grafos

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



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

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

Cuantos hay en total? Cuanto vale Ω_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 *ensamble* 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

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

$$\Omega_{n,m} = \begin{pmatrix} \binom{n}{2} \\ m \end{pmatrix}$$

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 *ensamble* 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

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

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

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

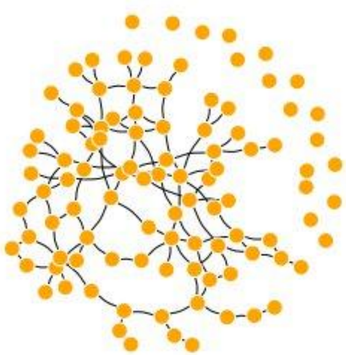
$$\langle k \rangle = \sum_G P(G)\, \langle k \rangle_G = \sum_G P(G)\, \frac{2m}{n} = \frac{2m}{n}$$

Teoría de redes aleatorias

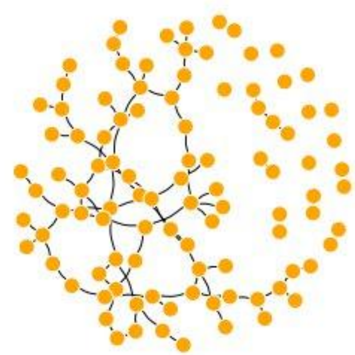
Grafo aleatorio: es un grafo obtenido muestreando un *ensamble* de grafos

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**

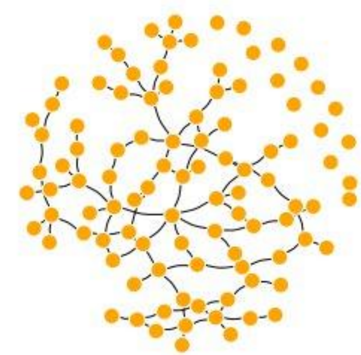
n=100, p=0.02



m=108



m=90



m=96

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

Teoría de redes aleatorias

Grafo aleatorio: es un grafo obtenido muestreando un *ensamble* 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** vertices y **m** enlaces

$$P(G) = \underbrace{p^m}_{\text{Prob de tener m enlaces}} \underbrace{(1-p)^{\binom{n}{2}-m}}_{\text{Prob de no tener los enlaces restantes}}$$

Teoría de redes aleatorias

Grafo aleatorio: es un grafo obtenido muestreando un *ensamble* 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** vertices y **m** enlaces

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

nro grafos con **n** vertices y **m** enlaces

Prob de observar en este ensamble **un** grafo con **n** vertices 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** vertices en este ensamble, sigue una **distribucion Binomial** de probabilidades

Repaso matematico…

G(n,p)

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

Supongamos que en una prueba exito/fracaso la probabilidad de éxito es *p.*
La probabilidad de tener *x* exitos en *N* pruebas independientes :

$$\underbrace{10001000011011100.......10}_{\text{R pruebas (con } x \text{ 1´s)}}$$

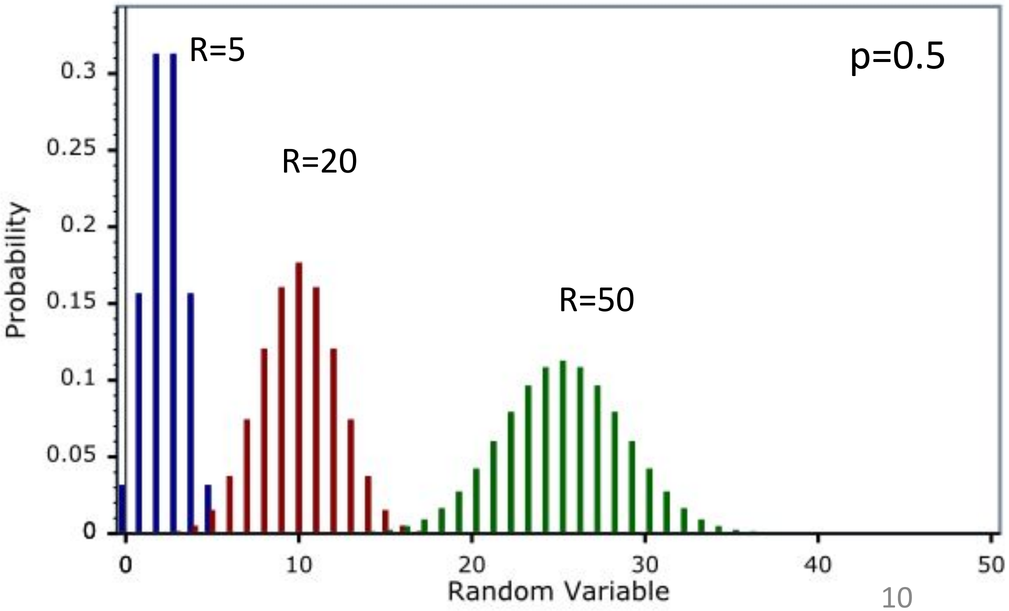
$$P(x)=\binom{R}{x}p^x(1-p)^{R-x}$$

$$\langle x\rangle=\sum xP(x)=Rp$$

$$\langle x^2\rangle=\sum x^2P(x)=p\,(1-p)R+p^2R^2$$

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

$$\frac{\sigma}{\langle x\rangle}\sim\frac{1}{\sqrt{R}}$$

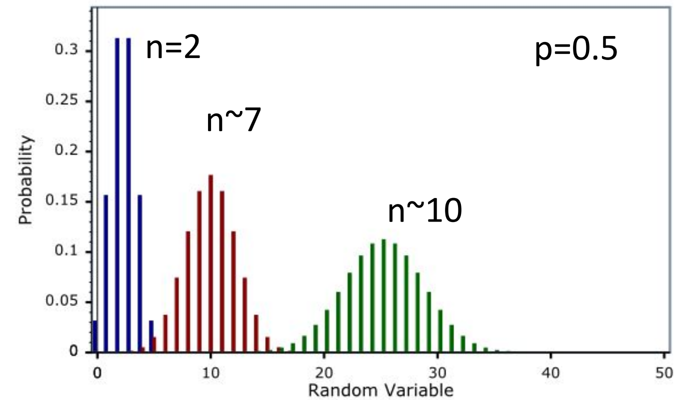


Propiedades de G(N,p)

G(n,p)

R

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



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

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

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

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

$$\langle x\rangle = Rp$$

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

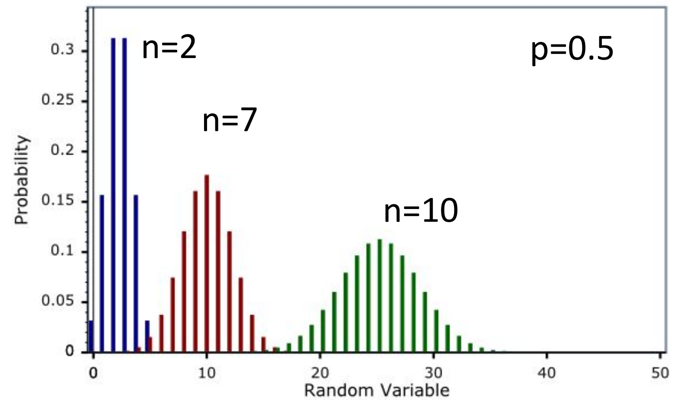
$$\sigma = \sqrt{\langle x^2\rangle - \langle x\rangle^2} = \sqrt{p\,(1-p)R}$$

