

Redes Complexas: Internet, Web, algoritmos e outras aplicações em computação

-Bloco #4

Modelo “Small World”
2º semestre de 2008

Virgílio A. F. Almeida
Setembro de 2008



Departamento de Ciência da Computação
Universidade Federal de Minas Gerais

Tópicos

1. Experimento Milgram ✓
2. Watts & Strogatz: modelo “small world”
3. Kleinberg: modelo “small world”
4. Modelos de redes SW: exemplos

Por que estudar o fenômeno de *small world*?

1. Curiosidade:

- Por que world small (o mundo é pequeno)?
- Como as pessoas são capazes de rotear mensagens?
informações

2. Social Networking como negócio:

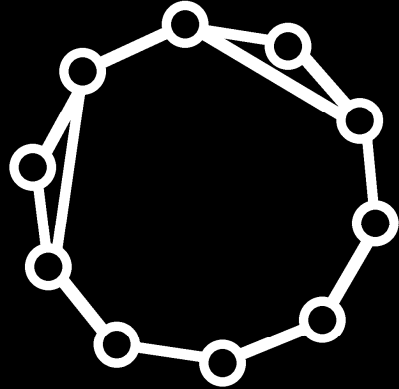
- Friendster, Orkut, MySpace
- LinkedIn, Spoke, VisiblePath

3. Aplicações

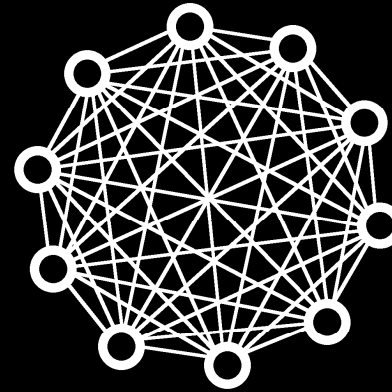
- Roteamento em redes sem fio, sensores, etc
- Buscas de pessoas, objetos, etc...
- Estruturas de topologias de rede (ex. P2P ...)

4.

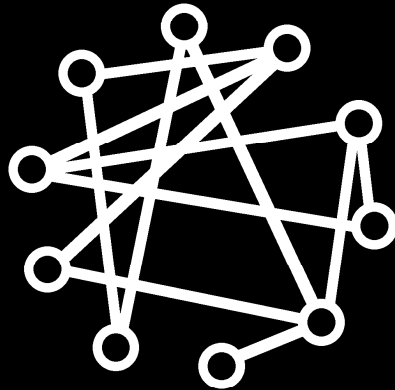
Tipos de Redes



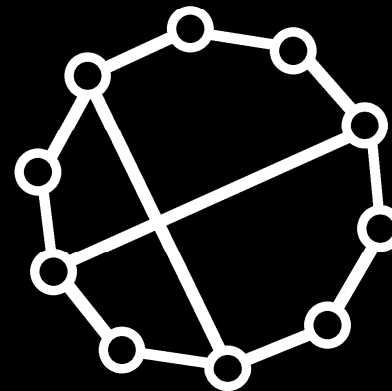
Regular Lattice



Fully Connected



Random

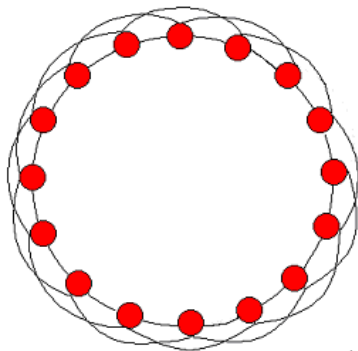


Small World

O Modelo de small world

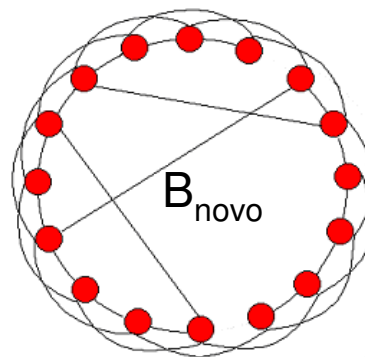
Alto grau de agrupamento (*high CC*): amigos de meus amigos tendem ser meus amigos... (e inimigos?)

Watts & Strogatz (1998) – uns poucos links randomicos num **grafo denso** provem caminho mínimo médio próximo daquele de um random graph (RG).

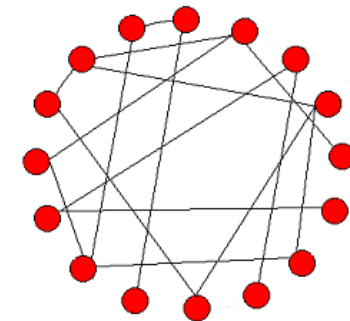


order

$p = 0$



$0 < p < 1$



randomness

$p = 1$

Redes na natureza (observações empíricas)

- rede neural de *C. elegans*,
- rede semântica e rede de linguagem
- grafo de colaboração de atores,
- food webs.

caminho mínimo
médio



$$l_{\text{network}} \approx \ln(N)$$



Olhar Newman MEJ,
The structure and
function of complex
networks, pag 11.

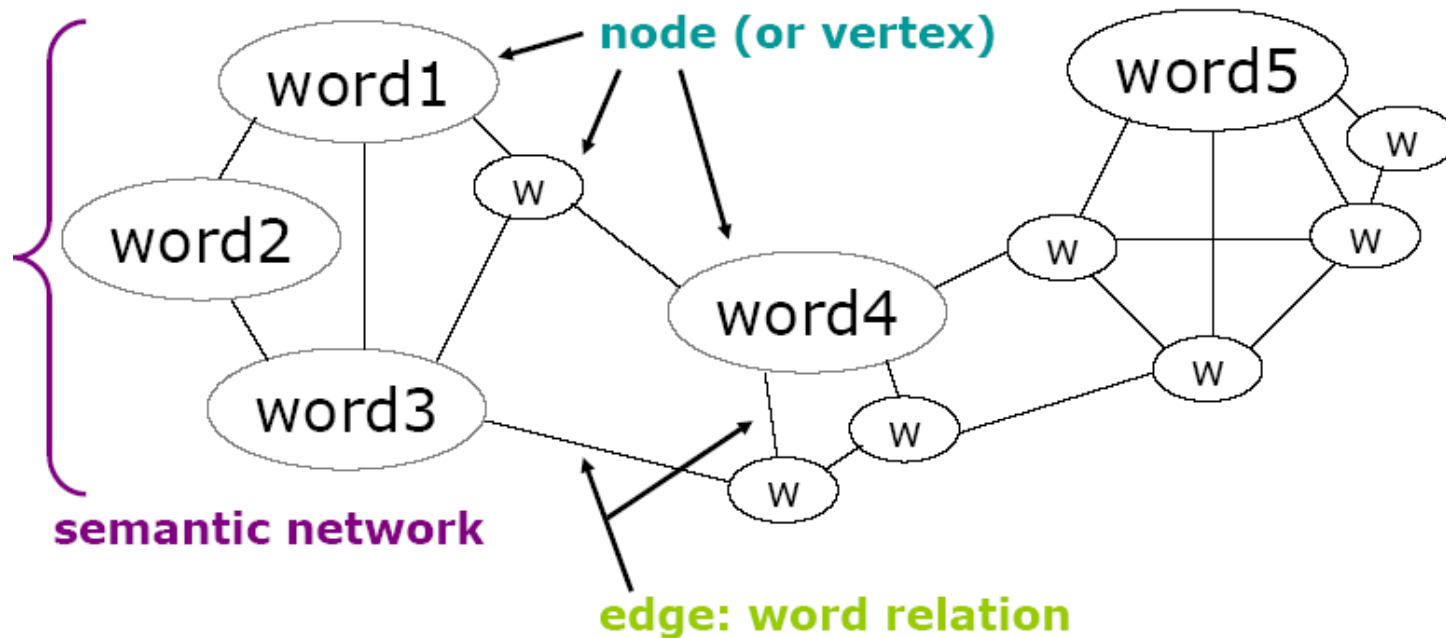
$$C_{\text{network}} \gg C_{\text{random graph}}$$




coeficiente de
clustering

Redes Semanticas

words & their relations → node & edge → semantic network



-  structures in large-scale word association data (Steyvers, 2003)
 1. combination of highly clustered neighborhoods
 2. a short average path length (3.03 in the undirected associative network of Nelson et al, 5.43 in Roget's thesaurus, 10.61 in WordNet)

