Mecanismos de Formación

Aníbal Olivera Morales

Mecanismos de Formación

1. Redes aleatorias
    1. Número de enlaces
    2. Grado medio
    3. Distribución de grado
    4. Coeficiente de clusterización
    5. Componente principal
2. Redes de mundo pequeño
    1. Propiedad de mundo pequeño
    2. Modelo Watts-Strogatz
3. Redes libres de escala
    1. Propiedad libre de escala
    2. Significado libre de escala
    3. Propiedad mundo pequeño
    4. Modelo Albert-Barábasi

Redes aleatorias

Redes aleatorias

- Nomenclatura:

Redes aleatorias == Erdös–Rényi model == Gilbert model

- Definiciones:

$G(N, L)$ Model — *N* nodos y *L* enlaces, *ala* Erdös–Rényi

$G(N, p)$ Model — *N* nodos y *p* probabilidad de enlace, *ala* Gilbert

- Definición:

A random network consists of N nodes where each node pair is connected with probability p.

1) Comienza con *N* nodos aislados.
2) Selecciona un par de nodos y genera un número aleatorio entre *0* y *1*. Si el número excede *p*, conecta el par de nodos seleccionado con un enlace; de lo contrario, déjalos desconectados.
3) Repite el paso (2) para cada uno de los *N(N-1) / 2* pares de nodos.

Redes aleatorias

- Erdös papers:

On random graphs I.

Dedicated to O. Varg

By P. ERI

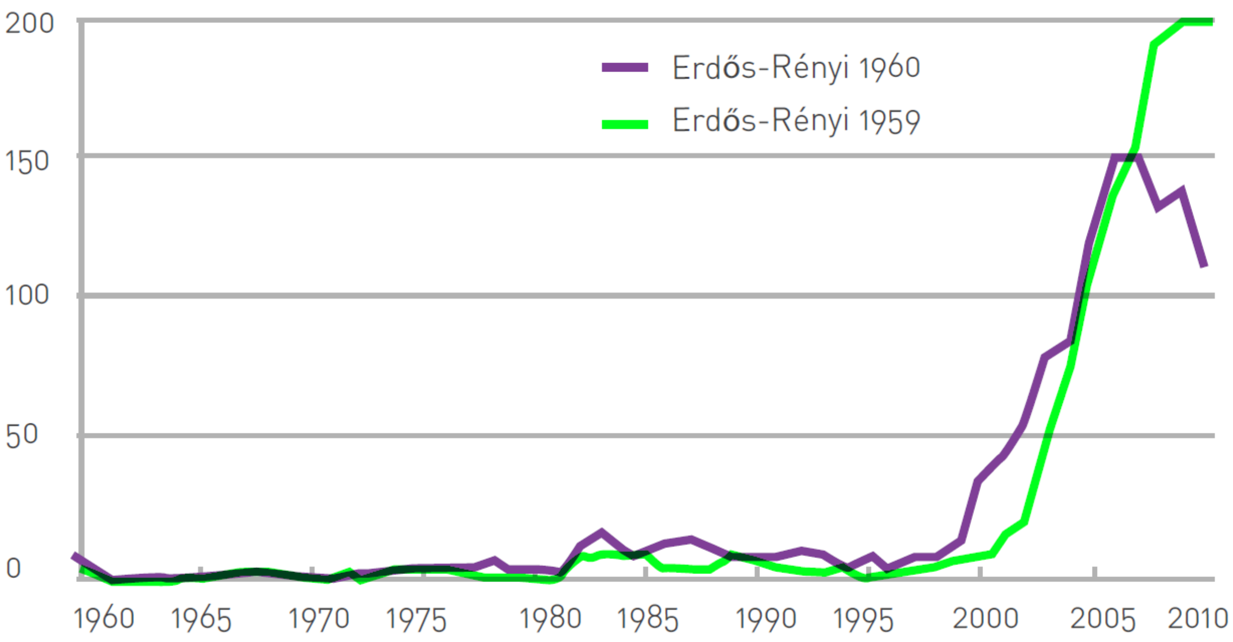
Let us consider a "rand vertices and N edges; in othe

probabilities) one of the $\left(\binom{n}{2} \atop N\right.$

the n (labelled) vertices P_1, I

possible edges $\widehat{P_i P_j}$ $(1 \leq i <$

$\Gamma_{n,N}$ may be less than n, as



ON THE EVOLUTION OF RANDOM GRAPHS

by

P. ERDÖS and A. RÉNYI

Institute of Mathematics
Hungarian Academy of Sciences, Hungary

1. Definition of a random graph

Let $E_{n,N}$ denote the set of all graphs having n given labelled vertices $V_1, V_2, \cdots, V_n$ and N edges. The graphs considered are supposed to be not oriented, without parallel edges and without slings (such graphs are sometimes called linear graphs). Thus a graph belonging to the set $E_{n,N}$ is obtained by choosing N out of the possible $\binom{n}{2}$ edges between the points $V_1, V_2, \cdots, V_n$, and therefore the number of elements of $E_{n,N}$ is equal to $\binom{\binom{n}{2}}{N}$. A random graph $\Gamma_{n,N}$ can be defined as an element of $E_{n,N}$ chosen at random, so that each of the elements of $E_{n,N}$ have the same probability to be chosen, namely $1/\binom{\binom{n}{2}}{N}$. There is however an other slightly different point of view, which has some advantages. *We may consider the formation of a random graph as a stochastic process* defined as follows: At time $t=1$

Redes aleatorias – Número de enlaces

- Si una red tienen *N* nodos, entonces existen *N(N-1)/2* pares de nodos.
- Probabilidad que obtener *x* sellos al tirar *N* veces una moneda:

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

Número de maneras distintas de tener *x* éxitos en *N* lanzamientos

Probabilidad que *x* lanzamientos hayan sido sello

Probabilidad que *N* lanzamientos hayan sido cara

Redes aleatorias – Número de enlaces

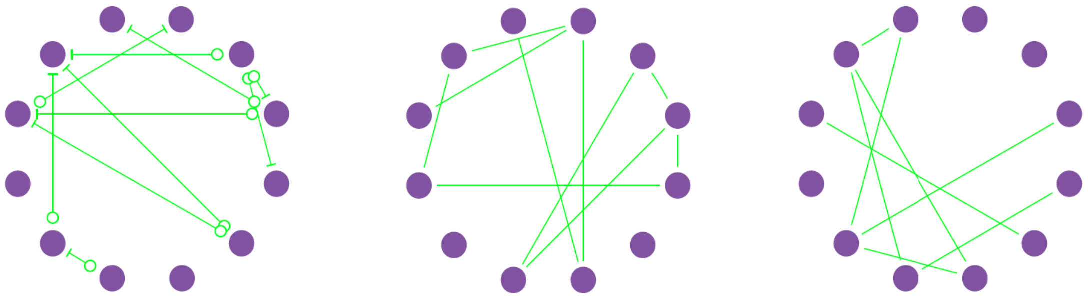
- Si una red tienen *N* nodos, entonces existen *N(N-1)/2* pares de nodos.
- Probabilidad que una realización tenga exactamente *L* enlaces:

$$p_L = \binom{\frac{N(N-1)}{2}}{L} p^L (1-p)^{\frac{N(N-1)}{2}-L}$$

Número de formas de colocar *L* enlaces entre *N(N-1)/2* pares

Probabilidad que *L* intentos hayan resultado en enlaces

Probabilidad que los restantes *N(N-1)/2 - L* pares no hayan terminado en con enlace



Redes aleatorias – Número de enlaces – Grado medio

- Por lo que el **número de enlaces** esperados en red aleatoria:

$$\langle L \rangle = \sum_{L=0}^{N} L\, p_L = p\frac{N(N-1)}{2}$$

- Y el **grado medio** en red aleatoria:

$$\langle k \rangle = \frac{2\langle L \rangle}{N} = p(N-1)$$

Que es equivalente al número de enlaces esperado para 1 solo nodo

Redes aleatorias – Distribución de grado

- En una red aleatoria, la probabilidad que un nodo *i* tenga exactamente *k* vecinos:

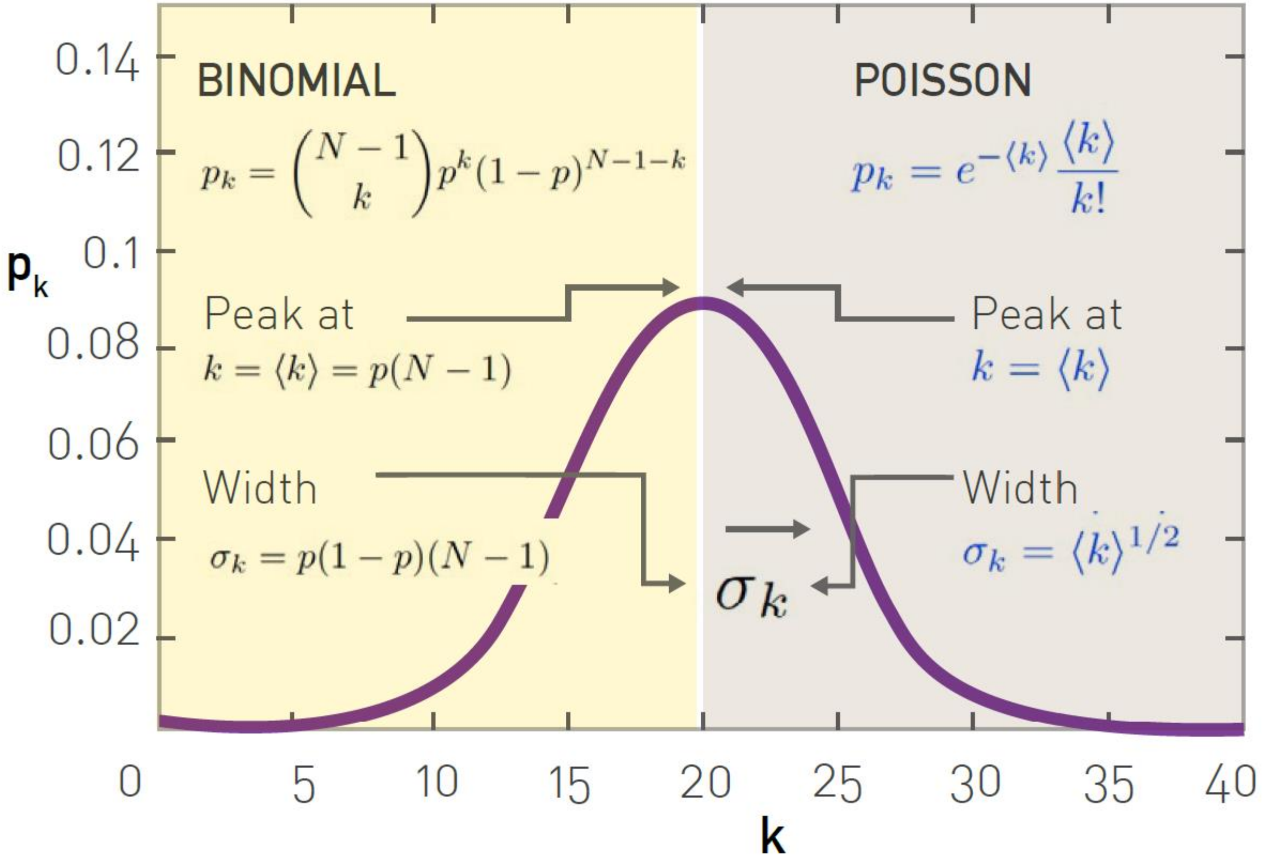
$$p_k = \binom{N-1}{k} p^k (1-p)^{N-1-k}$$

Número de formas de tener *k* enlaces entre *N-1* nodos

Probabilidad de tener *K* enlaces presentes en esa realización

Probabilidad que los restantes *N-1-k* nodos no presenten enlace

Redes aleatorias – Distribución de grado



En una red aleatoria, la probabilidad que un nodo *i* tenga exactamente *k* vecinos:

Redes aleatorias – Distribución de grado

- Ahora bien, la mayoría de las redes son dispersas: $< k > \ll N$

$$p_k = \binom{N-1}{k} p^k (1-p)^{N-1-k} \longrightarrow p_k = e^{-\langle k \rangle} \frac{\langle k \rangle^k}{k!}$$

- No depende de *N*, por lo que es agnóstica del tamaño.
- Depende de un solo parámetro, <*k*>.