Aunque el ensamble admite grafos con cualquier número de enlaces la mayor parte de los grafos poseen m~<m> si n>>1

Propiedades de G(N,p)

G(n,p)

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



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

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

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

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

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

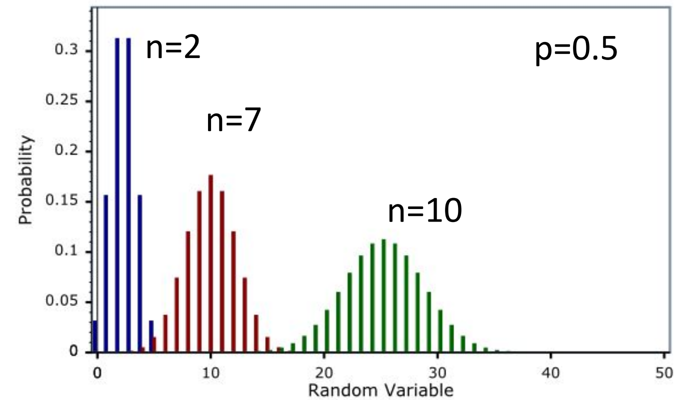
✔ Razonable: maximo nro de posibles enlaces de un nodo por la probabilidad **p** de que se establezca un enlace

La distribución de grado se angosta para redes gdes.

Propiedades de G(N,p)

G(n,p)

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



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

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

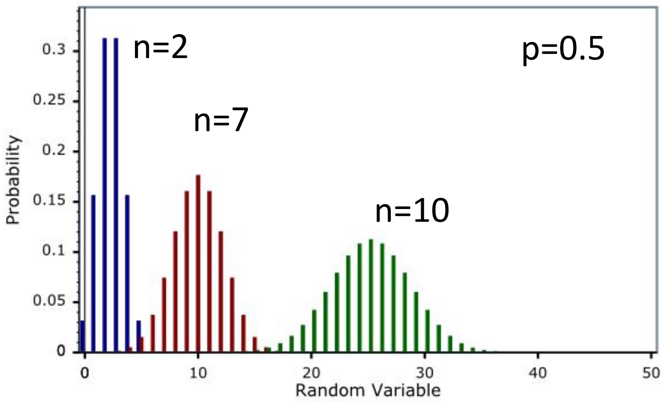
- $\langle k\rangle=\frac{2\langle m\rangle}{n}=(n-1)p$

✔ Razonable: maximo nro de posibles enlaces de un nodo por la probabilidad **p** de que se establezca un enlace

Propiedades de G(N,p)

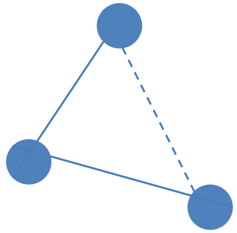
G(n,p)

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



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

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



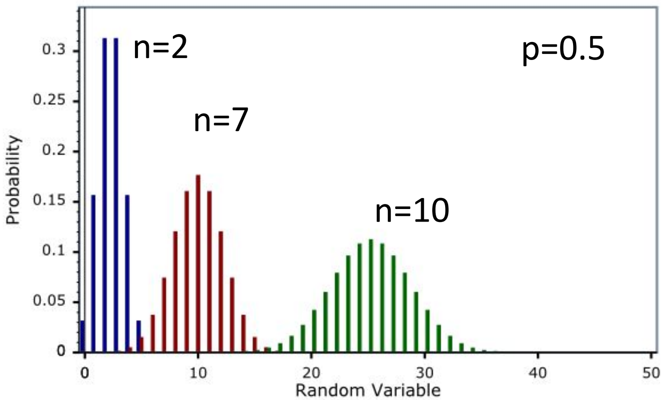
- **Coeficiente de Clustering:**

$$C_\Delta = 3\,\frac{\#triangulos}{\#tripletes} = 3\,\frac{\binom{N}{3}p^3}{\binom{N}{3}\binom{3}{2}p^2} = p$$

Propiedades de G(N,p)

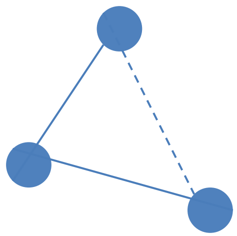
G(n,p)

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



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

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



- **Coeficiente de Clustering:**

$$C_\Delta = p$$

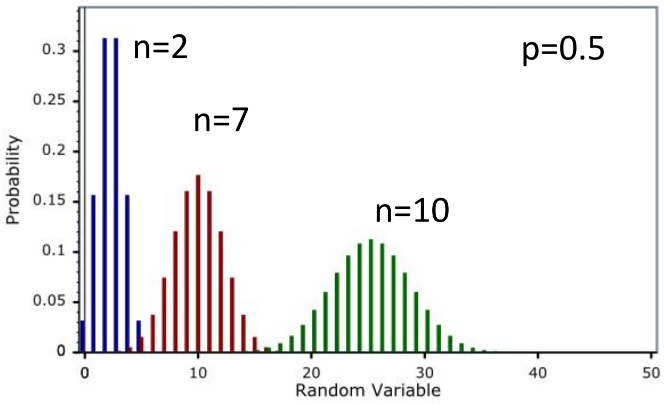
$$C_i = \frac{\#vec\ enlazados}{k_i(k_i-1)/2}$$

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

Propiedades de G(N,p)

G(n,p)

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



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

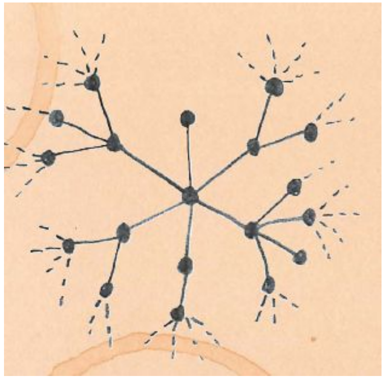
$C_\Delta = p$

$C_i = p$

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

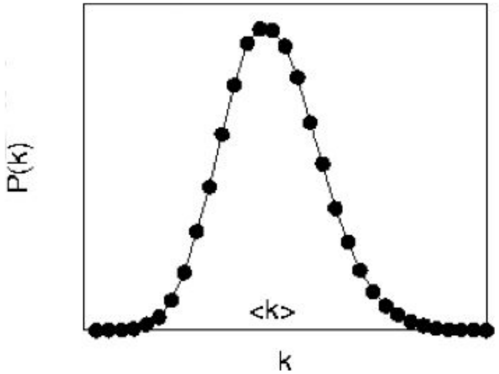
- $C = p$

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



G(N,p): Distribución de grado

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



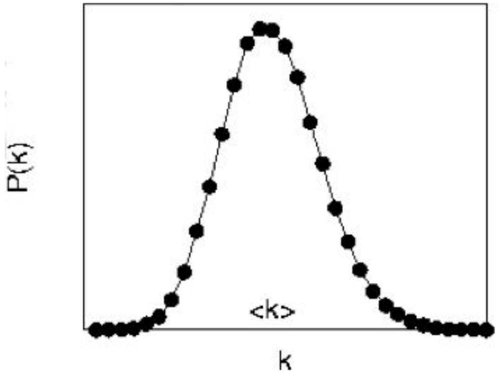
$$P_k = \binom{n-1}{k} p^k (1-p)^{n-1-k}$$