Small world
of Human Language
Ramon Cancho
Ricard V. Sole

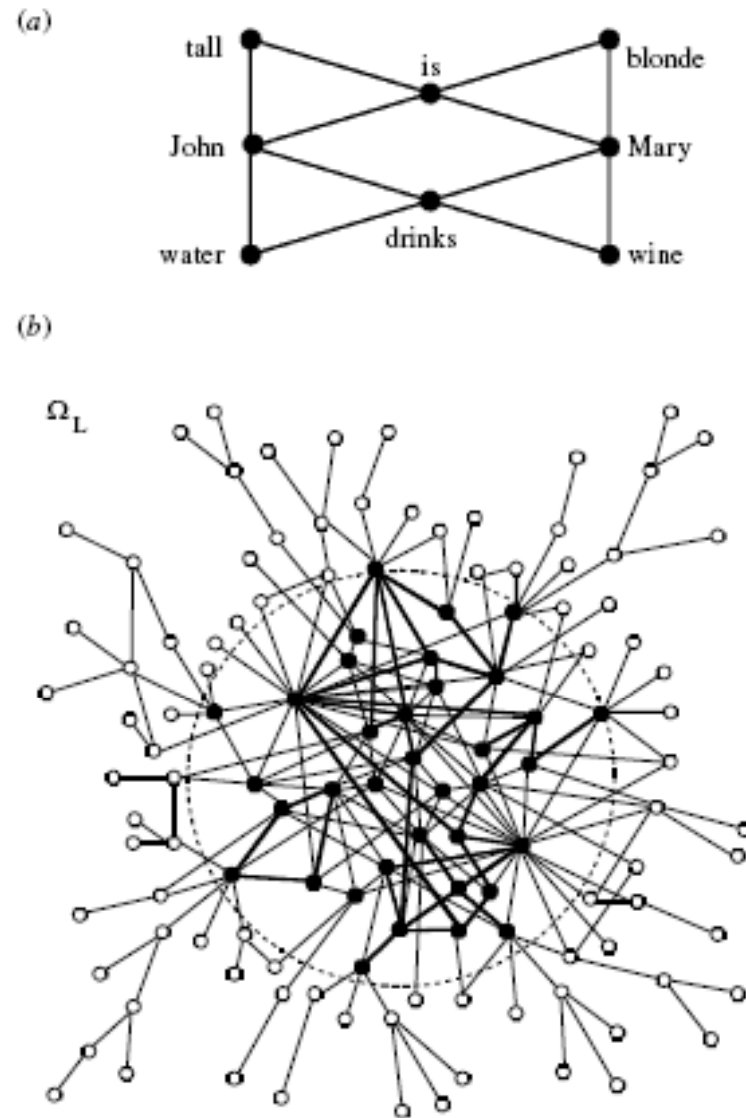


Figure 1. Word networks. (a) A toy network constructed with four sentences: 'John is tall,' 'John drinks water,' 'Mary is blonde,' 'Mary drinks wine.' (b) A possible pattern of wiring in Ω_L . Black nodes are common words and white nodes are rare words. Two words are linked if they co-occur significantly.

The Small World of Human Language

The small world of human language R. Ferrer i Cancho and R. V. Solé 2263

connections Table 1. Word network patterns.

(2.1) (It can be seen that $C \gg C_{\text{random}}$ and $d \approx d_{\text{random}}$, consistently in a SW network. All values are exact except for those marked with an asterisk, which have been estimated on a random subset of the vertices (after having processed 2% of the vertices, fluctuations in d^* as a function of the subset size clearly affected only the third decimal digit).)

graph	C	C_{random}	d	d_{random}
Ω_L (UWN)	0.687	1.55×10^{-4}	2.63*	3.03
Ω_L (RWN)	0.437	1.55×10^{-4}	2.67*	3.06

verage over

Bristish National Corpus

3. SCALING AND SMALL-WORLD PATTERNS

(2.2) The biggest connected component of the networks that results from the basic and improved methods will be called, respectively, the unrestricted word network (UWN) and the restricted word network (RWN). They have $N(\text{UWN}) = 478\,773$ and $N(\text{RWN}) = 460\,902$ nodes, with $E(\text{UWN}) = 1.77 \times 10^7$ and $E(\text{RWN}) = 1.61 \times 10^7$ edges, respectively. With average connectivities of $\bar{k}_{\text{uwn}} = 74, 2$ and $\bar{k}_{\text{rwn}} = 70, 13$, their clustering and path lengths are indicated in Table 1.

' neighbours
r.
1 two words
path length
verage path

Figure 2 shows the distribution of degree

<http://complex.upf.es/~ricard/>

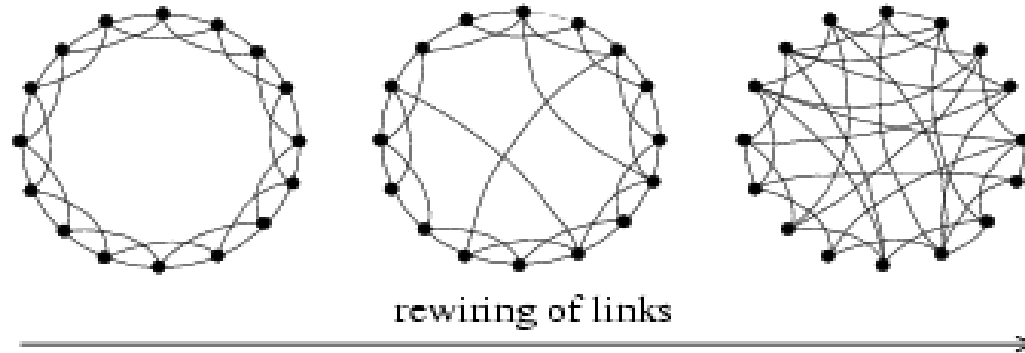
Modelo Proposto

1. Intermediário entre grafos regulares e grafos aleatórios
2. Adjustável (*tunable*)
3. Rede Small World com simultaneamente as seguintes características:
 1. Pequeno caminho mínimo médio
 2. Alto *clustering coefficient* (não é próprio dos RG)

Hipóteses do Modelo SW

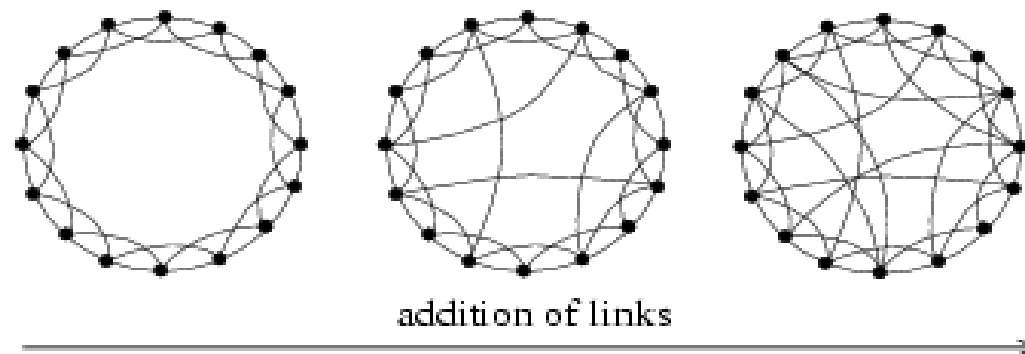
1. Arestas não orientadas
2. Grafo não ponderado: nenhuma anotação associada às arestas
3. Esparso:- número de arestas \ll número total de arestas de um grafo completo
4. Conectado: todos vértices são conectados (não há clusters isolados)