- Las redes sociales reales **no son aleatorias**.

$$\langle k \rangle \approx 1,000 \rightarrow \sigma_k = 31.62$$

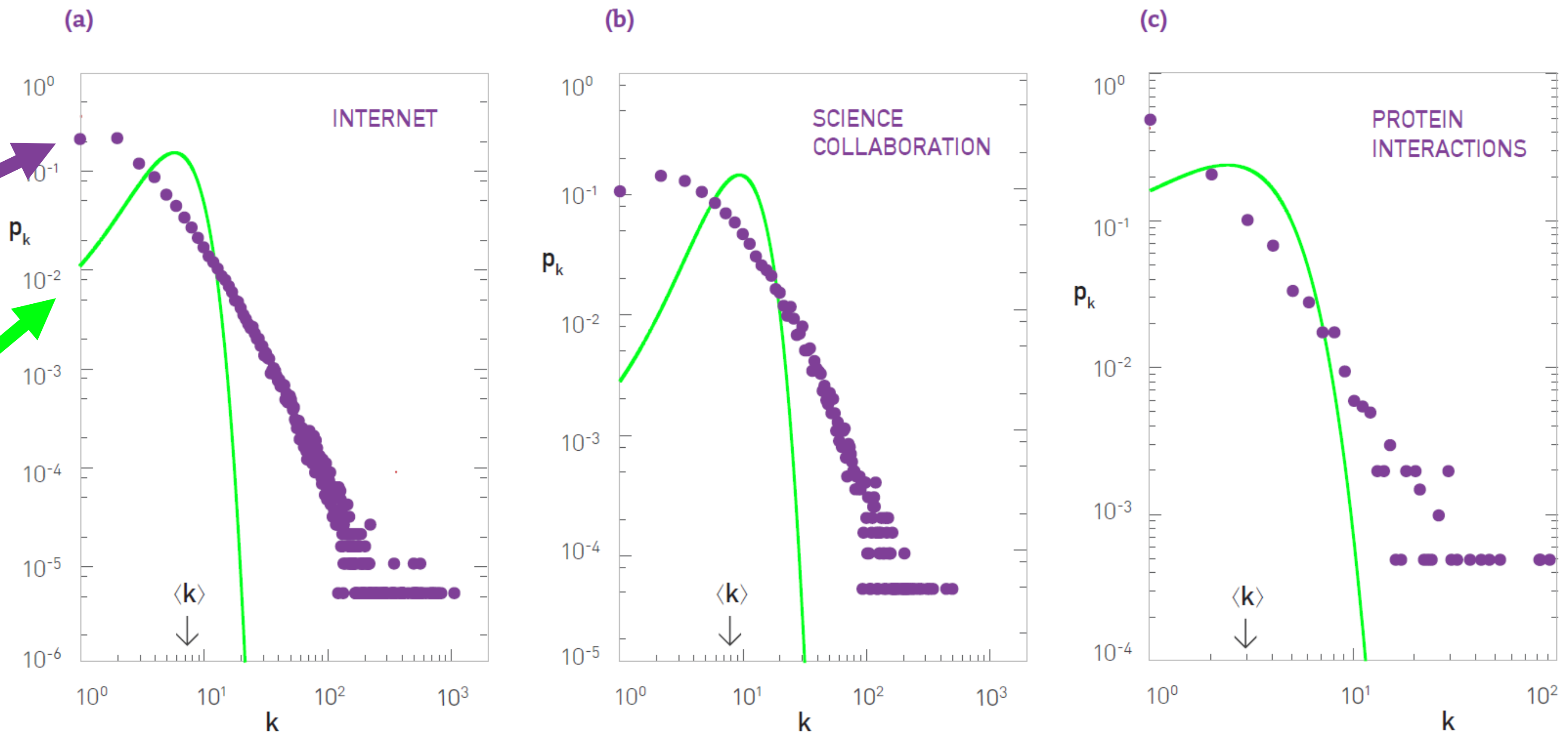
- O sea que las personas tienen entre 968 y 1032 conocidos, pero está documentado que Franklin Delano Roosevelt (US President) tenía ~22000 conocidos.

Redes aleatorias – Distribución de grado

- De hecho, no solo las redes sociales **no son aleatorias**:

Data Empírica

Predicción aleatoria



Redes aleatorias – Coeficiente de clusterización

- Recordemos que la clusterización local se definía como:

$$C_i = \frac{(\text{number of pairs of neighbors of } i \text{ that are connected})}{(\text{ number of pairs of neighbors of } i)}.$$

- Ahora, notemos que

$\langle L_i \rangle \rightarrow$ Número esperado de links entre los *k* vecinos de un nodo *i*

$\frac{1}{2}k_i(k_i - 1) \rightarrow$ Número de posibles links entre los *k* vecinos del nodo *i*

- Pero en una aleatoria simplemente se creará una fracción de los posibles

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

- Entonces

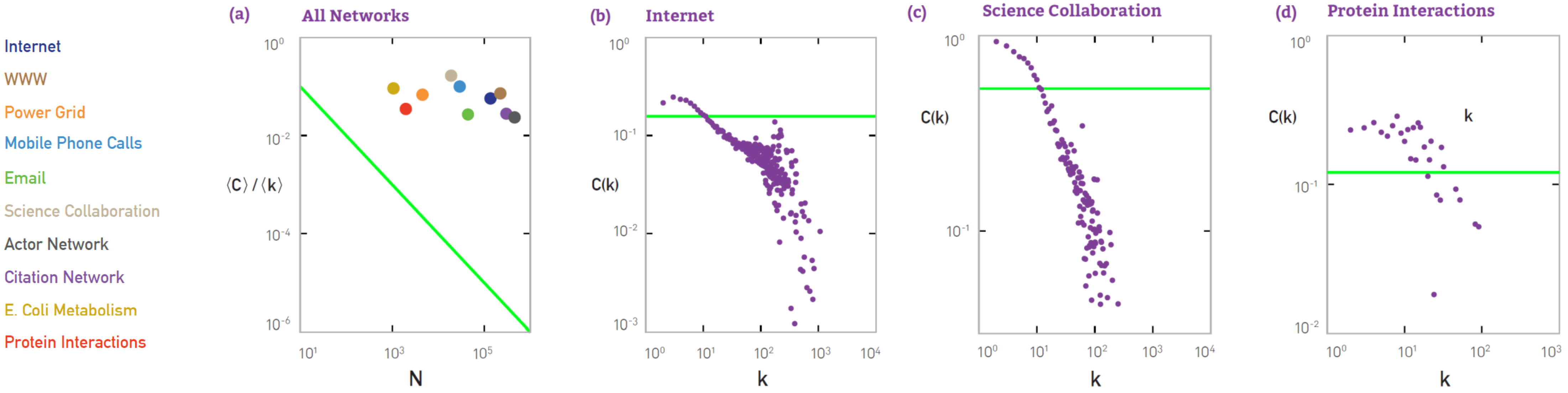
$$C_i = \frac{\langle L_i \rangle}{\frac{1}{2}k_i(k_i - 1)} = p = \frac{\langle k \rangle}{N}.$$