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

- P_k sigue una distribución binomial
- Si $n \to \infty$ y <k>=(n-1)p se mantiene finito, entonces $p \to 0$. Vamos a analizar P_k en éste limite

G(N,p): Distribución de grado

Quiero estimar $\boldsymbol{P_k}$ 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

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

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

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

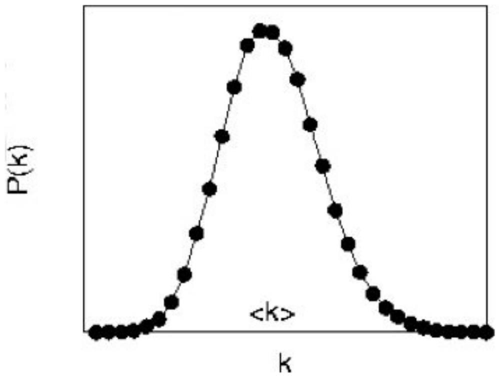
$$\log(1-p) \sim -p$$

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

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

G(N,p): Distribución de grado

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



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

Analizo los diferentes factores cuando

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

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

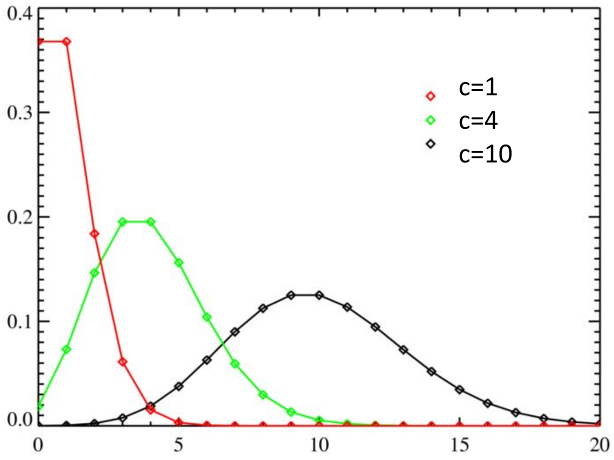
$$\underbrace{\binom{n-1}{k}} = \frac{(n-1)!}{(n-1-k)!\,k!} = \frac{(n-1)(n-2)\ldots(n-1-k+1)}{k!} \sim \frac{(n-1)^k}{k!}$$

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

Distribución de Poisson

Repaso matemático

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



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

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

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

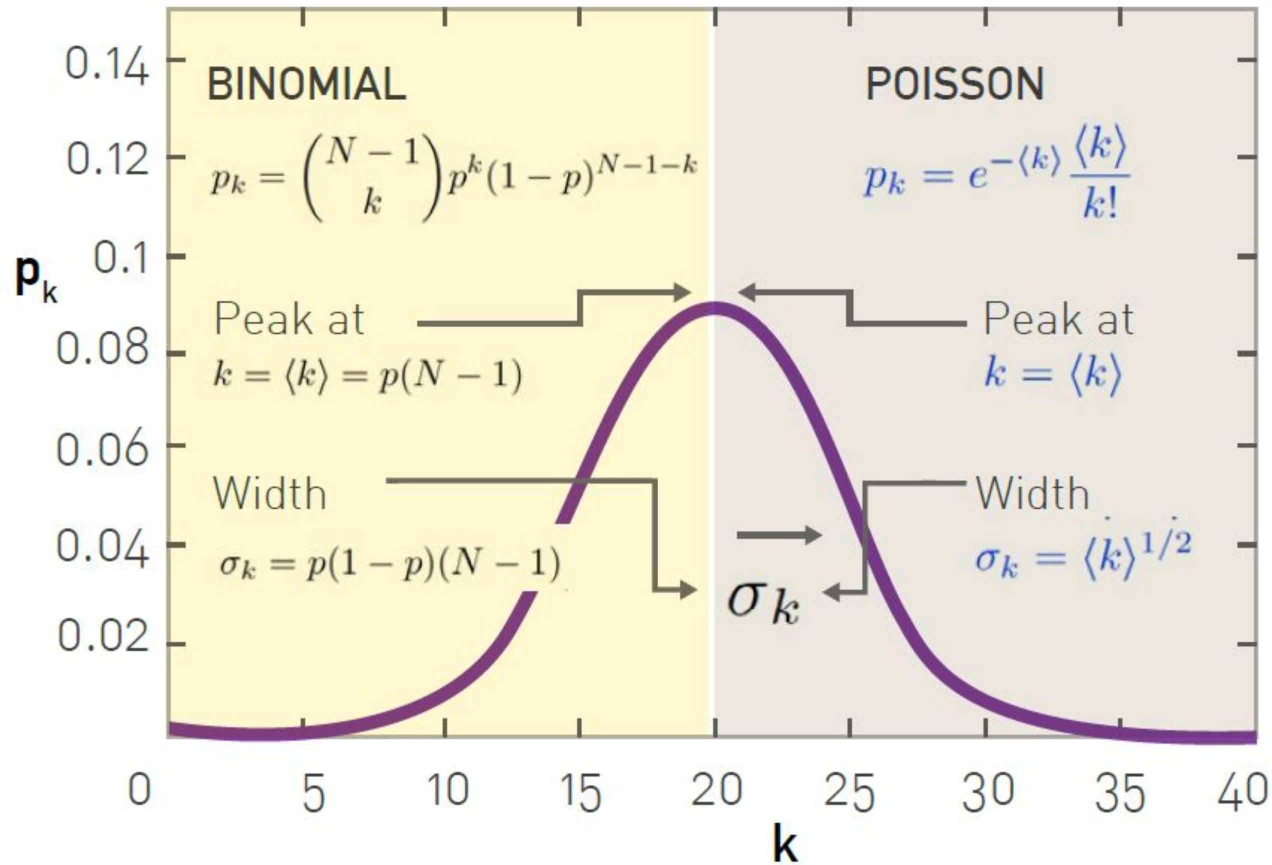
La varianza es igual a la media (!)