Duas formas de construir um grafo *small world*



a)

1. Selecione uma fração p de arestas,
2. Reposicione suas extremidades



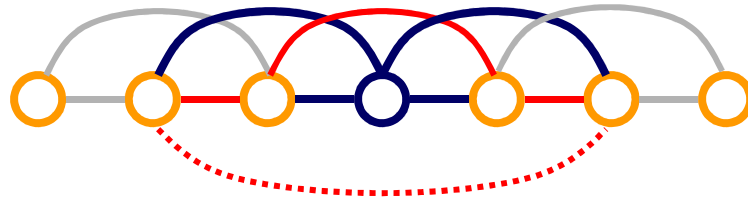
b)

1. Adicione uma fração p de aresta deixando o grafo regular subjacente inalterado.

1. Como em muitos algoritmos de geração de redes
 1. Não se permite arestas auto-referentes (self-loops)
 2. Não se permite múltiplas arestas (grafos ponderados)
2. A idéia é a inclusão dos “shortcuts”(arestas)

Modelo Original

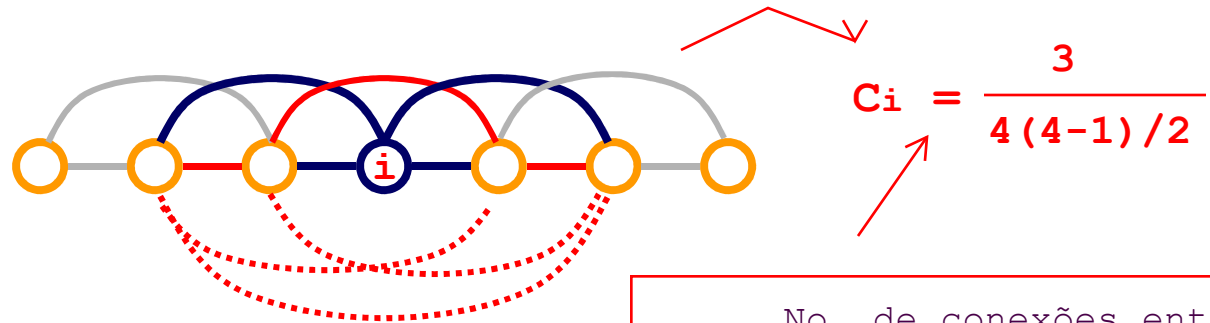
1. Cada nó tem $K \geq 4$ vizinhos mais próximos (locais)
2. Probabilidade p de “rewiring” vértices escolhidos aleatoriamente
3. p pequeno: malha regular (grafo regular)
4. p alto: grafo aleatório clássico



Grafo Regular (p=0 - lattice)

1. Calcule o clustering coefficient:

1. Cada nodo é conectado a K vizinhos, que podem ter $K*(K-1)/2$ conexões par-a-par entre eles.
2. Algumas conexões entre eles estão presentes no lattice (i.e., treliça)



Se $K = 4$ (conectado a 2 vizinhos de cada lado)

$$C = \frac{3*2}{4*3} = \frac{1}{2} \longrightarrow \text{m\u00e9dia dos } C_i, \text{ onde}$$

$$C_i = \frac{\text{No. de conex\u00f5es entre os vizinhos do n\u00f3 } i \text{ (arestas vermelhas)}}{\text{No. de **poss\u00edveis** conex\u00f5es entre os vizinhos do n\u00f3 } i}}$$

Obs: algumas vezes o lattice pode ser especificado como:

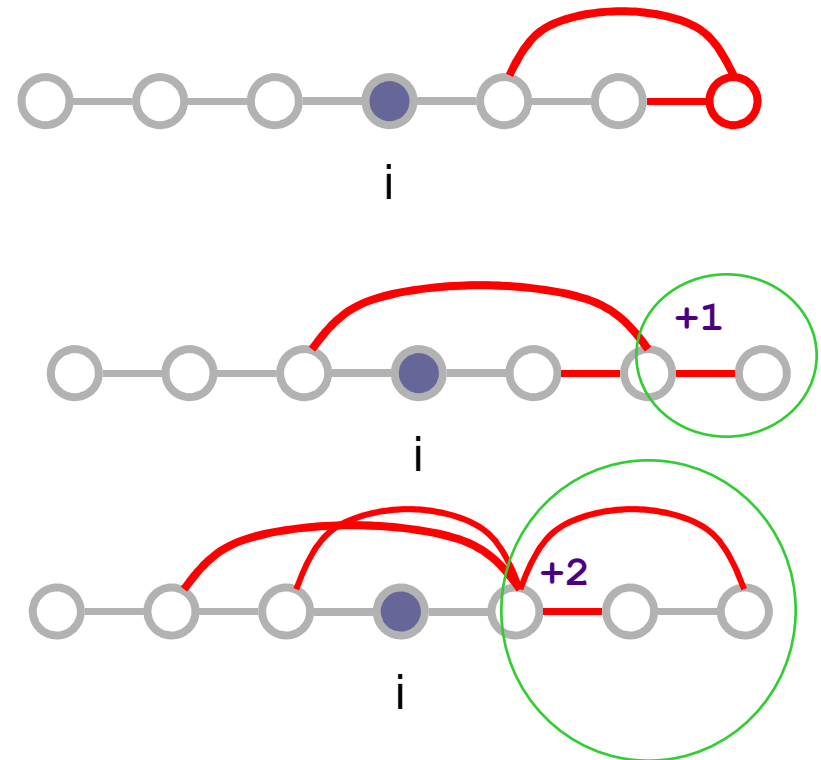
cada nodo conecta a K vizinhos mais pr\u00f3ximos

cada nodo conecta todos vizinhos mais pr\u00f3ximos numa dist\u00e2ncia k ($k = K/2$)

Clustering coefficient para grafo regular

$$K = 6$$

1. Em geral, pode ter qq K
2. Um vizinho que está $K/2$ hops away de i pode se conectar a $(K/2 - 1)$ dos vizinhos de i
3. Um vizinho que está $K/2 - 1$ hops away pode conectar a $(1 + K/2 - 1)$ vizinhos
4. $K/2 - 2$ hops away
 1. $(2 + K/2 - 1)$ vizinhos
5. 1 hop away
 1. $2*(K/2 - 1)$
6. Somando
 1. Multiplique por 2 porque i tem vizinhos dos dois lados
 2. divida por 2 porque arestas são não direcionadas.