Redes aleatorias – Coeficiente de clusterización

$$C_i = \frac{\langle k \rangle}{N}.$$

- Notemos que:
  - Para redes aleatorias, la clusterización **global** debe ser igual a la **local**.
  - Si las redes fueras aleatorias, la clusterización debería decrecer según ***1/N***.
  - La clusterización **no depende de *k***, sino que del término global *<k>.*

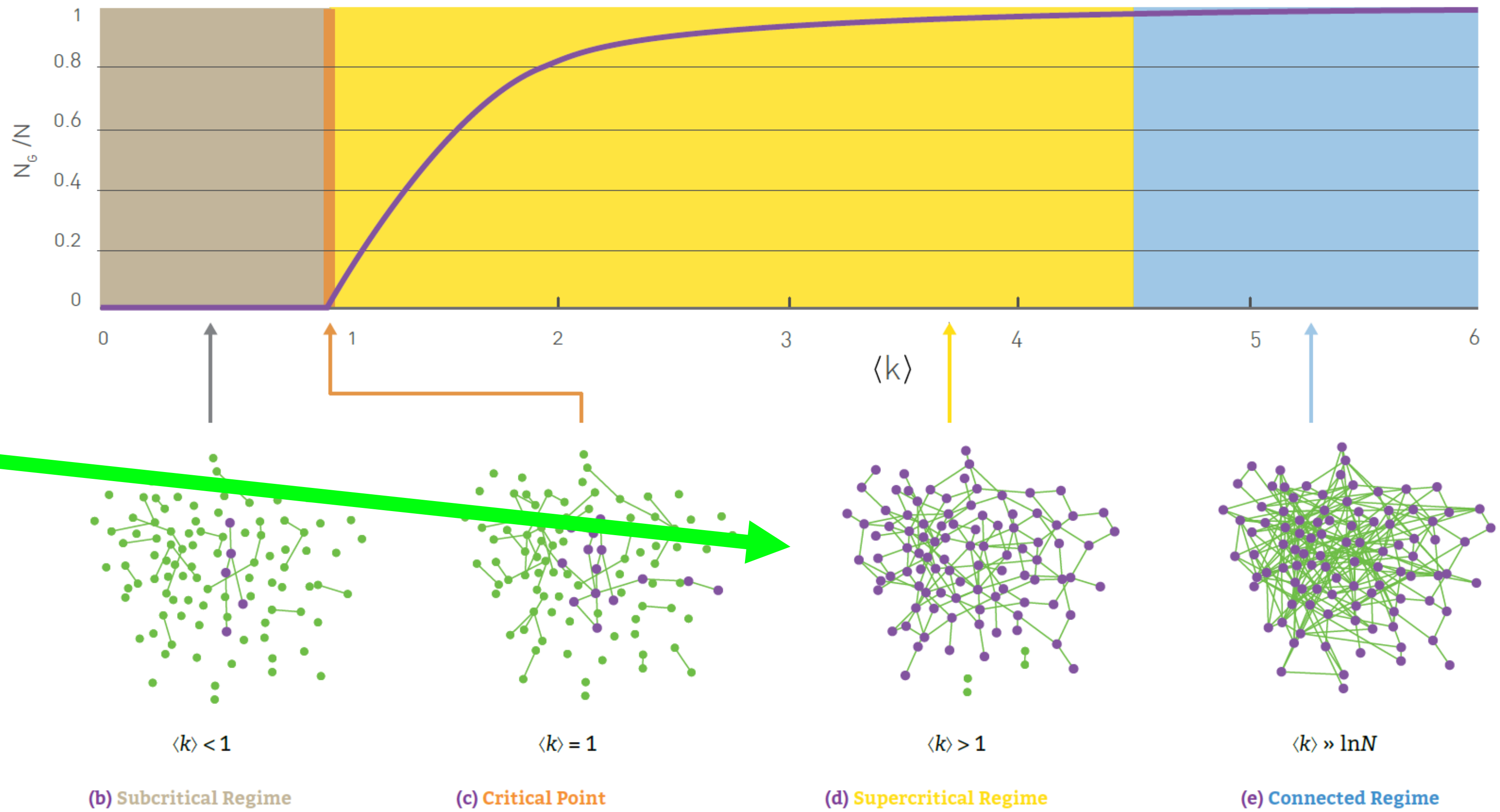
- Comparemos con redes reales:



Redes aleatorias – Componente principal

- En 1959, Erdös descubrió que la emergencia de una componente principal no es suave.