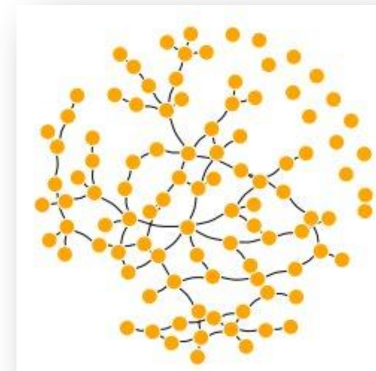
$$\frac{\sigma}{\langle k \rangle} = \frac{1}{\sqrt{\langle k \rangle}}$$

G(N,p): Distribución de grado

Resultado exacto

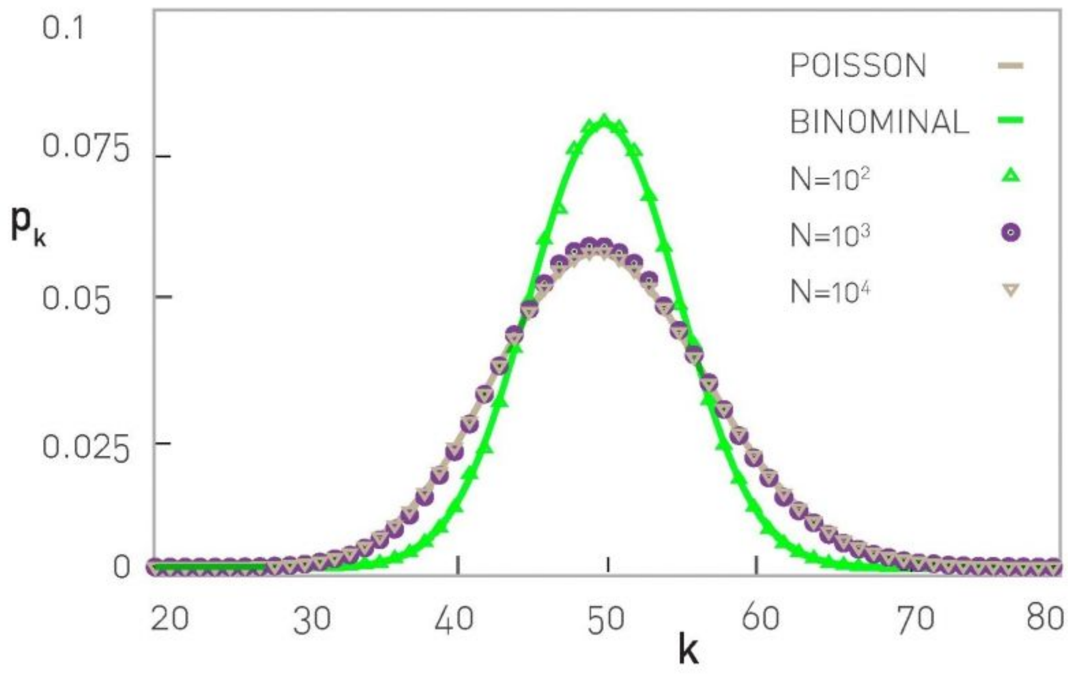
Límite N>>1, <k> finito



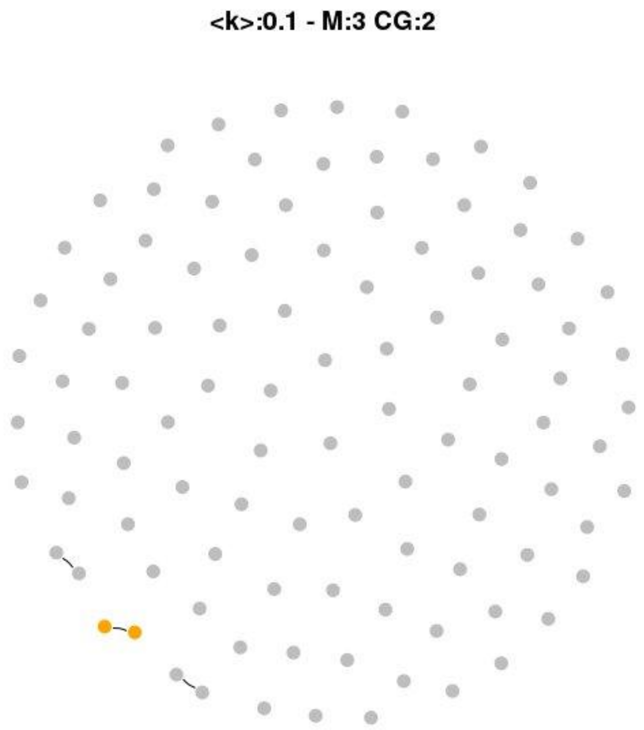


Pisson vs Binomial

La distribución de **Poisson** es una aproximación a la **binomial.**
Sin embargo hace más tratables algunas cuentas….

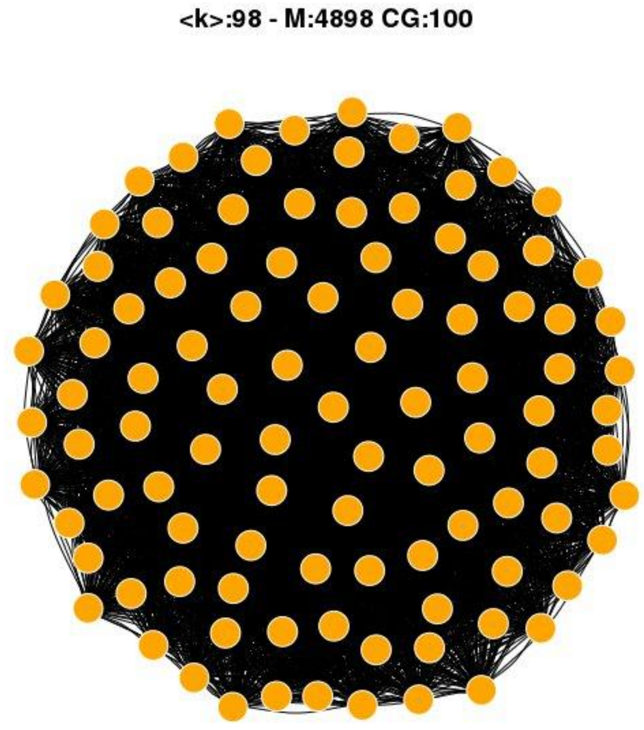


Conectividad de G(N,p)



Como se produce esta transición?

