Introdução à Teoria de Redes Complexas

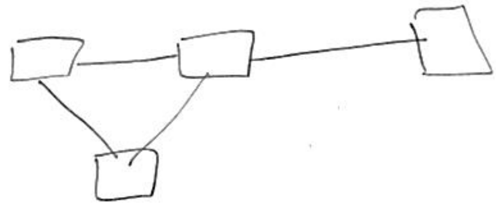
Marcus A M de Aguiar
6/2015

I- Conceitos Gerais

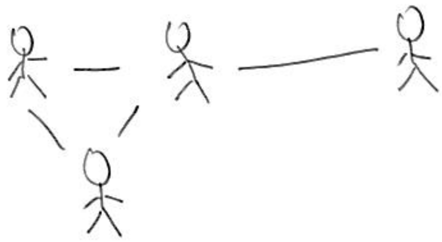
Sistemas complexos são definidos por terem muitos componentes que interagem de forma heterogênea. Alguns componentes interagem com vários outros enquanto que outros interagem com apenas um ou dois outros elementos do sistema. As redes de interação formadas pelos componentes são representadas por grafos.

Exemplos:

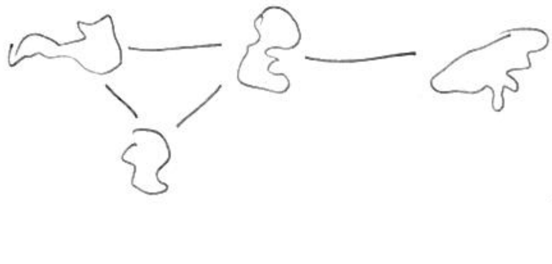
a) Computadores conectados



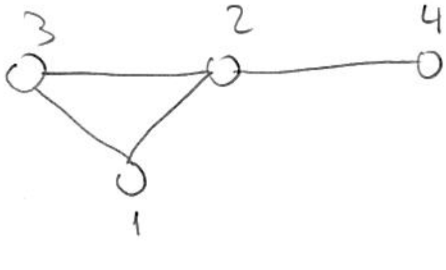
b) Rede social



c) Proteínas que interagem



GRAFO:



QUASE TUDO ESSE MATERIAL É RETIRADO DO LIVRO online de A.L. Barabasi barabasi.com/networksciencebook

Em ecologia, o modelo de Lotka-Volterra para um predador e uma presa é

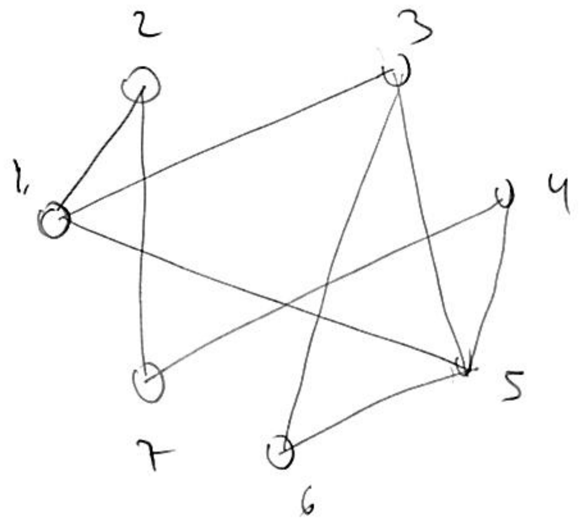
$$\frac{dx}{dt} = rx - axy \quad \rightarrow \text{presa}$$

$$\frac{dy}{dt} = -dy + bxy \quad \rightarrow \text{predador}$$

pode ser generalizado p/ N espécies na forma

$$\frac{dx_i}{dt} = b_i x_i + \sum_{j=1}^{N} a_{ij} x_i x_j$$

e pode ser representado como uma rede:



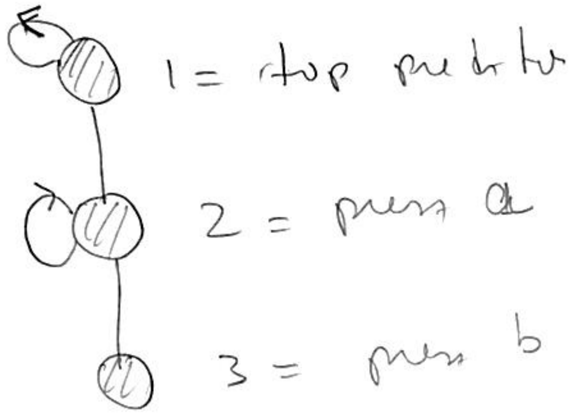
Se $a_{ij} > 0$ a espécie i se beneficia da interação.