Olhar material complementar para essa aula na wiki

Clustering coefficient para grafo regular

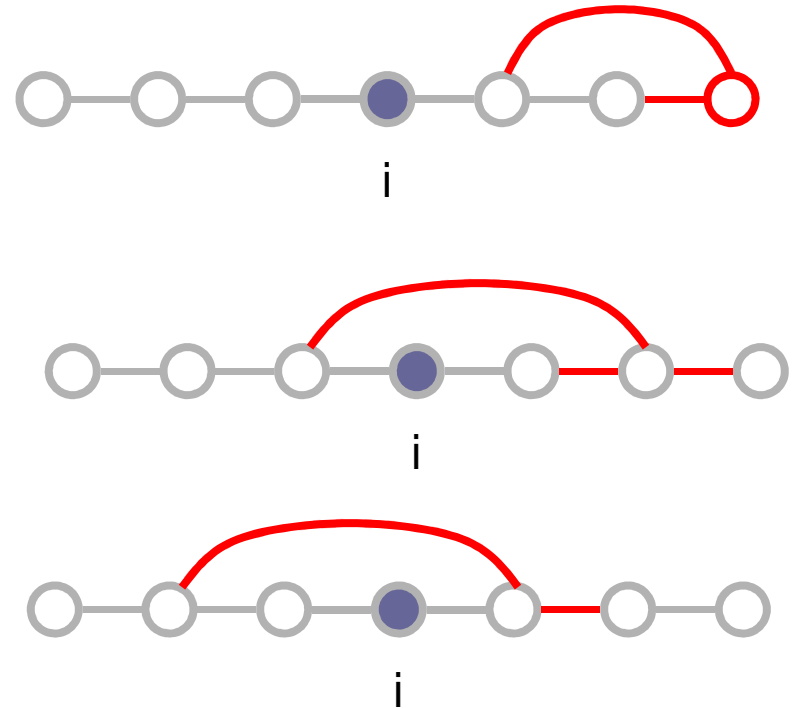
1. O número de conexões entre vizinhos é dado por:

$$\sum_{j=0}^{\frac{K-1}{2}} \left(\frac{K}{2} + j - 1 \right) = \frac{3}{8} K(K-2)$$

1. O máximo número de conexões é dado por: $K*(K-1)/2$

2. → clustering coefficient é

$$C = \frac{3(K-2)}{4(K-1)}$$



Olhar material complementar para essa aula na wiki

Caminho Mínimo Médio

(Average shortest path) - grafo regular

↙ ???

1. Nodo médio está $N/4$ hops away (um quarto de distancia no anel), e pode-se “pular” sobre $K/2$ nodos de cada vez:

$$l \approx \frac{N}{2K} \gg 1$$

Olhar Watts & Strogatz (1998), pag 440.

Grafo Randômico (p=1)

1. Observações

$$l \approx \frac{\ln N}{\ln K} \quad \text{pequeno}$$

$$C \approx \frac{K}{N} \quad \text{pequeno}$$

Watts & Strogatz (1998). Collective dynamics of 'small-world' networks. Nature 393: 440-442.

Existe uma média de K arestas por nodo.

A probabilidade que dois nodos quaisquer estejam conectados é $p = K/N$.

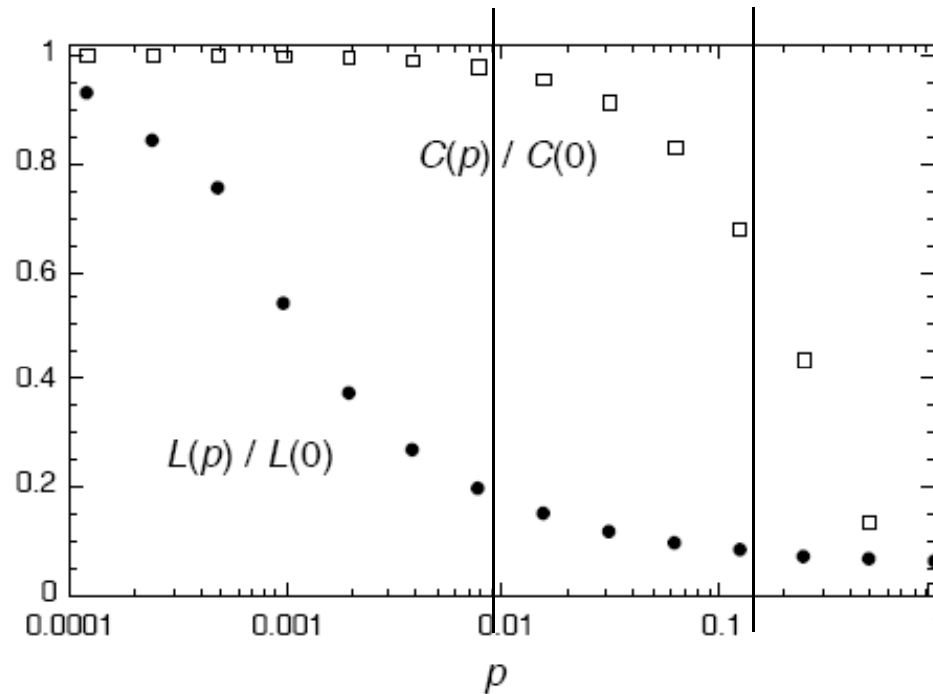
A probabilidade que dois nodos que compartilham um nodo comum são conectados entre si é a mesma que dois nodos randômicos: K/N

(na verdade $(K-1)/N$ por eles já tem uma aresta alocada na vizinhança comum).

O que acontece no meio do caminho?

1. Caminho mínimo pequeno significa baixo clustering?
2. Caminho mínimo grande significa alto clustering?
3. Através de simulação numérica, pode-se analisar variações
 - Quando aumenta-se p de 0 a 1
 1. Decrescimento rápido da distância média
 2. Decrescimento lento do clustering