$$\langle k \rangle = p(N-1)$$



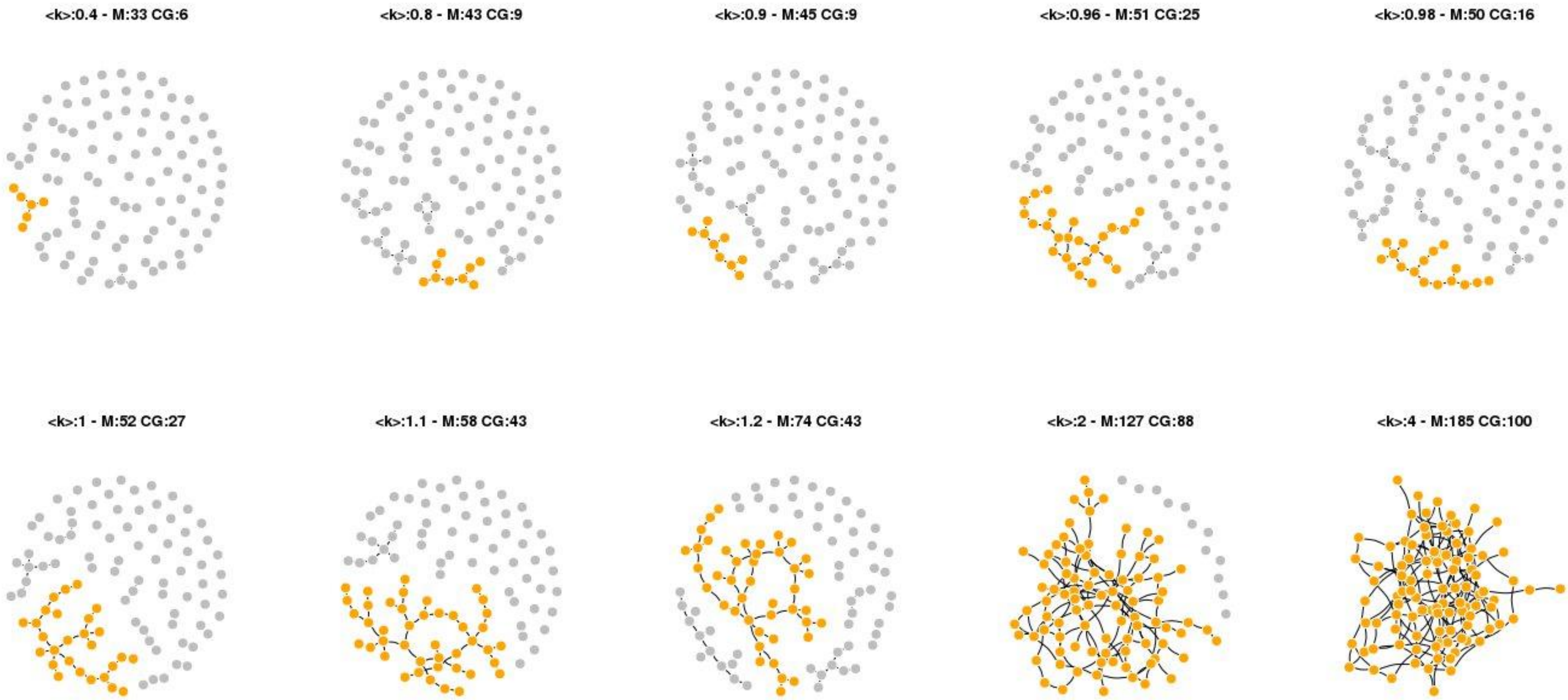
p ~ 0

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

p ~ 1

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

La transición



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

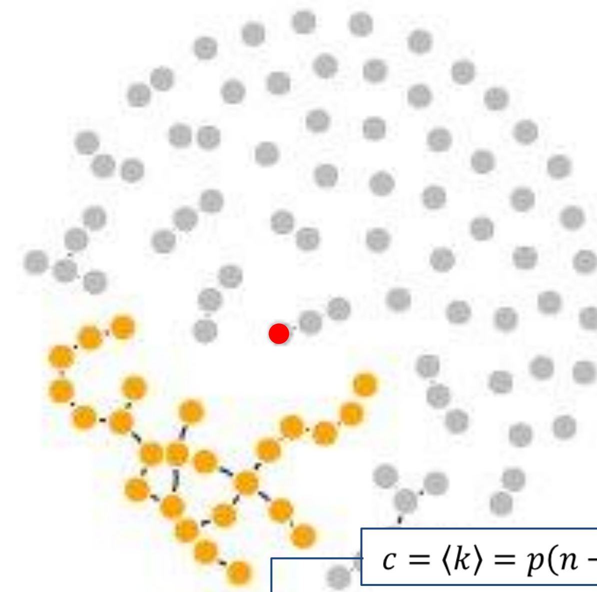
Tamaño de la **C**omponente **G**igante

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*



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

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

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

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

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

$$\log u = (n-1)\log\left[1 - \frac{c}{n-1}(1-u)\right] \sim -c\,(1-u)$$

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

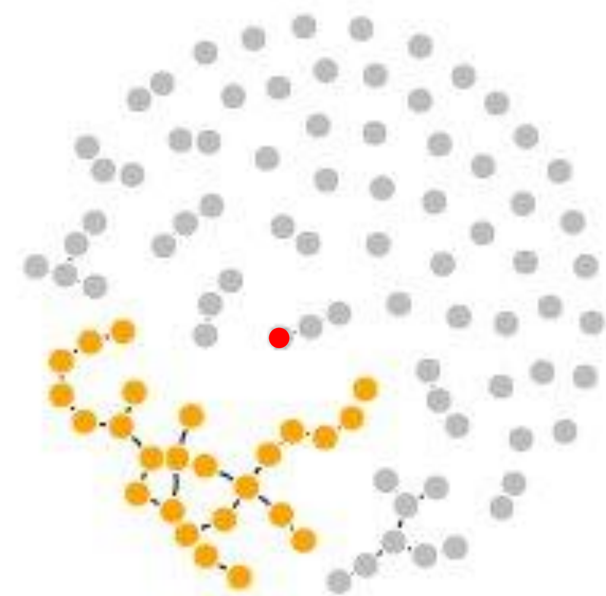
Tamaño de la **C**omponente **G**igante

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

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

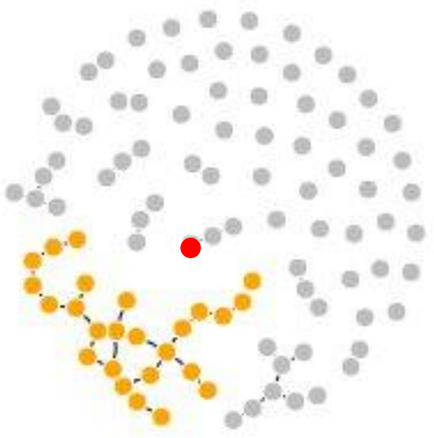
Entonces, la probabilidad de que nodo-i **no pertenezca** al CG resulta : $(1- p + p\, u)^{n-1}$