Se $a_{ij} >$ e $a_{ji} > 0$ então i e j tem interação mutualista

Se $a_{ij} > 0$ e $a_j < 0$, i = predador j = presa

Exemplo $N = 3$

$a_{11} = -a$	$a_{21} = -d$	$a_{31} = 0$
$a_{12} = b$	$a_{22} = -c$	$a_{32} = -f$
$a_{13} = 0$	$a_{23} = e$	$a_{33} = 0$

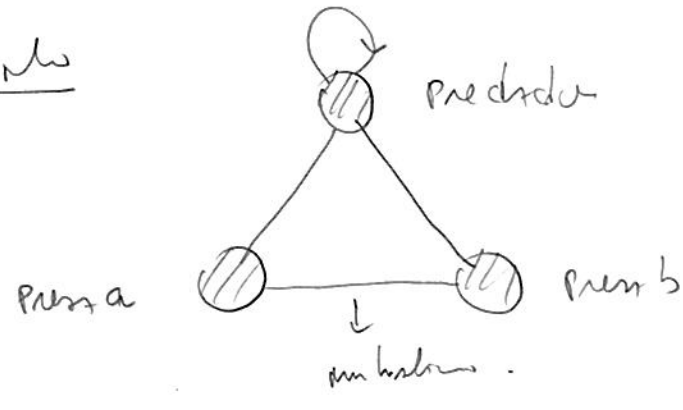


$$\frac{dx_1}{dt} = -b_1 x_1 - a x_1^2 + b x_1 x_2$$

$$\frac{dx_2}{dt} = b_2 x_2 - c x_2^2 - d x_1 x_2 + e x_2 x_3$$

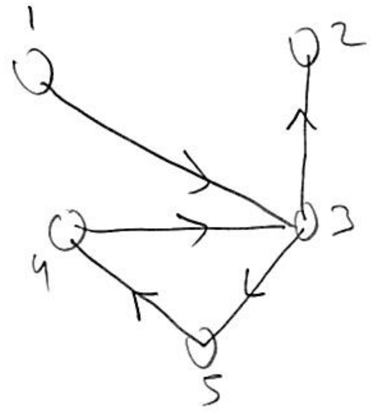
$$\frac{dx_3}{dt} = b_3 x_3 - f x_2 x_3$$

Exemplo



TIPOS DE REDE

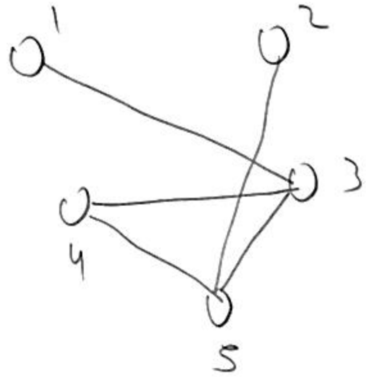
DIRECIONADA



NODO 1 influencia 3
MAS 3 NÃO influencia 1

Ex. paper 3 cita paper 1
que é mais antigo..
(WWW web pages)

NÃO-DIRECIONADA



NODO 1 influencia NODO 3
NODO 3 " NODO 1
(como em Lotka-Volterra)
ou internet

- sem peso: todas conexões tem peso 1
- com peso: LINK 1-3 é mais intenso que LINK 4-3.

NOTAÇÃO

N = número de nós da rede

L = " de links da rede

Nós são enumerados 1, 2, ... N

k_i = nº de links do nó i = grau do nó

veja que $L = \frac{1}{2}\sum_{i=1}^{N} k_i$

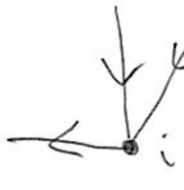
Grau médio

$$\langle K \rangle = \frac{1}{N} \sum_{i=1}^{N} k_i = \frac{2L}{N}$$

$$\langle K^2 \rangle = \frac{1}{N} \sum_{i=1}^{N} k_i^2$$

$$\sigma_K = \sqrt{\langle K^2 \rangle - \langle K \rangle}$$

Para redes direcionadas



$K^{in} = 2$

$K^{out} = 1$

$$K_i = K_i^{in} + K_i^{out}$$

$$L = \sum_{i=1}^{N} K_i^{in} = \sum_{i=1}^{N} K_i^{out}$$

$$\langle K_i^{in} \rangle = \frac{1}{N} \sum_{i=1}^{N} K_i^{in} = \frac{1}{N} \sum_{i=1}^{N} K_i^{out} = \langle K_i^{out} \rangle = \frac{L}{N}$$

Distribuição de Grau

$P_K \equiv$ probab. de um nó sorteado ao acaso ter grau K

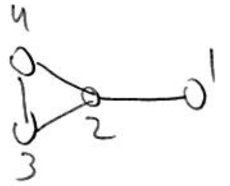
$$\sum_{K=1}^{\infty} P_K = 1$$

$N_K \equiv$ número de nós com grau K

$$P_K = \frac{N_K}{N}$$

$$\langle K \rangle = \frac{1}{N} \sum_{i=1}^{N} k_i = \frac{1}{N} \left[1N_1 + 2N_2 \ldots \right] = \frac{\sum K N_K}{N} = \sum K P_K$$