Clustering Coefficient - Characteristic Path Length



log-scale in p

Quando $p = 0$, $C = 3(k-2)/4(k-1) \sim 3/4$

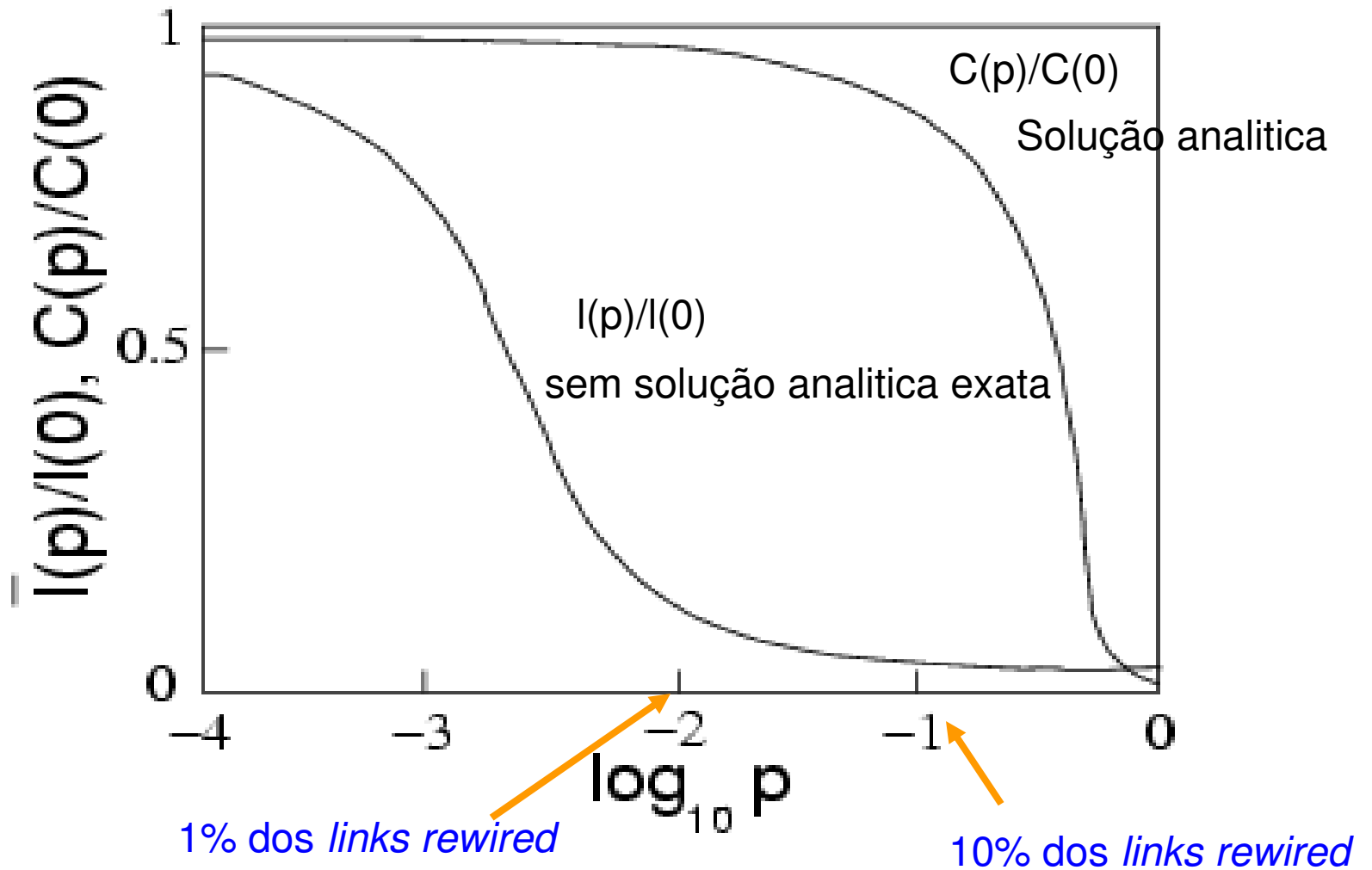
$L = n/k$

$$\frac{n}{2k}$$

Para pequenos p , $C \sim 3/4$

$L \sim \log n$

Mudanças no "clustering coefficient" e no comprimento do caminho mínimo médio como função de p (fração de "rewired edges")

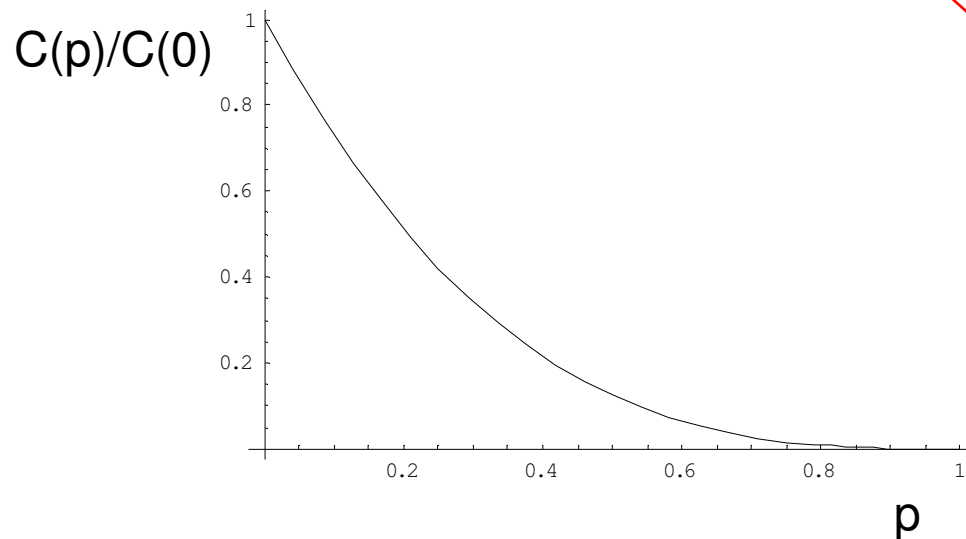


Olhar definição em Newman MEJ, The structure and function of complex networks.

Clustering coefficient para o modelo SW com "rewiring"

triângulo conectado

1. A probabilidade que uma ~~tripla conectada~~ permaneça conectada após o *rewiring*
 1. Probabilidade que nenhum das 3 arestas foi rewired $(1-p)^3$
 2. Probabilidade que as arestas fossem rewired de volta para cada um é mínima e pode ser ignorada
2. Clustering coefficient = $C(p) = C(p=0) \cdot (1-p)^3$



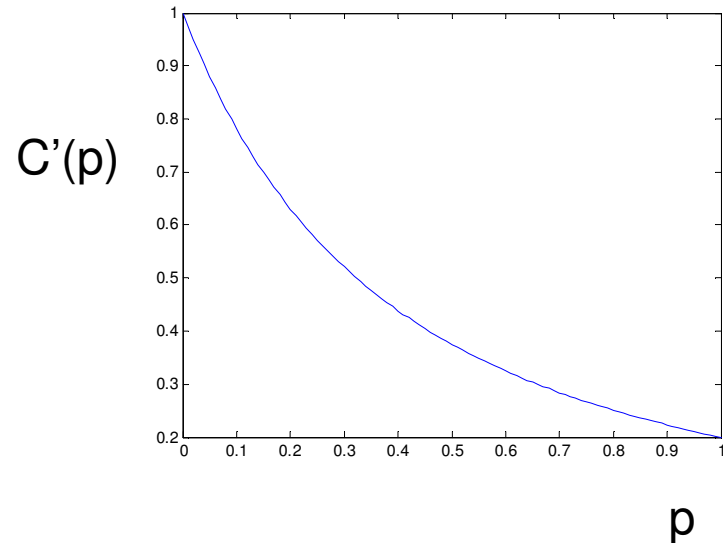
**Que viagem...
De onde surgiu?
Se alguém entender
isso, por favor
pode explicar?**

Clustering coefficient: adição de arestas randomicas

1. Quanto C depende de p?
2. $C'(p) = 3 \times \text{numero de triangulos} / \text{numero de triplas conectadas}$
3. $C'(p)$ computado analiticamente para modelo *small world* sem *rewiring*

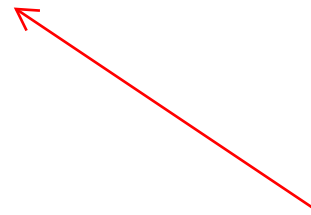
$$C'(p) = \frac{3(k-1)}{2(2k-1) + 4kp(p+2)}$$

De onde isso surgiu?



Distribuição de graus

1. $p=0$
2. $p>0$ amplia distribuição
3. Arestas deixadas intactas com probabilidade $(1-p)$
4. Arestas “rewired” para i com probabilidade $1/N$



ih...

O que danado é "i"???

Exemplos de Redes Reais (valores médios)

Rede	Tamanho	grau Vertice	shortest path	Shortest path in fitted random graph	Clustering (# triangles)	Clustering (averaged over vertices)	Clustering em grafo aleatorio
Atores de Filmes	225,226	61	3.65	2.99	0.20	0.79	0.00027
MEDLINE co-autoria	1,520,251	18.1	4.6	4.91	0.45	0.56	1.8×10^{-4}
E.Coli substrate graph	282	7.35	2.9	3.04		0.32	0.026
C.Elegans	282	14	2.65	2.25		0.28	0.05

Limitações do Modelo SW

1. SW somente endereça o aspecto de conectividade de uma rede
2. SW não faz uso das distribuições de graus da rede
3. SW não considera:
 1. Tipos de links.
 2. Aparecimento e desaparecimento de arestas.
 3. Grau médio de arestas

Aplicações em Redes de Computadores

Degree of separation: social networking



The screenshot shows a website header with navigation links: Home, Accessibility, Sitemap, and O2 sites. Below the header is a dark blue navigation bar with links for Media centre, Investor centre, O2 in society, O2 services, and Working. The main content area features a breadcrumb trail: You are here: Media centre > Group press releases. The title of the press release is 'The six degrees of separation is now three', dated 19/08/2008. The text discusses a study by O2 showing that the average person is now connected to others in just three degrees, down from the traditional six degrees. A yellow box highlights the sentence: 'However, the O2 study reveals that within a shared 'interest' network (i.e. hobbies, sport, music, religion, sexuality etc), the average person is connected by just three degrees.' An arrow points from this box to the text below.

