Redes Complexas

Modelagem de Sistemas Complexos

EACH 2016

O que são redes?

- Abstração que permite codificar relacionamentos entre **pares** de objetos

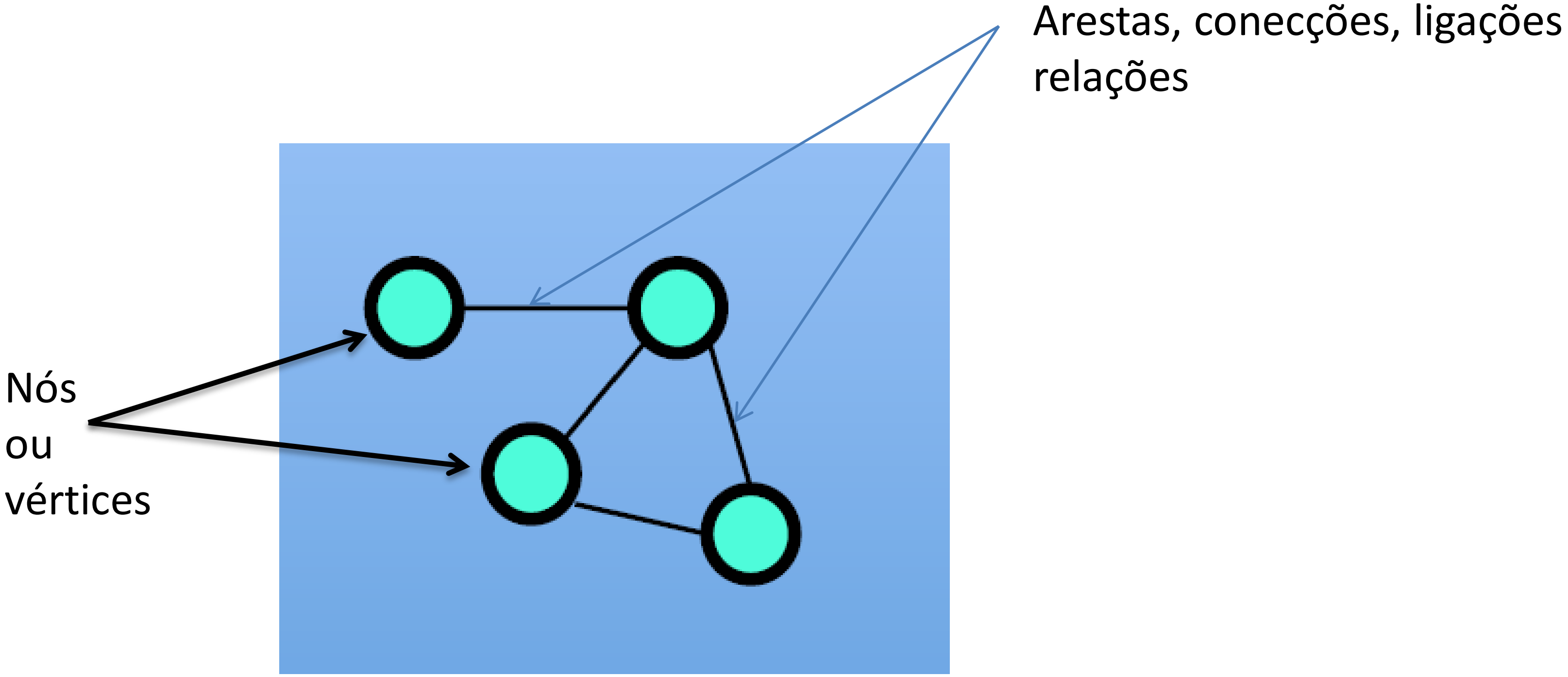
objetos → vértices do grafo

relacionamentos → arestas do grafo

Os **objetos** podem ser: pessoas, cidades, computadores, paginas web Filmes, empresas, bancos, proteinas, animais, etc...