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

Tamaño de la **C**omponente **G**igante

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

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



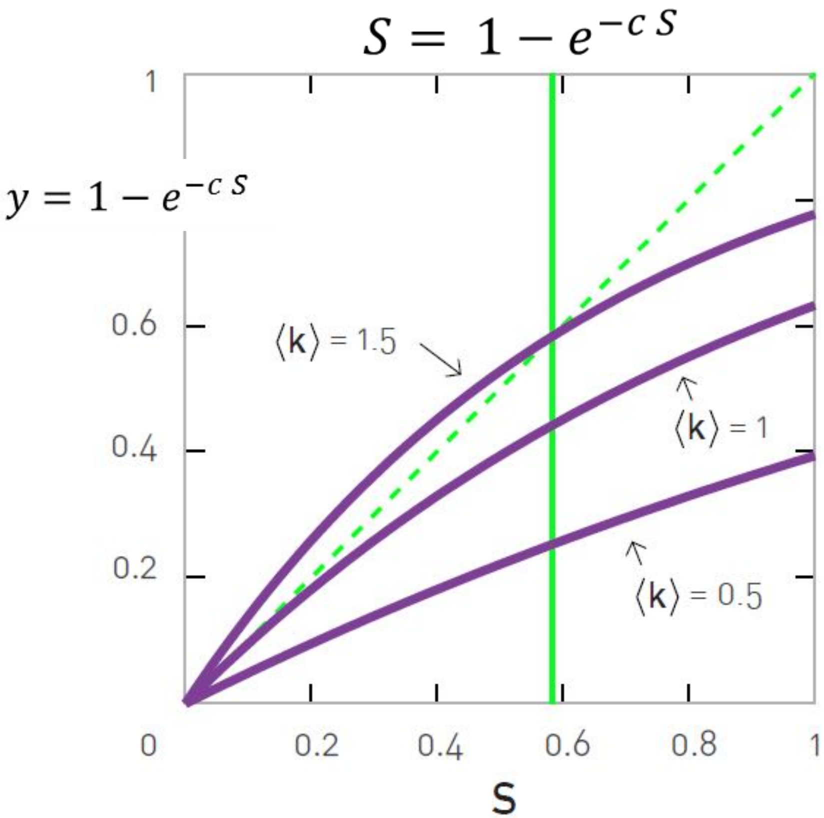
S = 1-u , fracción de nodos en la CG

$$1 - S \sim e^{-c\,S}$$

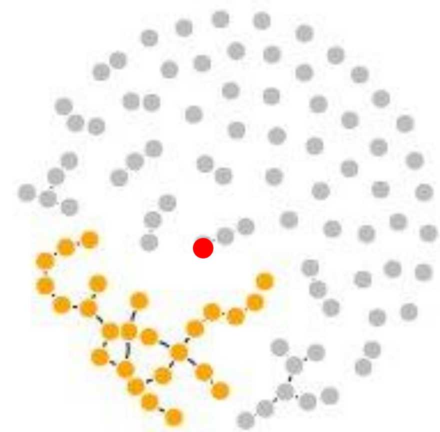
$$\underline{S = \ 1 - e^{-c\,S}}$$

(ER 1959)

Tamaño de la **C**omponente **G**igante



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



$$\langle k \rangle = p_c(n-1) = 1$$

$$p_c \sim \frac{1}{n}$$

En redes gdes la CG comiena a aparecer para *p* chicas

- y(S) es monotona 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**

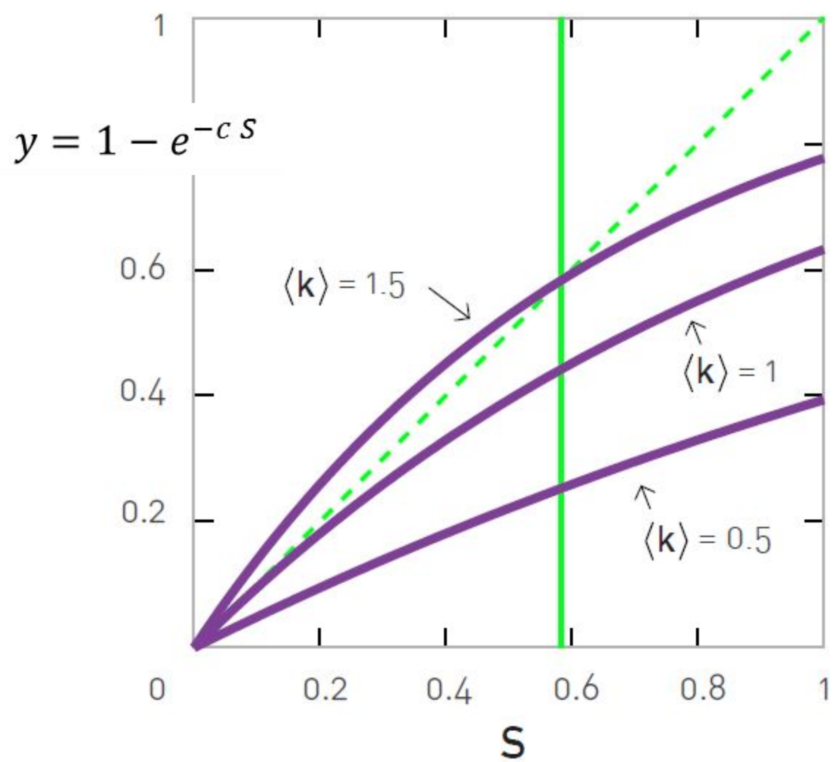
$$\left.\frac{dy}{dS}\right|_{S=0} = \left. ce^{-cS}\right|_{S=0} = c$$