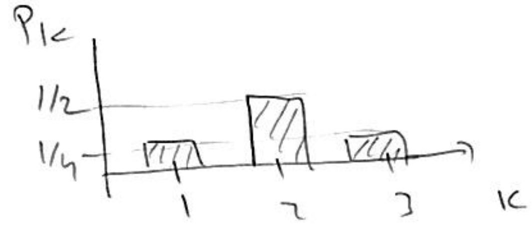
Exemplos

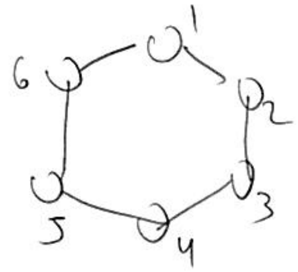


$K_1 = 1$ $\quad P_1 = 1/4$

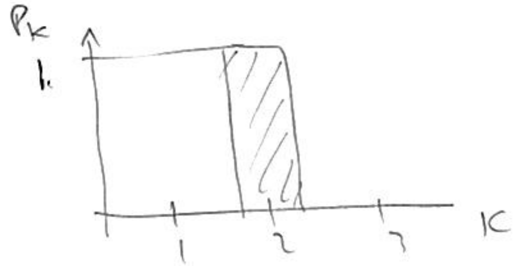
$K_2 = 3$ $\quad P_2 = 2/4$

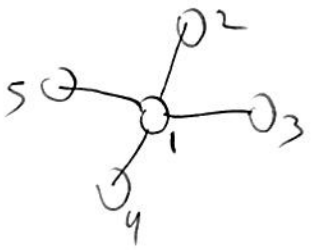
$K_3 = K_4 = 2$ $\quad P_3 = 1/4$





$K = 2$



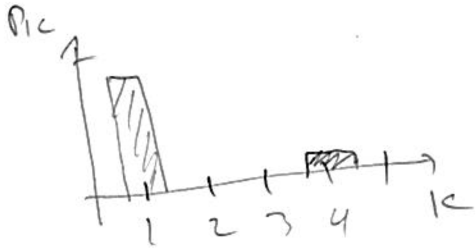


$K_1 = 4$

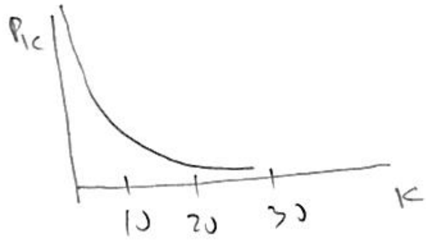
$K_2 = K_3 = K_4 = K_5 = 1$

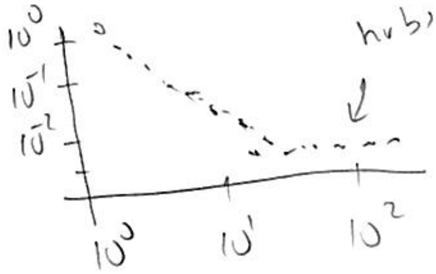
$P_1 = \frac{4}{5}$

$P_4 = \frac{1}{5}$



Rede de proteínas





MATRIZ DE ADJACÊNCIAS (redes não-direcionadas, sem peso)

$$A_{ij} = \begin{cases} 1 & \text{se } i \longleftrightarrow j \text{ conectados} \\ 0 & \text{se } i \longleftrightarrow j \text{ não conectados} \end{cases}$$

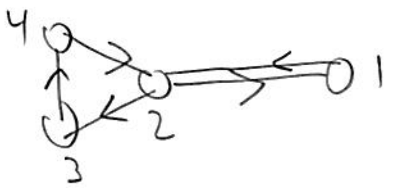
Em geral $A_{ii} = 0$, mas nem sempre. Veja Lotka-Volterra.

Para os exemplos anteriores

$$\begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 \end{pmatrix} ; \quad \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 0 & 1 & 0 \end{pmatrix} ; \quad \begin{pmatrix} 0 & 1 & 1 & 1 & 1 \\ 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \end{pmatrix}$$