Cuando $\langle k \rangle > 1$ se reconoce una red

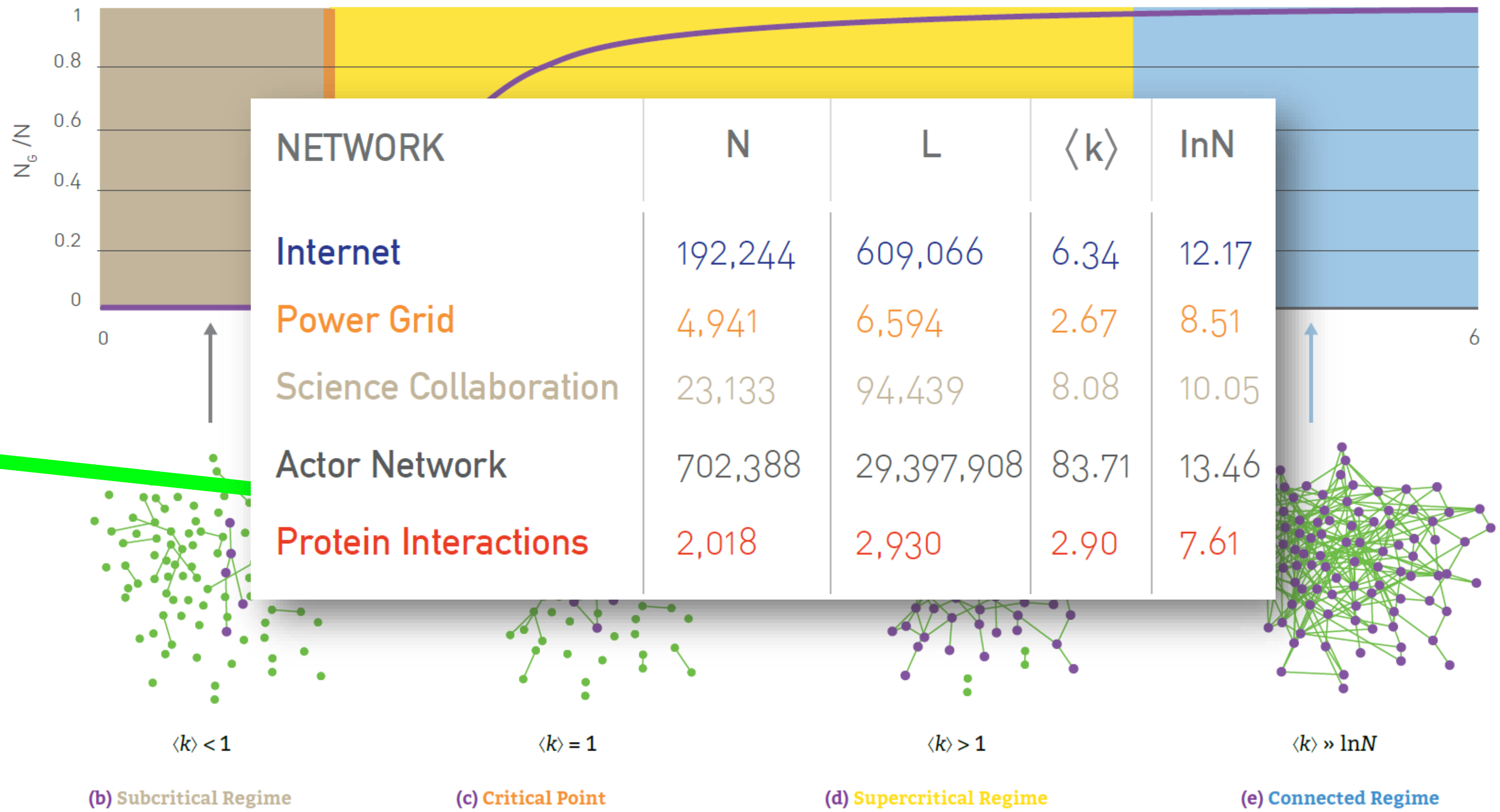


Cuando $\langle k \rangle \gg \ln N$, se espera solo 1 componente

Redes aleatorias – Componente principal

- En 1959, Erdös descubrió que la emergencia de una componente principal no es suave.

Cuando $\langle k \rangle > 1$ se reconoce una red



NETWORK	N	L	⟨k⟩	lnN
Internet	192,244	609,066	6.34	12.17
Power Grid	4,941	6,594	2.67	8.51
Science Collaboration	23,133	94,439	8.08	10.05
Actor Network	702,388	29,397,908	83.71	13.46
Protein Interactions	2,018	2,930	2.90	7.61

Cuando $\langle k \rangle \gg \ln N$, se espera solo 1 componente

Redes de Mundo Pequeño

Redes mundo pequeño – *propiedad* mundo pequeño

- Milgram: Seis grados de separación -> longitud de camino corta -> ¿comparado con qué?

- En una red aleatoria, ¿cuántas personas están a una distancia *d*?

$< k >$ nodos a distancia uno $(d = 1)$.
$< k >^2$ nodos a distancia dos ($d = 2$).
$< k >^3$ nodos a distancia tres $(d = 3)$. ..
..$< k >^{\mathrm{d}}$ nodos a distancia d.

- Por lo tanto, el número **total** de nodos que están como máximo a distancia *d*

$$N(d) \approx I + \langle k \rangle + \langle k \rangle^2 + \ldots + \langle k \rangle^d = \frac{\langle k \rangle^{d+1} - 1}{\langle k \rangle - 1}$$

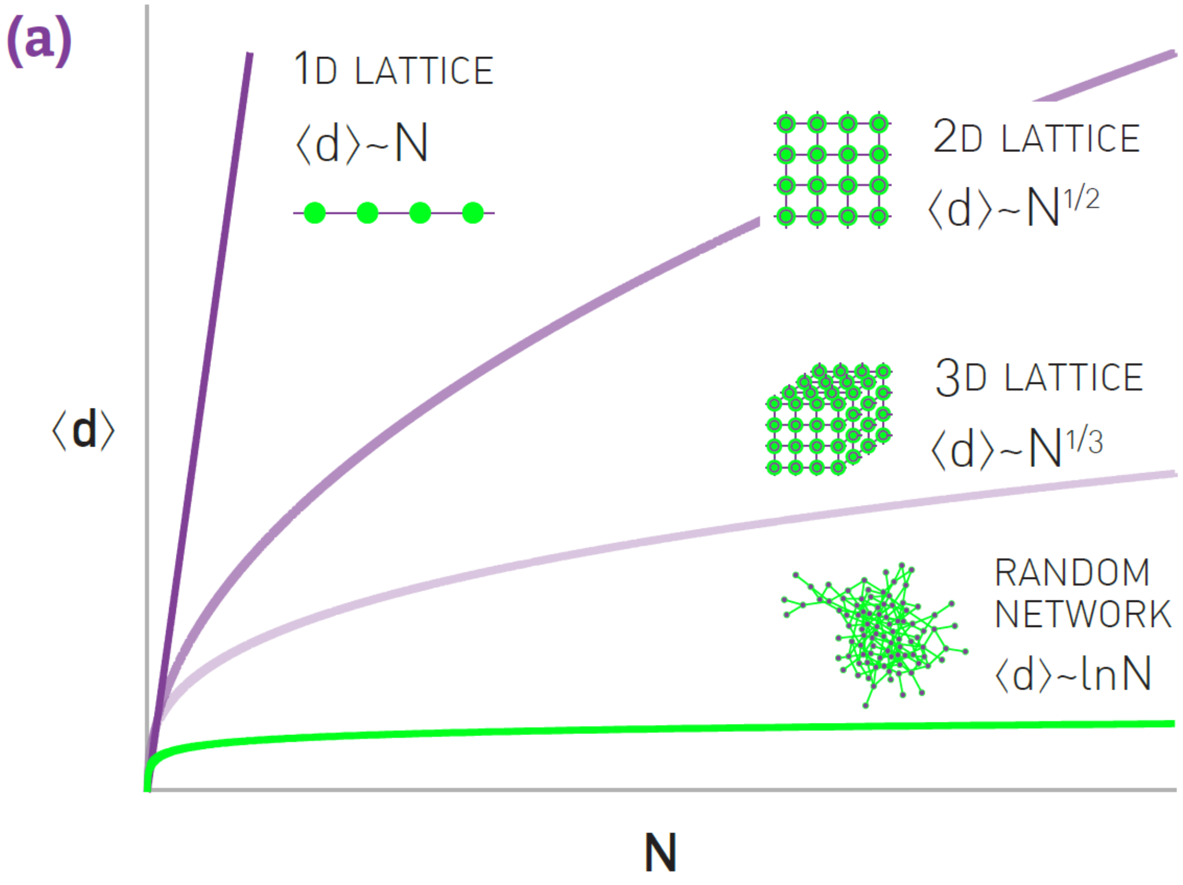
$$\to \langle k \rangle^{d_{\max}} \approx N. \qquad \to d_{\max} \approx \frac{\ln N}{\ln \langle k \rangle}$$