c<1 S*=0

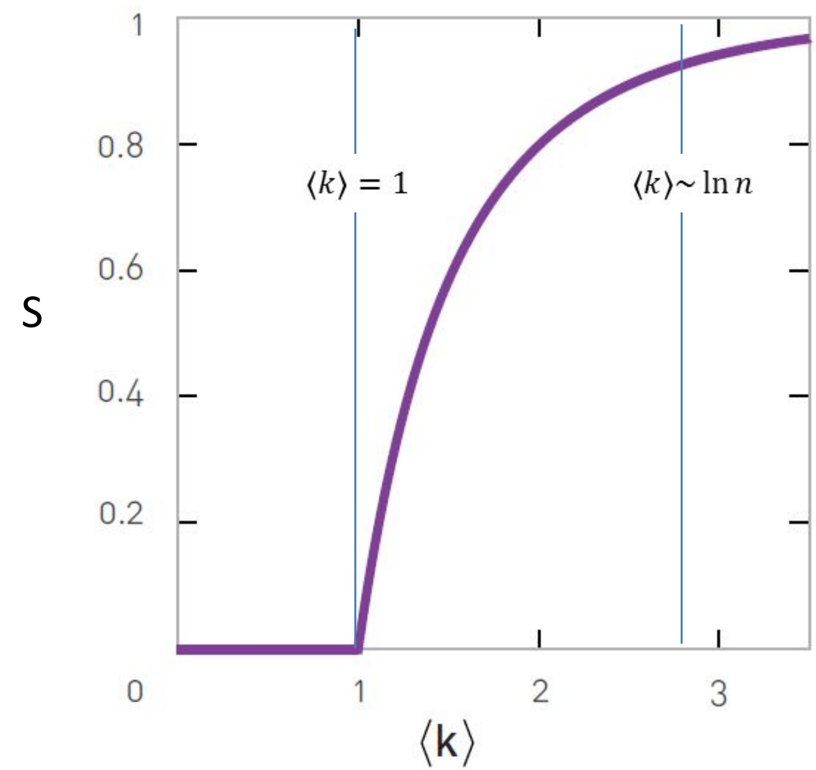
c≥1 S*>0 ademas de la trivial

Tamaño de la **C**omponente **G**igante

$$S = 1 - e^{-c\,S}$$



Es posible distinguir 4 regimenes



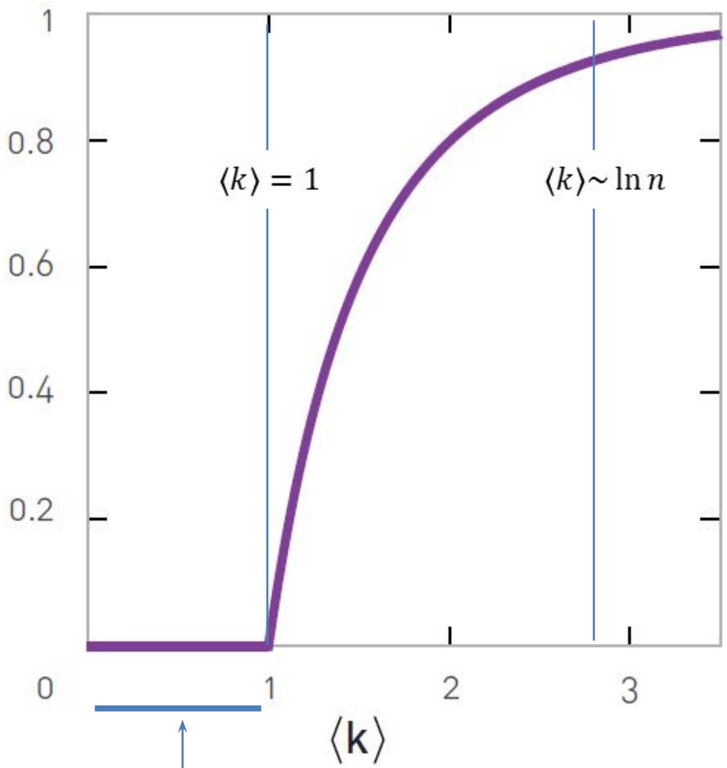
- y(S) es monotona creciente y tiene curvatura negativa
- Hay un **cambio cualitativo** en las soluciones S*=y(S*) para los casos en que la **pendiente en el origen** toma valor **uno**

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

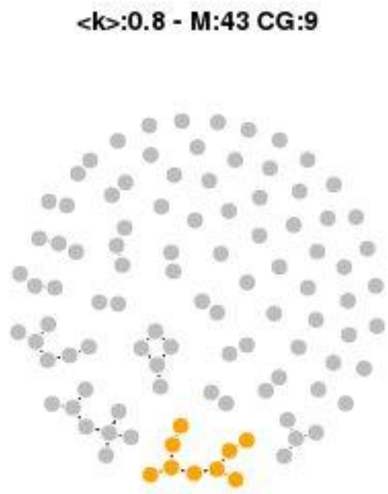
c<1 S*=0

c≥1 S*>0 ademas de la trivial

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



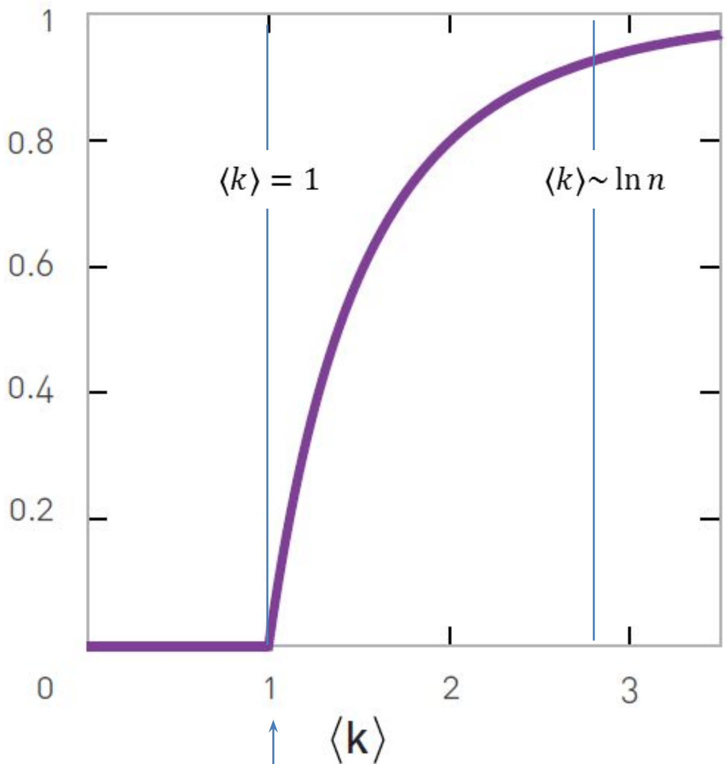
régimen
subcrítico



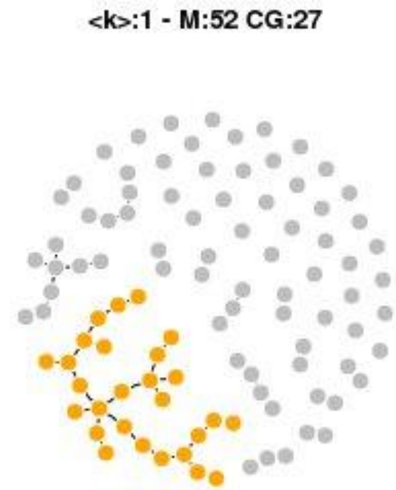
- no hay componente gigante
- componente más gde de tamaño ~ *ln n*
- Fraccion de nodos en GC $\sim \frac{\ln n}{n} \rightarrow 0$
- componentes aislados con distribución de tamaños exponencial

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

Conectividad G(N,p). Régimen crítico



régimen
crítico



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

igual…comparar con regimen subcritico:

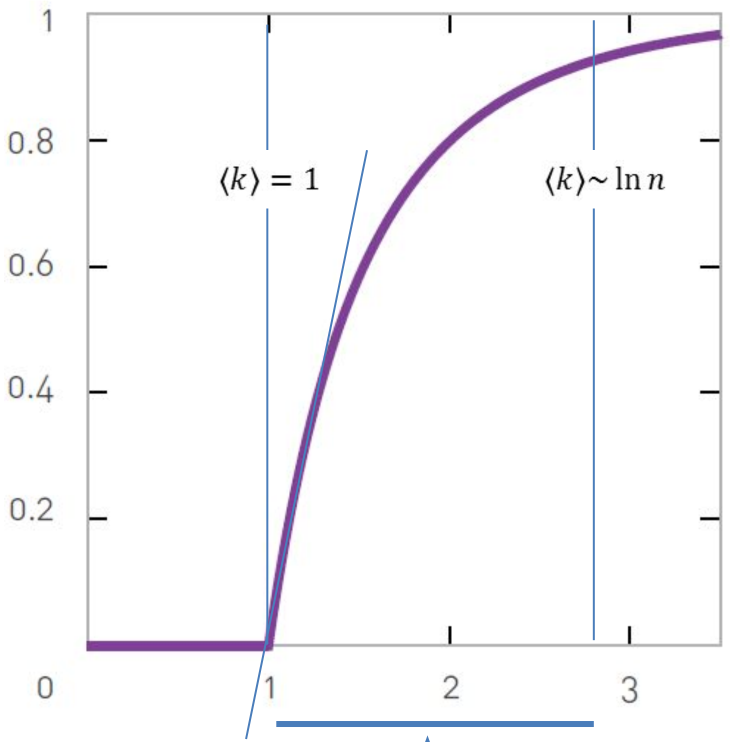
n=1000: ln n ~ 7 vs $n^{2/3}$ ~ 95

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

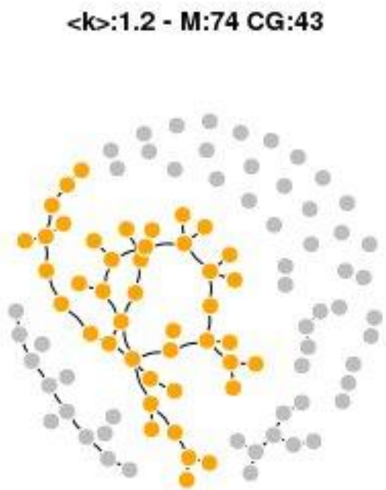
- componentes no-gigantes sin loops
- distribución de tamaño: ley de potencias