$$K_i = \sum_{j=1}^{N} A_{ij} = \sum_{j=1}^{N} A_{ji}$$

REDES DIRECIONADAS

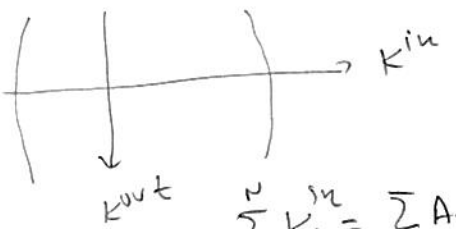


$$\begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}$$

NA LINHA i ENTRA QUEM INFLUÊNCIA O NÓ i

$$K_i^{in} = \sum_{j=1}^{N} A_{ij}$$

$$K_i^{out} = \sum_{j=1}^{N} A_{ji}$$



$$\sum_{i=1}^{N} K_i^{in} = \sum_{i,j} A_{ij} = L$$

$$2L = \sum_i K_i = \sum_{i,j} A_{ij}$$

Em geral, para redes reais $L << L_{max} = \frac{N(N-1)}{2}$

Exemplo WWW tem 1.5×10^6 links

$N \simeq 300.000 = 3 \times 10^5$

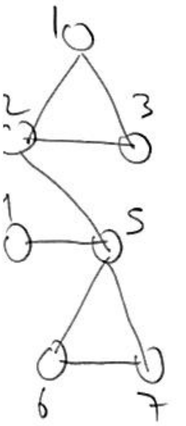
$L_{max} = 10^{10}$

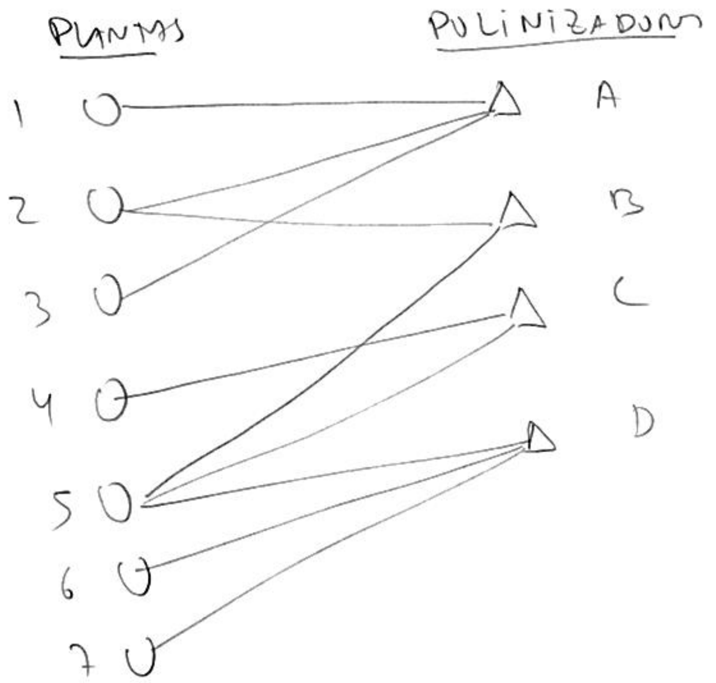
$L / L_{max} \simeq 10^{-4}$

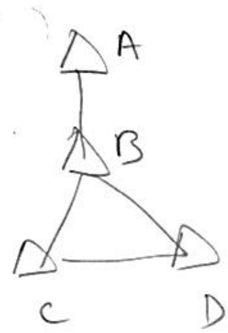
REDES COM PESOS

$A_{ij} \longrightarrow W_{ij}$ = número real, como em L-V.

REDES BIPARTIDAS

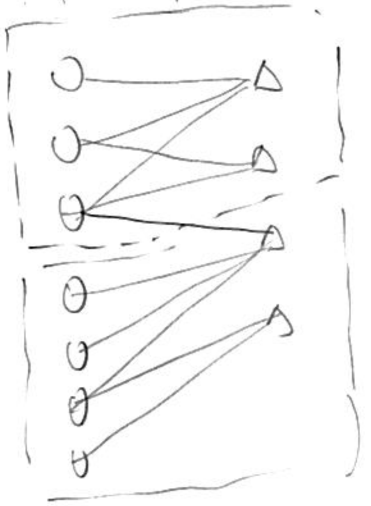




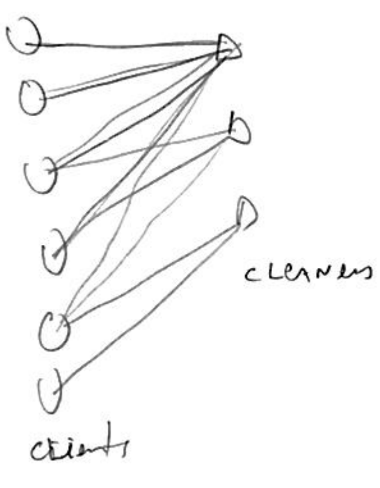


ECOLOGIA: redes de mutualismo

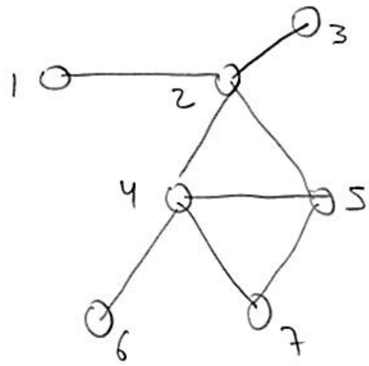
MODULARIDADE (interações que co-evoluíram)