Diámetro de red aleatoria

<d> ~ *ln N* **propiedad mundo pequeño**

Redes mundo pequeño – *propiedad* mundo pequeño

- Comparemos el diámetro esperado en **redes aleatorias** vs **redes regulares**

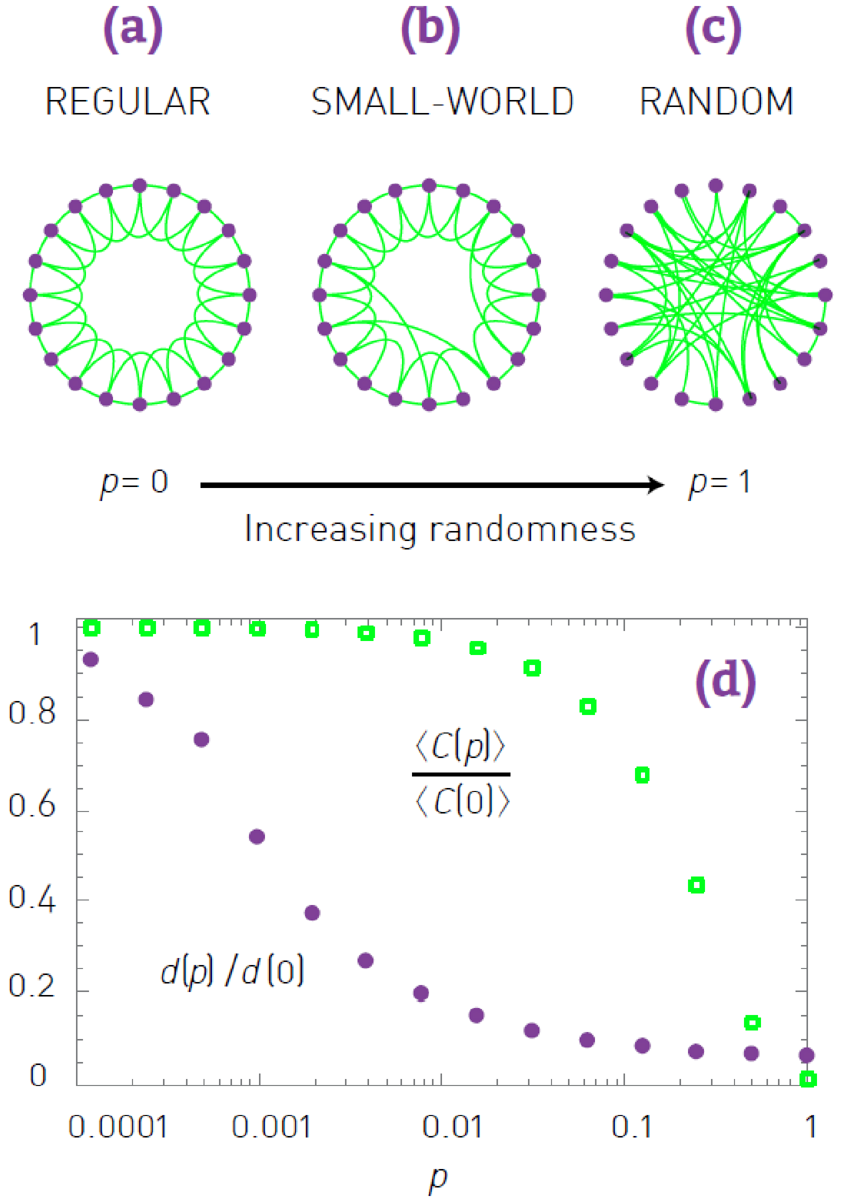


En general, para una red regular *d*-dimensional

$$\langle d \rangle \sim N^{1/d}$$

Redes mundo pequeño – modelo Watts-Strogatz

- Animados por:
  - Coeficiente de clusterización alto para redes reales según varía *N*.
  - Diámetro con crecimiento logarítmico con respecto a *N*.
- Notaron que:
  - Redes regulares
    alta clusterización, baja propiedad mundo pequeño
  - Redes aleatorias
    baja clusterización, alta propiedad mundo pequeño
- Sin embargo adolece de:
  - Distribución de grado Poisson similar a red aleatoria
    → no hubs
  - Clusterización similar a red regular
    → no dependencia en *k*.