Home → Accessibility → Sitemap → O2 sites

Media centre | Investor centre | O2 in society | O2 services | Working

You are here: Media centre > Group press releases

Press release

19/08/2008

The six degrees of separation is now three

Six degrees of separation has fallen to three due to the impact of social networking and developments in technology, according to a study carried out by O2.

O2 commissioned social organisational specialist Jeff Rodrigues to examine the impact of technology on how connected people are. The research included over 50 hours of in-depth interviews with adults across three different age groups, (18-25, 35-45, 55+) and found that the conventional notion of six degrees of separation is out of date.

The term was coined by US psychologist Stanley Milgram following a 1967 experiment. The six degrees theory was upheld in a 2006 Microsoft study of instant messenger conversations. However, the O2 study reveals that within a shared 'interest' network (i.e. hobbies, sport, music, religion, sexuality etc), the average person is connected by just three degrees.

Rodrigues finds that we are usually part of three main networks based on family, friendship and work. Outside of these we are, on average, part of five main shared 'interest' networks based on a range of personal interests from hobbies, sport, music and the neighbourhood we live in, to religion, sexuality and politics. It is the growth of these shared interest networks and the influence of technology on them that has led to the reduction in the number of degrees of separation.

All respondents were asked to make contact with an unknown person from destinations selected at random from across the globe using only personal connections. By using their shared interest networks the participants were able, on average, to make the connection in three person-to-person links.

Aplicações em Computação

1. Links entre webpages.
2. Freenet.
3. Internet.
4. Redes de Sensores
 1. Sharma, G. and Mazumdar, R.; Hybrid sensor networks- A small world, Mobihoc 2005, Urbana, June 2005
5. Redes Wireless.
 1. A. Helmy, "Small Worlds in Wireless Networks", *IEEE Communications Letters*, pp. 490-492, Vol. 7, No. 10, October 2003.

Sobre o Experimento de Milgram

“The geographic movement of the [message] from Nebraska to Massachusetts is striking. There is a progressive closing in on the target area as each new person is added to the chain”

S.Milgram 'The small world problem', Psychology Today 1,61,1967

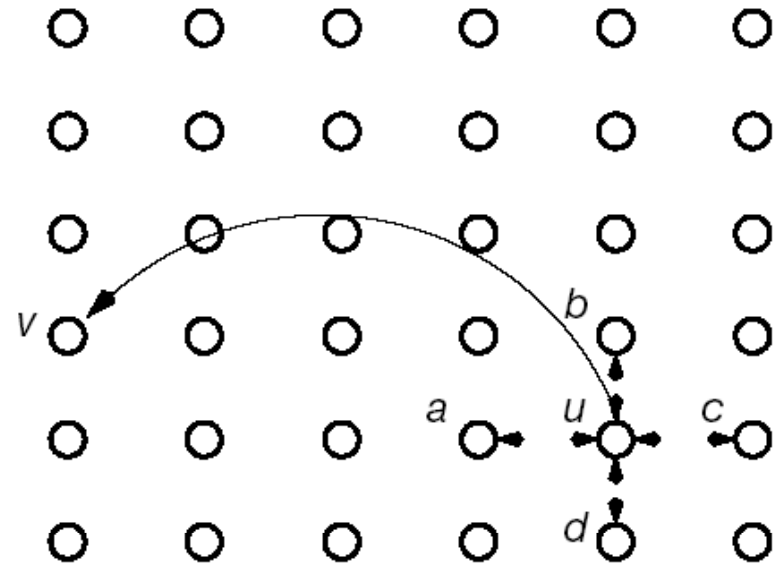


- 1) O experimento de Milgram descobriu a existência de “short paths”
- 2) Pessoas em sociedade, com conhecimento local de amigos, foram capazes de coletivamente uma carta rapidamente para um destino remoto.
- 1) Visto em termos computacionais, estamos falando de roteamento

Modelo Geográfico de Small World de Kleinberg

Perguntas:

- O experimento mostra que o modelo SW é rico em conexões curtas estruturadas e tem poucas conexões aleatórias de longo-alcance.
- Característica de SW é o diâmetro limitado por $\log N$.
- Existe sempre um caminho curto entre 2 nodos, mas a questão é se um algoritmo descentralizado com informações locais é capaz de descobrir os caminhos curtos entre dois nodos?



Nodos são colocados num grafo regular (lattice-treliça) e conectados aos vizinhos mais próximos

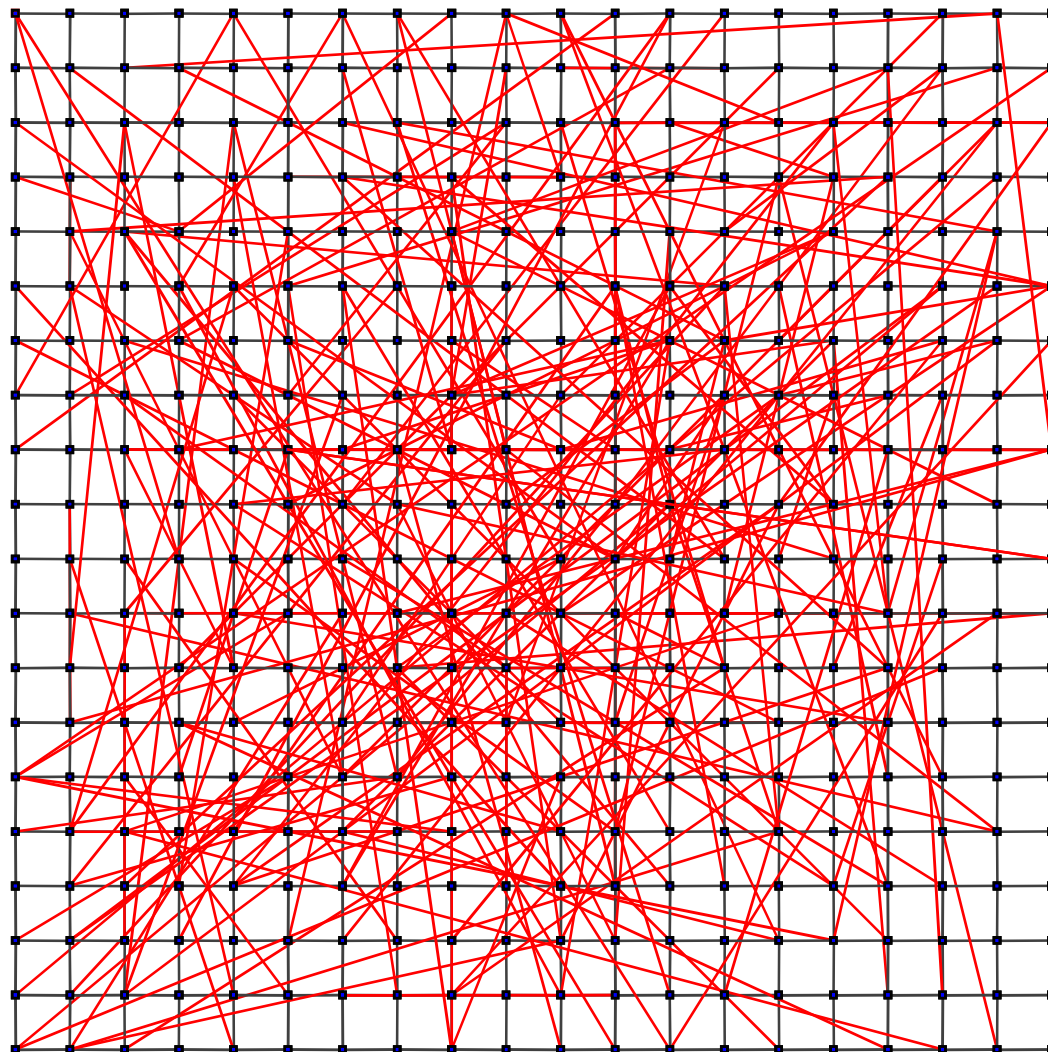
Links adicionais são colocados com $p_{u,v} \approx d_{u,v}^{-r}$