NESTED (ANINHADAS)



CAMINHOS ENTRE NÓS



CAMINHOS entre (6) ⟷ (1)

a) 6 – 4 – 5 – 2 – 1

b) 6 – 4 – 7 – 5 – 2 – 1

c) 6 – 4 – 2 – 1

CAMINHO mínimo entre 1 e 6 tem 3 passos:

$$d_{16} = 3$$

– CAMINHO MÍNIMO MÉDIO

$$d = \frac{1}{N(N-1)} \sum_{i,j} d_{ij}$$

– DIÂMETRO DA REDE

$$d_{max} = \text{maior dos } d_{ij}$$

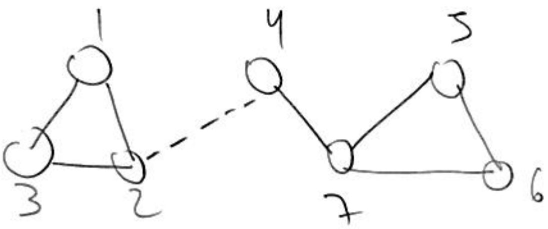
– NÚMERO de CAMINHOS de comprimento 1 entre $i \leftrightarrow j$ é

$$N_{ij}^{(1)} = A_{ij} \quad (\text{ou } 1 \text{ ou } 0)$$

– NÚMERO de caminhos de comprimento 2 é

$$N_{ij}^{(2)} = \sum_{k=1}^{N} A_{ik} A_{kj} = \left(A^2\right)_{ij}$$

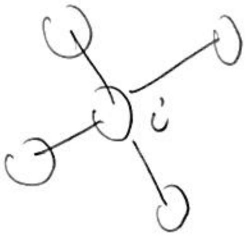
COMPONENTES DE UMA REDE



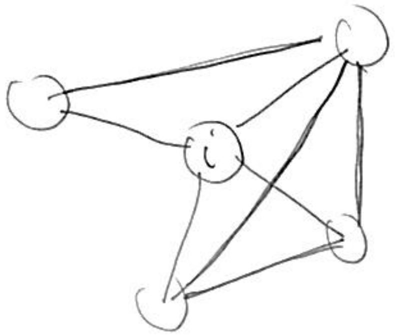
$$
\left(\begin{array}{ccc|cccc}
0 & 1 & 1 & 0 & 0 & 0 & 0 \\
1 & 0 & 1 & \boxed{0} & 0 & 0 & 0 \\
1 & 1 & 0 & 0 & 0 & 0 & 0 \\
\hline
0 & \boxed{0} & 0 & 0 & 0 & 0 & 1 \\
0 & 0 & 0 & 0 & 0 & 1 & 1 \\
0 & 0 & 0 & 0 & 1 & 0 & 1 \\
0 & 0 & 0 & 1 & 1 & 1 & 0
\end{array}\right)
$$

AGREGAMENTO (CLUSTERING)

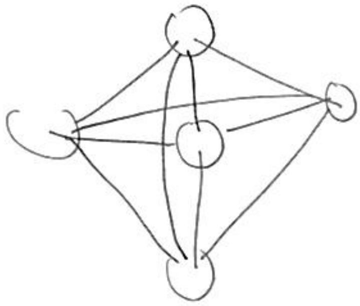
"QUANTOS DOS MEUS AMIGOS SÃO AMIGOS entre si"



$C_i = 0$



$C_i = \frac{4}{6}$



$C_i = 1$

$$C_i = \frac{2L_i}{k_i(k_i-1)}$$ onde

L_i = número de links entre os vizinhos de i

$\frac{k_i(k_i-1)}{2}$ = número máximo de links possíveis.

II – A rede Aleatória

Def. Uma rede aleatória consiste de N nós onde cada par é conectado com probabilidade p : $G(N,p)$

Como construir:

- comece com N nós isolados
- selecione um par de nós e x = número aleatório $\in (0,1)$
  se $x > p$ conecte os nós, senão não conecte.
- faça isso para todos os pares de nós.