$$p(size) \sim s^{-3/2}$$

Conectividad G(N,p). Régimen supercrítico



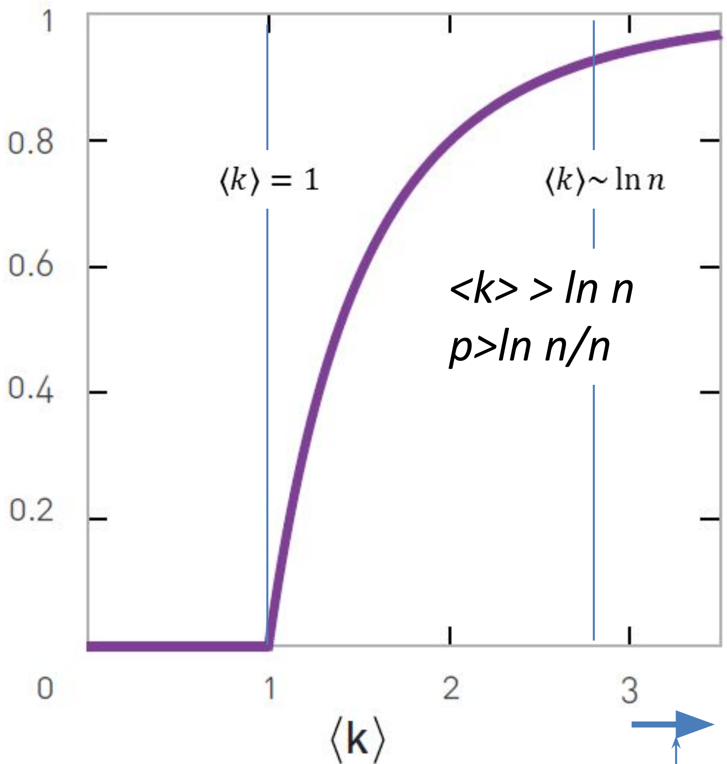
régimen supercrítico

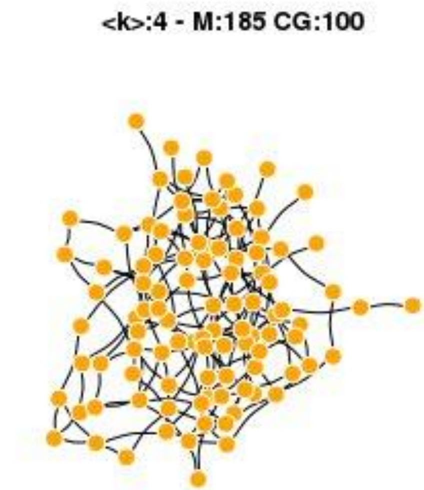


- componente gigante único $n_{CG} \sim (p - p_c) n$
- componente gigante tiene loops
- otros componentes aislados con distribución de tamaños exponencial

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

Conectividad G(N,p). Régimen conectado

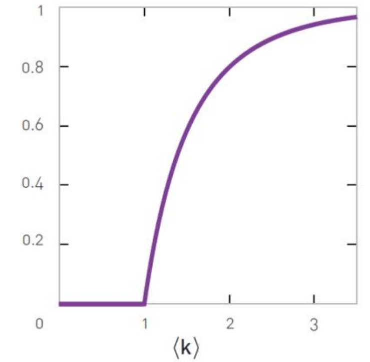




régimen conectado

- 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: $S_1.n$ y $S_2.n$
- 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 se conecte debe ser nula

$$q = (1-p)^{S_1 S_2 n^2} = \left(1 - \frac{c}{n-1}\right)^{S_1 S_2 n^2} \quad \sim e^{-ncS_1S_2}$$

$$\log q = S_1 S_2 \lim_{n\to\infty} n^2 \log\left(1 - \frac{c}{n-1}\right)$$

$$\sim - S_1 S_2 \lim_{n\to\infty} n^2 \frac{c}{n-1} \sim - ncS_1S_2$$

La probabilidad de que **no** exista vínculo entre S1 y S2 decae exponencilmente

Redes Aleatorias

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

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

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

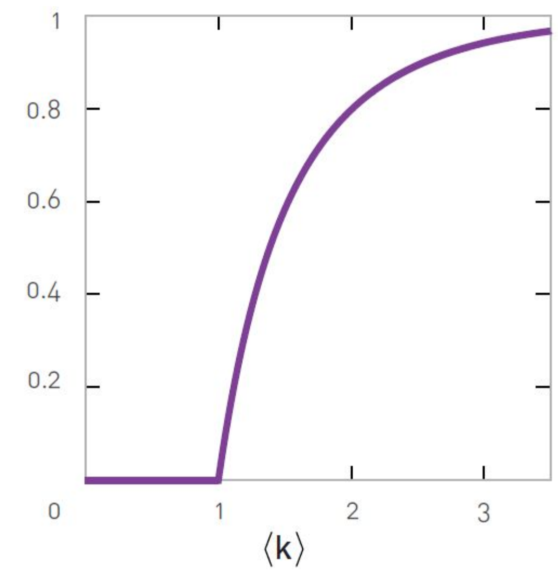
distribucion de grado

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

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

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

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



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

Redes Aleatorias

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

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

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

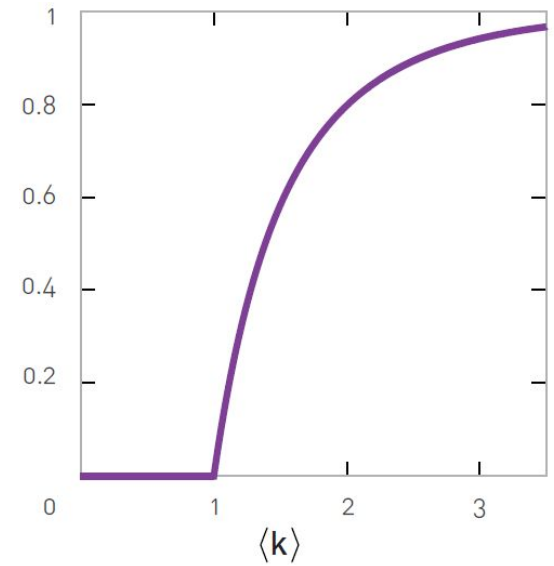
distribucion de grado

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

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

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

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



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