[Kleinberg, 'The Small World Phenomenon, An Algorithmic Perspective' \(Nature 2000\)](#)

Sem localidade

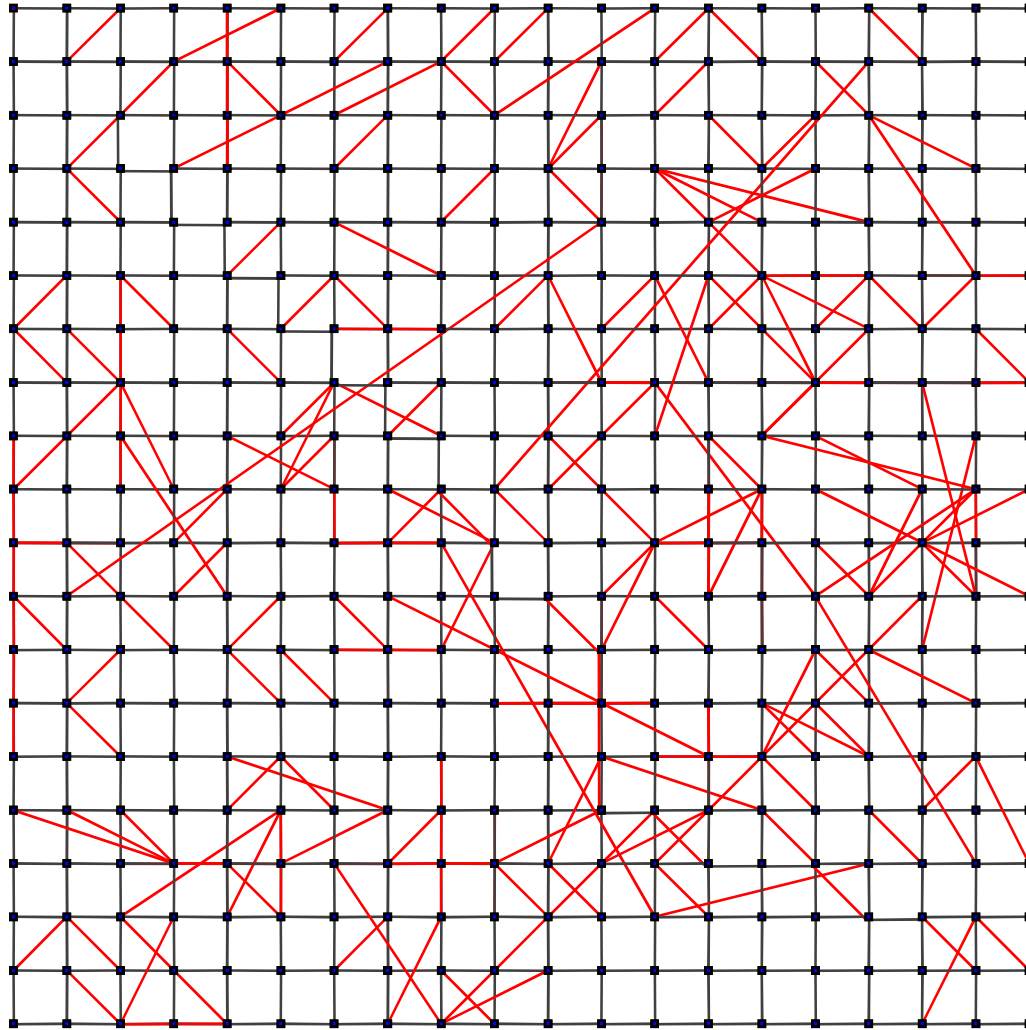
quando $r=0$, são randomicamente distribuidos, $ASP \sim \log(n)$, n tamanho do grid.