Nomenclatura: GRAFO ALEATÓRIO, REDE ALEATÓRIA, rede de Erdös-Rényi

PROPRIEDADES:

(i) Probabilidade que a rede tenha exatamente L links

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

= DISTRIBUIÇÃO BINOMIAL.

$$\langle L \rangle = \text{número médio de links é} = \sum_{L=1}^{\infty} L P_L$$

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

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

BOX: DISTRIBUIÇÃO BINOMIAL E DE POISSON

$$P_K = \binom{N}{K} p^K (1-p)^{N-K} = \frac{N!}{(N-K)!\,K!} p^K (1-p)^{N-K}$$

$$\sum_{K=0}^{N} P_K = 1$$

$$\langle K \rangle = \sum_{K=1}^{N} K P_K = \sum_{K=1}^{N} \frac{N!}{(N-K)!\,(K-1)!} p^K (1-p)^{N-K}$$

$$= pN \sum_{K=1}^{N} \frac{(N-1)!}{((N-1)-(K-1))!\,(K-1)!} p^{K-1} (1-p)^{(N-1)-(K-1)}$$

$$= pN \sum_{K=0}^{N-1} \binom{N-1}{K} p^K (1-p)^{(N-1)-K} = pN$$

DA mesma forma mostramos que

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

$$\sigma^2 = \langle K^2 \rangle - \langle K \rangle^2 = p(1-p)N$$

PARA $N \to \infty$

$$P_K = \frac{N(N-1)\cdots(N-K+1)}{K!} \left(\frac{\langle K \rangle}{N}\right)^K \left(1 - \frac{\langle K \rangle}{N}\right)^{N-K}$$

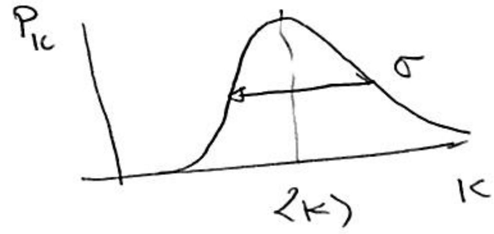
$$\simeq \frac{N^K}{K!} \frac{\langle K \rangle^K}{N^K} \left(1 - \frac{\langle K \rangle}{N}\right)^N = \frac{\langle K \rangle^K}{K!} e^{-\langle K \rangle}$$

= Distrib de Poisson

Pico ocorre em $\langle K \rangle$ e $\sigma^2 = \langle K \rangle$.

(ii) Probabilidade que o nó i tenha exatos k links

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



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

$$\sigma^2 = p(1-p)(N-1)$$

Para N grande

$$P_k \simeq e^{-\langle k \rangle} \frac{\langle k \rangle^k}{k!}$$

$$\sigma = \sqrt{\langle k \rangle} \sim \sqrt{N}$$

$$\frac{\sigma}{N} = \frac{1}{\sqrt{N}} \to 0$$

<u>Exemplo</u> humanos, $N = 7 \times 10^9$, $\langle k \rangle = 1000$

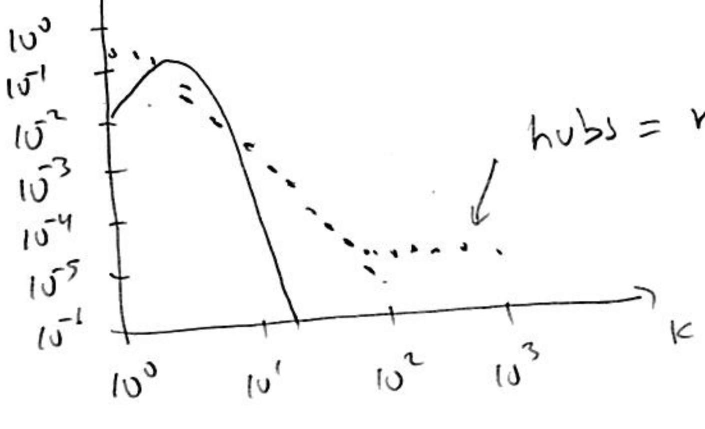
$\sigma \simeq 32 \Rightarrow 3\sigma \simeq 100$

indivíduos mais conectados ~1100 amigos — <u>ERRADO</u>

" menos " ~900 amigos

$\boxed{\text{Redes reais não são aleatórias}}$

tipicamente



hubs = nós altamente conectados

TRANSIÇÕES DE FASE

Podemos pensar nas redes aleatórias $G(N,p)$ como sendo função do parâmetro p. Se N é muito grande podemos imaginar que diferentes redes geradas com o mesmo p devam ter propriedades bem similares.

Sabemos que:

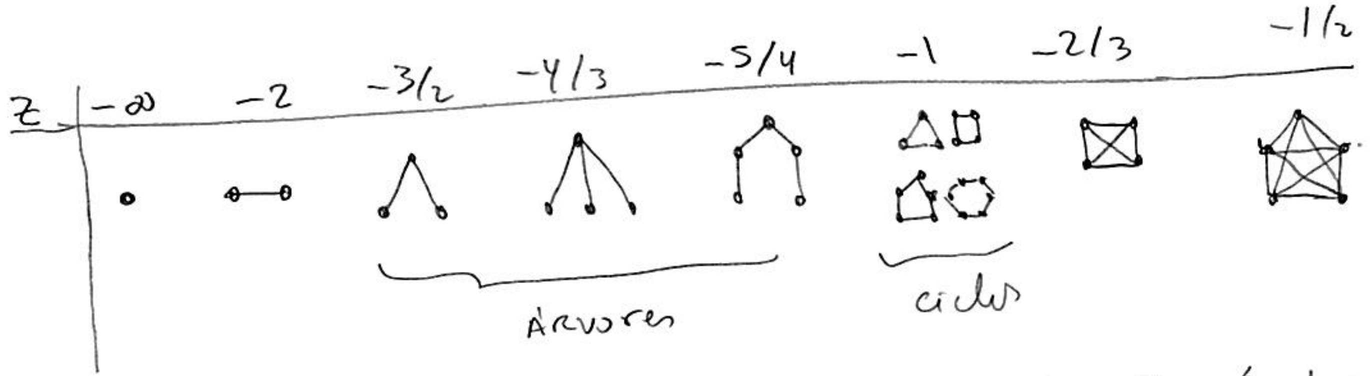
a) $p=0 \Rightarrow$ nenhum nó está conectado