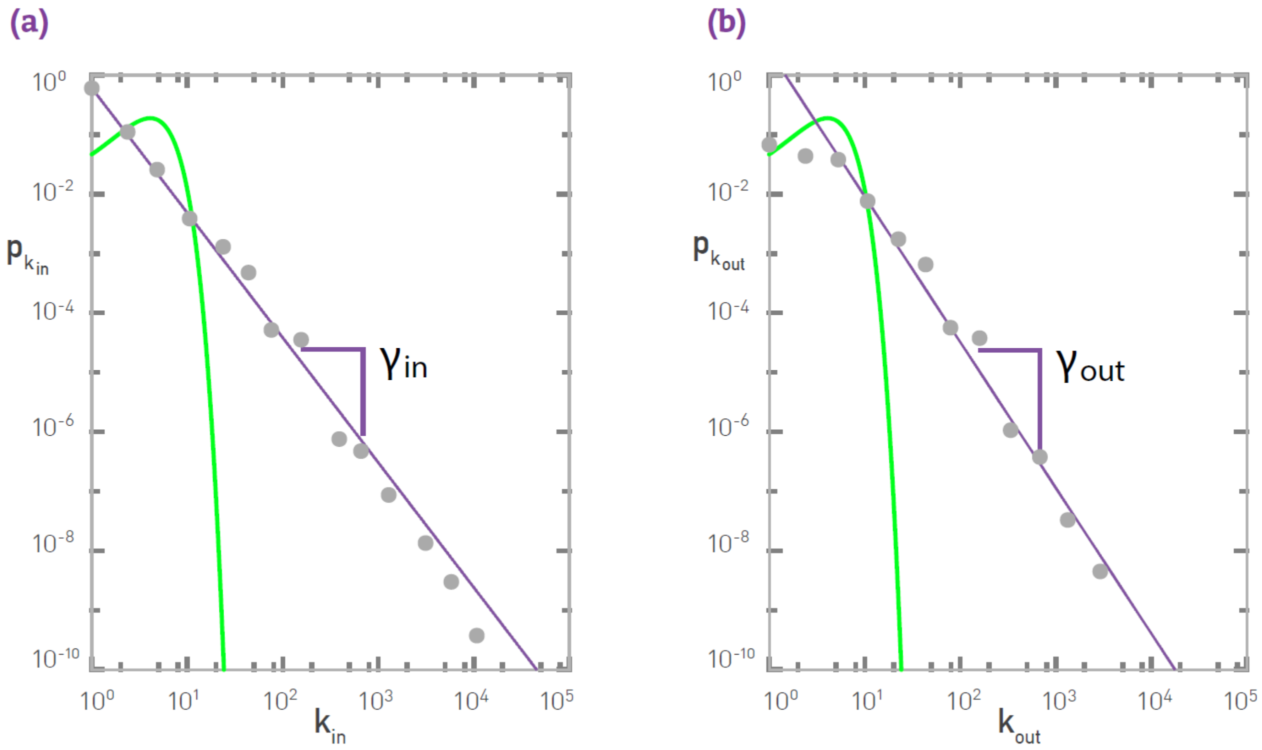


Redes Libres de Escala

Redes libres de escala – *propiedad* libre de escala

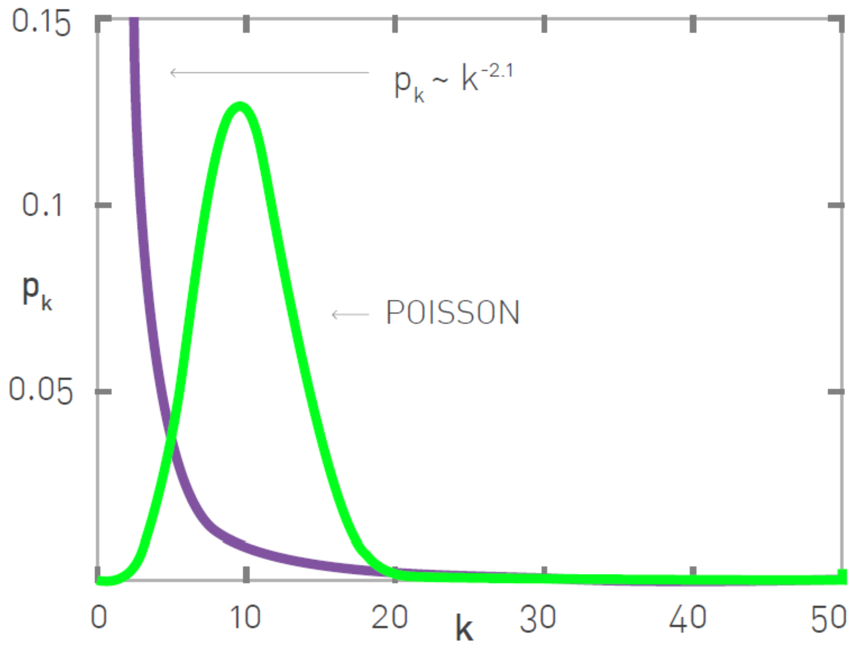
- Al ver la distribución de grado de la red WWW en un plot log-log, vemos que quedaría bien descrita por

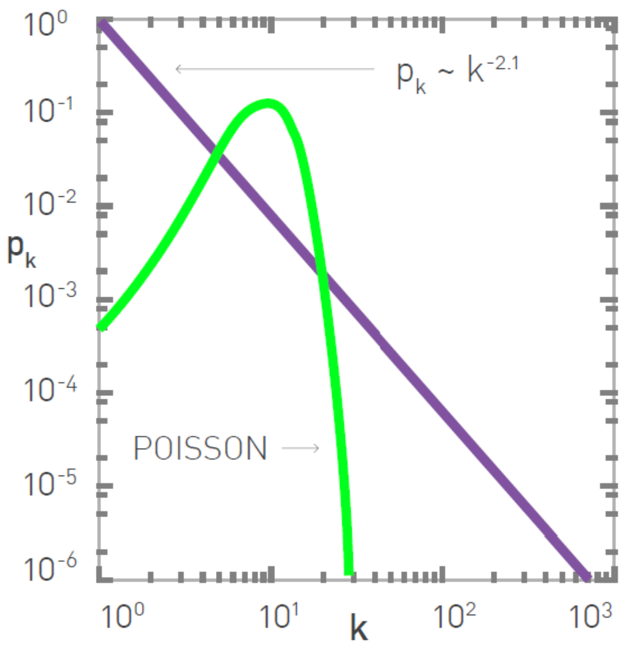
$$\log p_k \sim -\gamma \log k \longrightarrow p_k \sim k^{-\gamma}$$



Redes libres de escala – *propiedad* libre de escala

- La red aleatoria
  - Sub-estima la cantidad de nodos con *low-degree*.
  - Sobre-estima la cantidad de nodos con $k \sim \langle k \rangle$.
  - Sub-estima el número de nodos con *high-degree* (hubs).
- El hecho que la *recta* no dependa del *N* de la red, hace que se llame *libre de escala*.