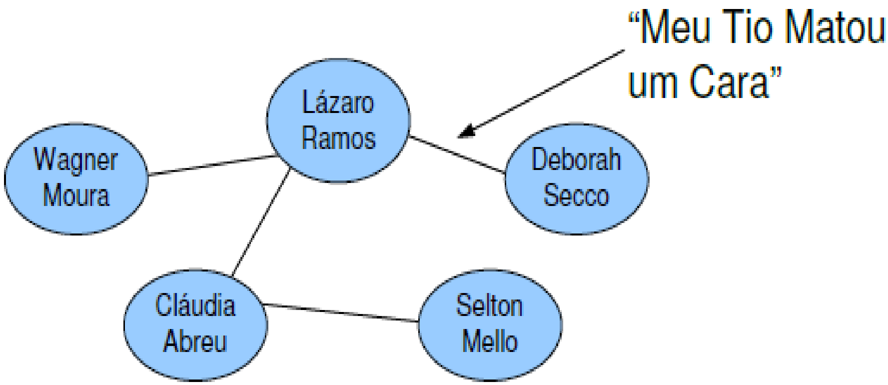
Os **relacionamentos** podem ser: amizade, colaboração científica, língua faladas (dialetos), voos, produção, função biológica, sentido da predação etc..

O que são redes?

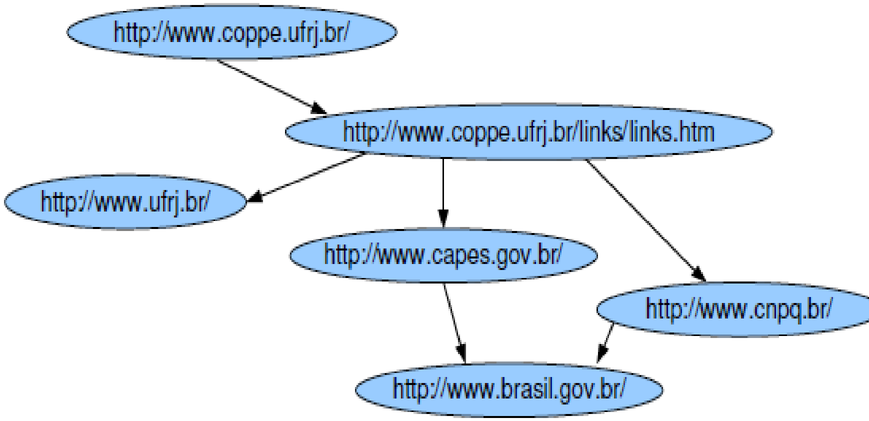


Rede captura estrutura do sistema

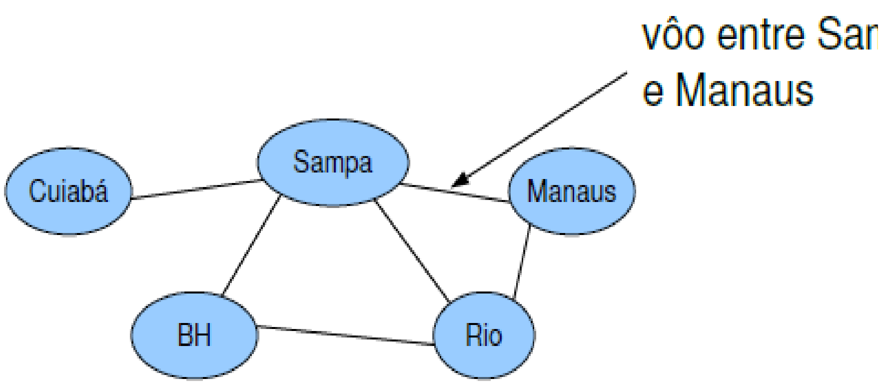
- objeto: atores
- relacionamento: atores atuam em um mesmo filme



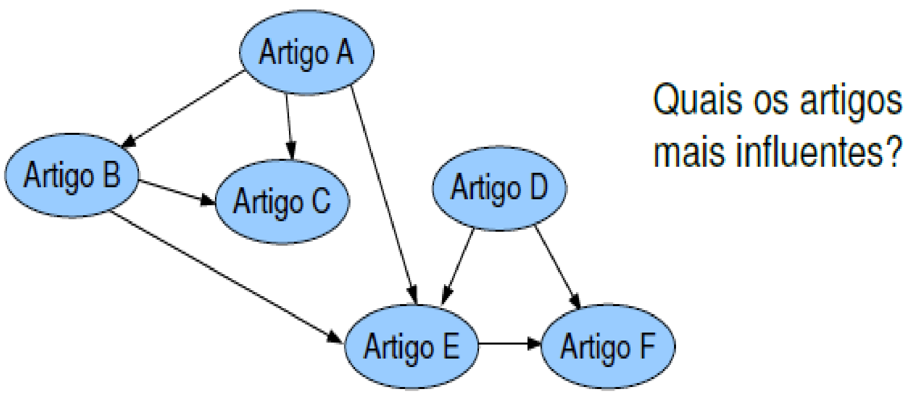
- objeto: páginas web
- relacionamento (direcionado): *link* de uma pagina para outra