b) $p=1 \Rightarrow$ todos os nós se conectam a todos os outros.

Vamos tentar entender o que acontece conforme p aumenta de zero até um. Escrevendo

$$p = N^{z}$$

vamos mostrar que as seguintes sub-redes aparecem abruptamente conforme p passa dos limites indicados:



Uma sub-rede $\tilde{R}(\tilde{K},\tilde{l})$ de uma rede R é tal que todos os $\tilde{K}$ nós de $\tilde{R}$ são nós de R e todos os $\tilde{l}$ links de $\tilde{R}$ também são links de R.

RIS

Queremos determinar qual é a probabilidade crítica $p_c(N)$ tal que quase toda rede com N nós e $p > p_c(N)$ tenha, por exemplo, tenha uma árvore de ordem 3.

$$\text{Árvore de ordem } K \equiv \begin{cases} K \text{ nós} \\ K-1 \text{ links} \end{cases} \quad \text{(nenhum ciclo)}$$

Seja então um grafo aleatório $G(N,p)$ e uma sub-rede F_{kl} com K nós e l links. Vamos primeiro determinar quantas sub-redes F_{kl} existem em $G(N,p)$ se N é grande:

X = número de sub-redes F_{kl} em $G(N,p)$

$E(X)$ = " médio de subredes F_{kl}

$$E(X) = \binom{N}{K} p^l \frac{K!}{a}$$

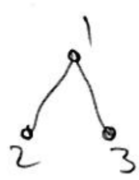
$\binom{N}{K}$: # de maneiras de escolher K entre N

p^l: probabilidade de l links

$K!$: permutações dos K nós

a: nº de grafos isomórficos

<u>NOTA</u>: No caso de uma árvore com $K=3$



e

representam a mesma subrede

$\Rightarrow a = 2$

Para distribuir l links dentre K nós temos

$$\binom{K}{l} = \frac{K!}{l!(K-l)!} = \frac{K!}{a(K,l)}$$

Para N grande

$$E(X) \simeq \frac{N!}{K!\,(N-K)!}\,\frac{K!}{a}\,p^{e} = \frac{N!}{(N-K)!}\,\frac{p^{e}}{a}$$

$$\approx N^{K}\,\frac{p^{e}}{a}$$

i) Se $p(N)$ for tal que $\lim_{N\to\infty} N^{K}\frac{p(N)^{e}}{a} \to 0$

então a subrede F NÃO será encontrada

ii) Se $p(N) = c\,N^{-K/e}$, $E(X) \simeq c^{e}/a \equiv \lambda$

iii) Vamos supor que as subredes F tenham uma distribuição de Poisson, de forma que

$$P(X=r) = e^{-\lambda}\frac{\lambda^{r}}{r!}$$ = prob. que uma rede $G(N,p)$ tenha r subredes F

- A prob. de NÃO ter nenhum é
$$P(X=0) = e^{-\lambda}$$

- A prob. de ter pelo menos um é
$$P = 1 - e^{-\lambda}$$

- Se aumentamos p de tal forma que

$$\frac{p^{\ell} N^{k}}{a} \to \infty \quad , \quad \lambda \to \infty \quad e \quad P = 1$$

Portanto $p(N) = c\, N^{-\ell/k}$ é de fato a probabilidade crítica p/ uma subrede $F(k, \ell)$ aparecer.