Redes libres de escala – *significado* de libre de escala

- Para una red con distribución *power-law*, su *n*-ésimo momento:

$$\langle k^n \rangle = \int_{k_{\min}}^{k_{\max}} k^n p(k) dk = C \frac{k_{\max}^{n-\gamma+1} - k_{\min}^{n-\gamma+1}}{n-\gamma+1}$$

- Queremos ver el comportamiento para *k_max* → *infty*
  - Todos los momentos con $n < \gamma - 1$ son finitos.
  - Todos los momentos con $n > \gamma - 1$ son infinitos.
- La desviación estándar corresponde al momento *n=2.*
- Típicamente, estas redes tienen $2 < \gamma < 3$.
- Por lo tanto, el grado de un nodo seleccionado aleatoriamente

$$k = \langle k \rangle \pm \sigma_k \in [^{-}\infty, ^{+}\infty]$$

$$\langle x \rangle = \sum_{x=0}^{N} x p_x$$

$$\langle x^2 \rangle = \sum_{x=0}^{N} x^2 p_x$$

$$\sigma_x = \sqrt{\langle x^2 \rangle - \langle x \rangle^2}$$

Redes libres de escala – *significado* de libre de escala

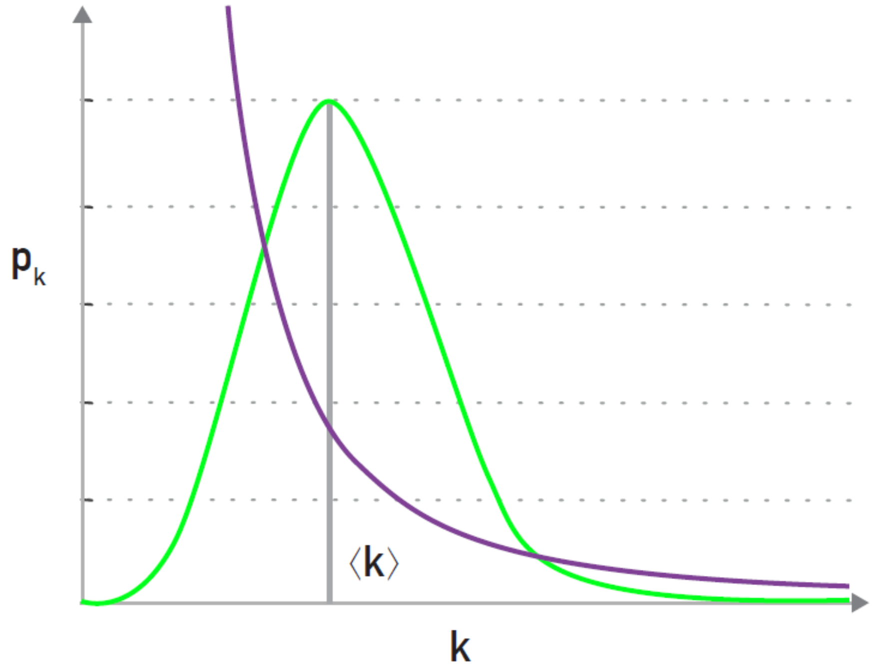
- Para una red con distribución *power-law*, su *n*-ésimo momento:

$$\langle k^n \rangle = \int_{k_{\min}}^{k_{\max}} k^n p(k) dk = C \frac{k_{\max}^{n-\gamma+1} - k_{\min}^{n-\gamma+1}}{n-\gamma+1}$$

- Queremos ver el comportamiento para *k_max* → *infty*
  - Todos los momentos con $n < \gamma - 1$ son finitos.
  - Todos los momentos con $n > \gamma - 1$ son infinitos.

- La desviación estándar corresponde al momento *n=2.*
- Típicamente, estas redes tienen $2 < \gamma < 3$.
- Por lo tanto, el grado de un nodo seleccionado aleatoriamente

$$k = \langle k \rangle \pm \sigma_k \in [^{-}\infty, ^{+}\infty]$$



Random Network
Randomly chosen node: $k = \langle k \rangle \pm \langle k \rangle^{1/2}$
Scale: $\langle k \rangle$

Más pequeño que <*k*>

Scale-Free Network
Randomly chosen node: $k = \langle k \rangle \pm \infty$
Scale: none

Redes libres de escala – Modelo Albert-Barábasi

- Nomenclatura:

  Modelo Albert-Barábasi == Modelo BA == Modelo *scale-free*

- Es un *mecanismo de formación* que sigue un *preferential-attachment*.
  Antigua idea del fenómeno *rich-gets-richer*.

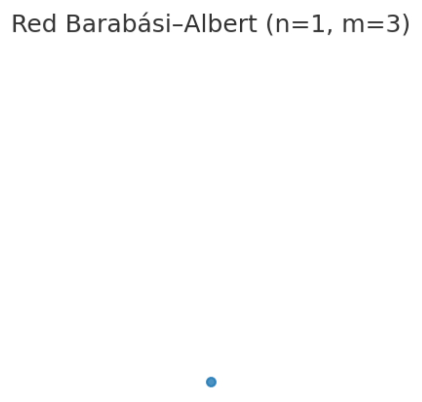
- Modelo:
  - Comenzamos con *q* nodos y *q* enlaces.
  - En cada paso un nuevo nodo se conecta a otros *m* nodos ya existentes, siguiendo *preferential-attachment:*

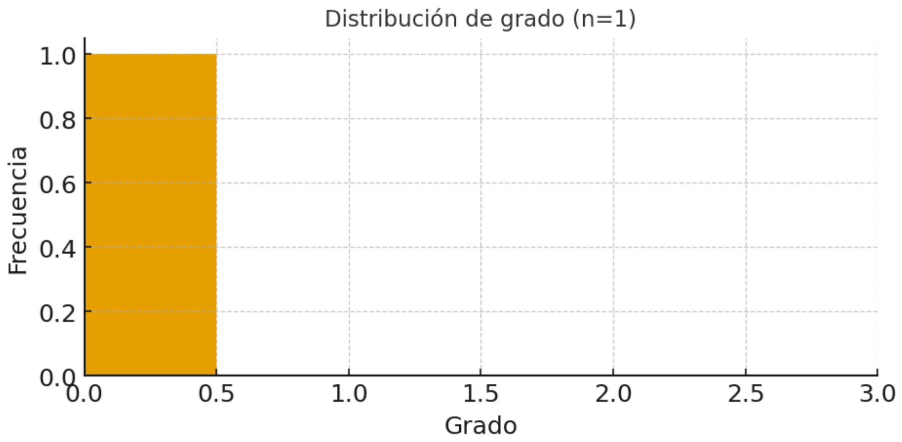
$$\Pi(k_i) = \frac{k_i}{\sum_j k_j}$$

  - Después de *t* pasos, tenemos un total de nodos $N=q+t$, y un total de enlaces $q+mt$.

Redes libres de escala
– Modelo Albert-Barábasi

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





Fin