$$p \sim p_0$$



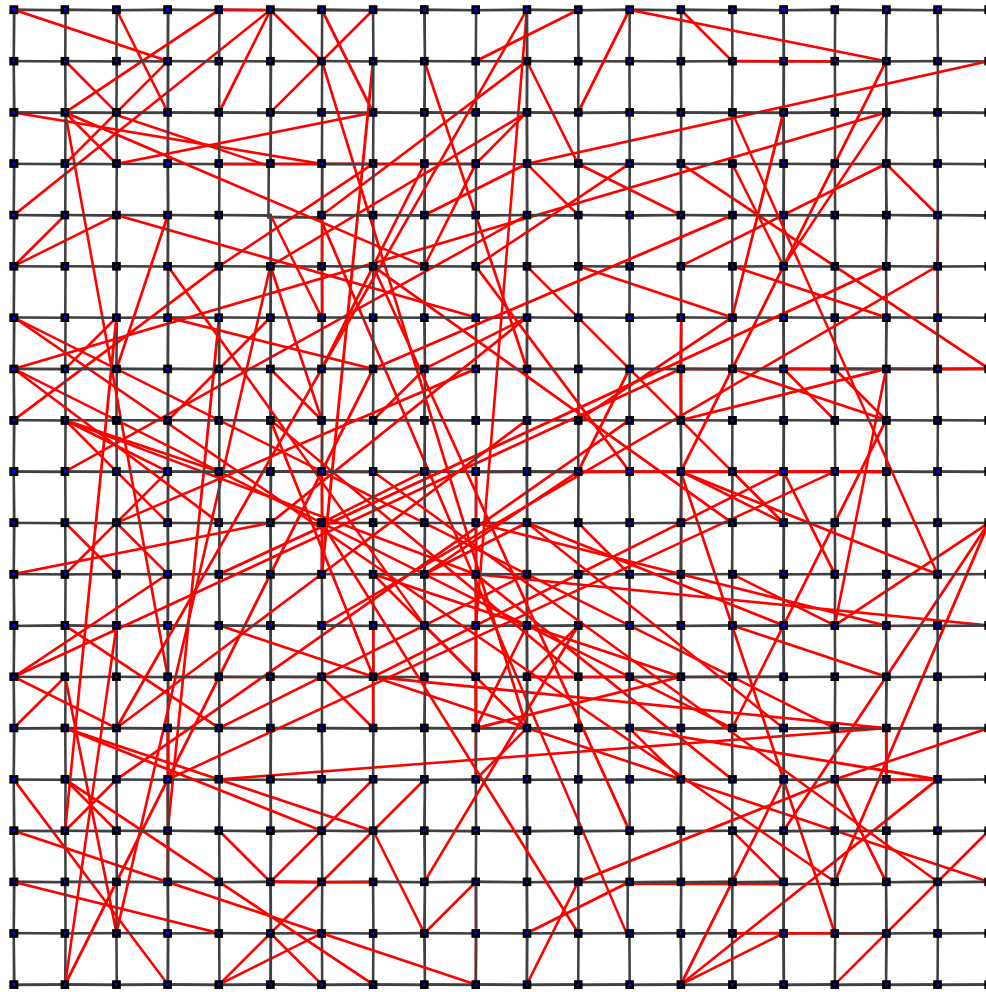
Links altamente localizados no lattice

$$p \sim \frac{1}{d^4}$$

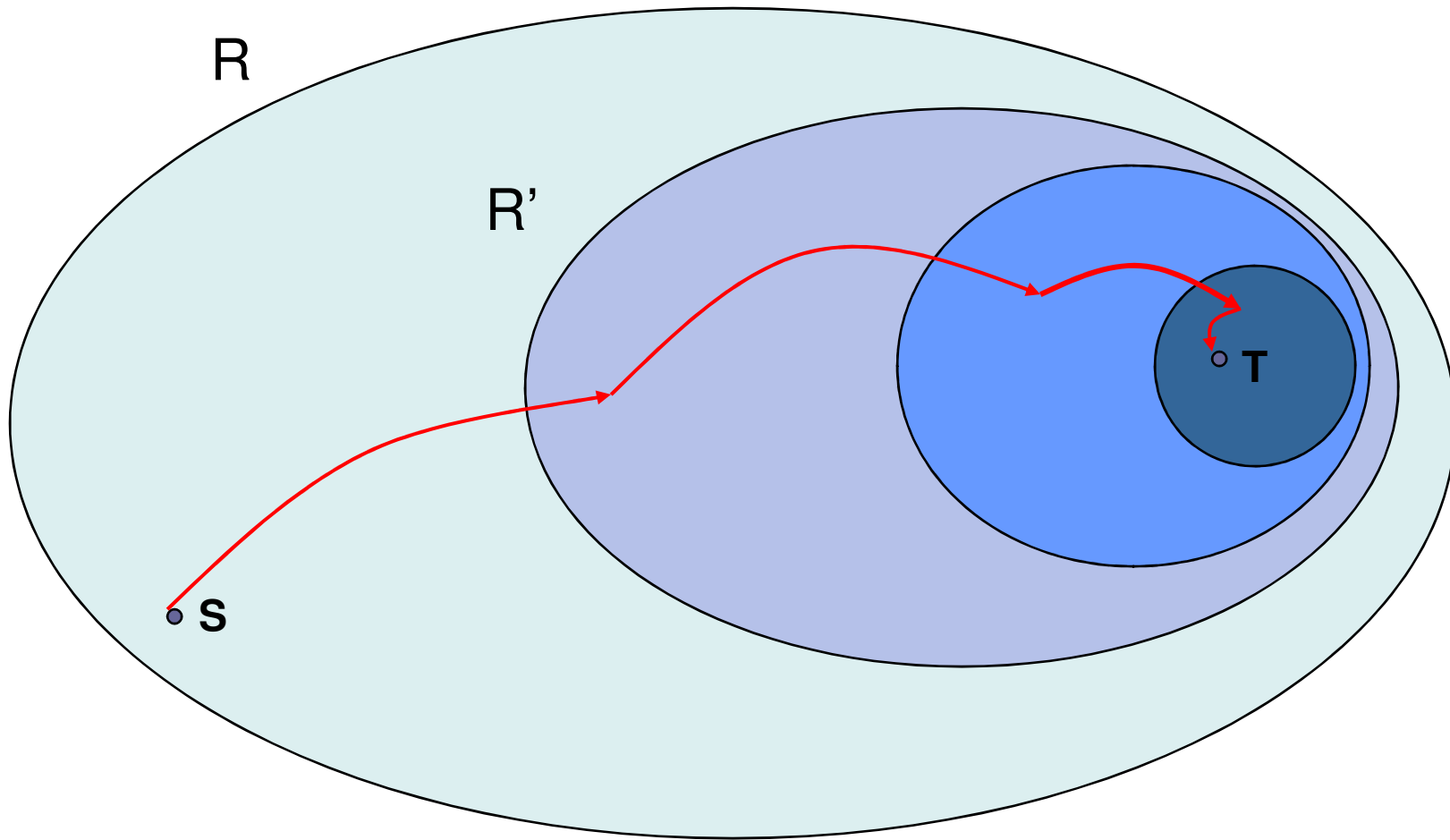


Links equilibrados entre o curto e longo alcance

$$p \sim \frac{1}{d^2}$$



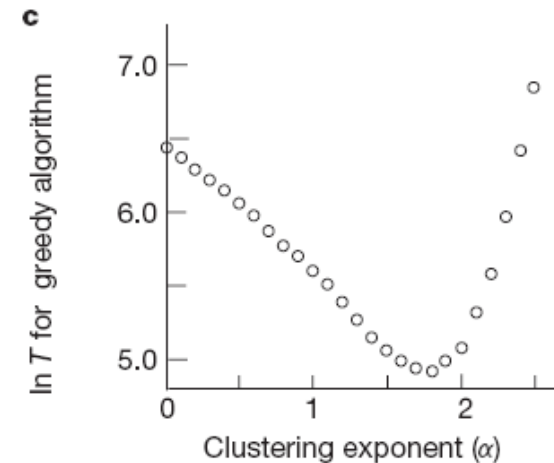
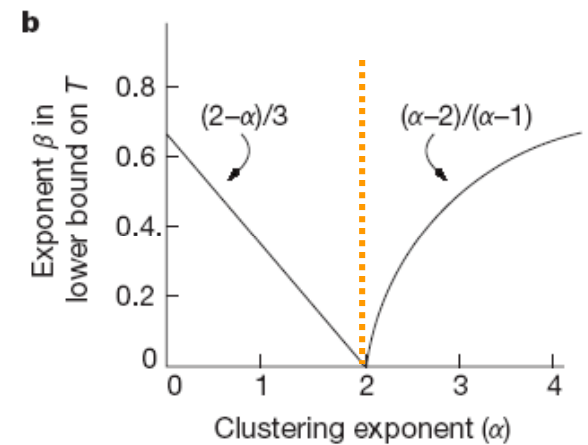
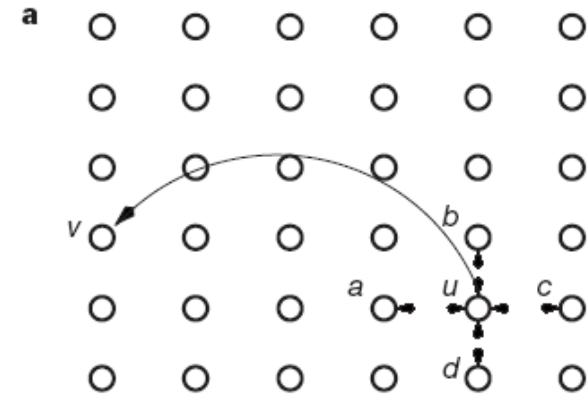
Como o fenômeno small world aparece? Explicações....



Busca em Redes SW

Kleinberg (2000)

- a) Variação do modelo β de Watts's
1. Lattice é duas dimensões ($d=2$).
 2. Um link randomico por nodo.
 3. Parametro α controla probabilidade do link randomico – maior para nodos proximos.
- b) Para $d=2$, time-to-search em $\alpha=2$
1. Para baixos α , RG; sem correlação “geographic” nos links.
 2. Para altos α , não é small world; sem atalhos (short paths) para encontrar.
- c) Searchability em $\alpha=2$, por simulação



Exercício

1. Um voluntario “sorteado*” de uma urna apresentará o artigo abaixo, com seus comentários sobre a teoria do modelo Small World, bem como fará conjecturas sobre aplicabilidade desses fundamentos:
2. [Geographic routing in social networks](#)
D Liben-Nowell, J Novak, R Kumar, P Raghavan, A ... -
Proceedings of the National Academy of Sciences, 2005 –
www.pnas.org
PNAS | August 16, 2005 | vol. 102 | no. 33 | 11623-11628 .

* Caso o aluno sorteado não esteja preparado, perder os pontos e um novo é sorteado, assim sucessivamente....