Em outras palavras, para

$$p(N) = c\, N^{-k/\ell + \varepsilon}$$

$$E(X) = N^{k} \frac{c^{\ell}}{a} N^{-k + \ell\varepsilon} = \frac{c^{\ell}}{a} N^{+\varepsilon \ell} \begin{cases} \to 0 & \text{se } \varepsilon < 0 \\ \to \infty & \text{se } \varepsilon > 0 \end{cases}$$

<u>Exemplos</u>

a) Se $p(N) < \frac{1}{N^{2}} \Rightarrow$ só pontos desconexos

b) Se $p(N) > \frac{1}{N^{2}} \Rightarrow$  $(k=2, \ell=1)$

c) Se $p(N) > \frac{1}{N^{3/2}} \Rightarrow$  $(k=3, \ell=2)$

d) Se $p(N) > \frac{1}{N^{4/3}} \Rightarrow$  ,  $(k=4, \ell=3)$

e) Se $p(N) > \frac{1}{N} \Rightarrow$  etc $k = \ell$, todos os ciclos aparecem.

f) Para sub-redes totalmente conectados de ordem K, $\ell = \frac{K(K-1)}{2}$

$$P(N) = c\,N^{-2/(K-1)}$$

$\underline{K=5}$ 

RESULTADOS GERAIS

I - Regime Sub-crítico ; $p < \frac{1}{N} \Rightarrow \langle k \rangle < 1$

- Rede tem pequenos aglomerados isolados
- tamanho do maior aglomerado $N_G \simeq \ln N$ e $\lim_{N\to\infty} \frac{N_G}{N} \to 0$.

  ↳ "AGLOMERADO GIGANTE"

II - Ponto crítico ; $p = \frac{1}{N} \Rightarrow \langle k \rangle = 1$

- $N_G \simeq N^{2/3} \Rightarrow \lim_{N\to\infty} \frac{N_G}{N} \to 0$

Para $N = 7 \times 10^9$:

sub-crítico , $N_G \simeq 22.7$

crítico , $N_G \simeq 3 \times 10^6$

} salto gigantesco embora o aglomerado gigante ainda não pegue toda a rede.

III - Regime Supercrítico

$p > \frac{1}{N}$, $\langle k \rangle > 1$.

$$N_G \simeq \langle k \rangle - 1 \quad \text{ou} \quad N_G = (p - p_c)N$$

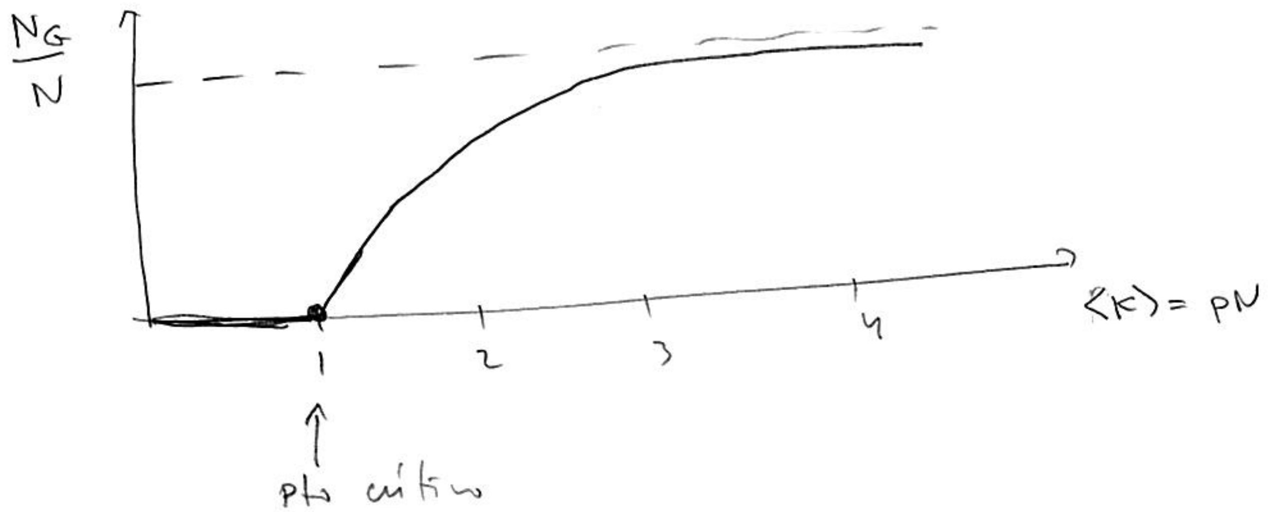
$$p_c \equiv \frac{1}{N} \; ; \quad p = \frac{\langle k \rangle}{N}$$

IV - Connected Regime

$p > \frac{\ln N}{N}$; $\langle k \rangle > \ln N$

A maioria dos nós faz parte da componente gigante

se $p = \frac{N-1}{N}$, $\langle k \rangle = N-1$ → rede é totalmente conect.



DIÂMETRO

- se $p > \frac{1}{N}$, $d = \frac{\ln(N)}{\ln(pN)} = \frac{\ln(N)}{\ln\langle k \rangle}$ = diâmetro do Aglom. gigante

CLUSTERING

$$\langle C \rangle = p = \frac{\langle k \rangle}{N}$$

$L_{max} \simeq \frac{\langle k \rangle (\langle k \rangle - 1)}{2}$

$L_i = p \cdot$ (# de conexões possíveis)

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