- Transporte aéreo
  - objeto: cidades
  - relacionamento: vôo comercial entre duas cidades



- objeto: artigos científicos
- relacionamento (direcionado): citação entre os artigos



Histórico

1736- Euler introduz a idéia de grafo

1926- Boruska propõe árvore geradora mínima –rede elétrica em Morávia

1930- Moreno aplica redes na sociologia - sociogramas

1950- Freeman – medidas de centralidade

1957- Redes aleatórias por Rapoport

1959-1960 -Redes aleatórias por Erdos e Reny

1965- Price propõe modelo matemático ligação preferencial (Herbet Simon)

1967- Stanley Milgram e o Mundo Pequeno com 6 graus de separação

1968- Merton publica trabalho sobre efeito Mathew

1998 – Modelo Watson-Strogatz para mundo pequeno e Page Rank (Google)

1999- Modelo Albert-Barabasi introduz as redes livre de escala.

Ciência das Redes

Por que estudar as redes?

Ciência das Redes

Muitas redes possuem propriedades não triviais do ponto de vista topológico

Como assim?

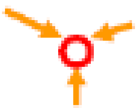
Elas não são regulares e nem são aleatórias!

Contudo, muitas redes possuem propriedades topológicas iguais. Algumas são Redes que representam o mesmo sistema, mas em locais diferentes, enquanto outras redes representam sistemas completamente distintas,

Como isso é possível ? Por que? Como estudá-las?

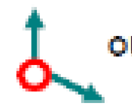
Nós

- Propriedades dos nós da rede
  - Das conexões imediatas (locais)
    - indegree
      quantas arestas orientadas que chegam a um nó



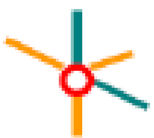
indegree=3

    - outdegree
      quantas arestas orientadas originam em um nó



outdegree=2

    - Grau: Degree (in ou out):
      número de arestas que incidem em um nó



degree=5

  - Do grafo inteiro (globais)
    - centrality (betweenness, closeness)

Como Falar sobre Redes?

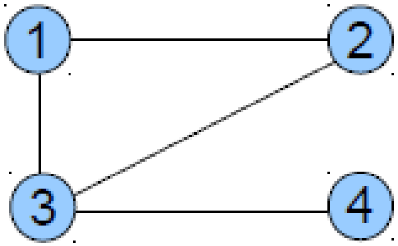
- Desenho da rede: “figura vale por mil palavras”
- Não funciona para redes grandes

Descrever estrutura da rede!

- **Matriz de adjacência**
- Matriz *n* x *n* (*n* é número de vértices)
  - $a_{ij} = 1$, se existe aresta entre vértices *i* e *j*
  - $a_{ij} = 0$, caso contrário
- Codifica todos os relacionamentos da rede
- Generaliza com pesos, pode ser assimétrica

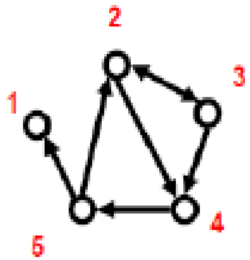
Matriz de Adjacência

- Exemplo

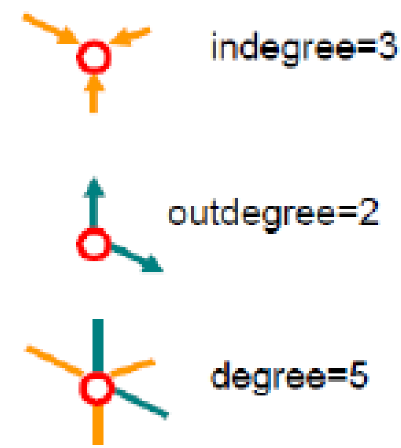


	1	2	3	4
1	0	1	1	0
2	1	0	1	0
3	1	1	0	1
4	0	0	1	0

Grafo direcionado



Grau de Nós da Matrix de Valores



- Outdegree = $\sum_{i=1}^{n} A_{ij}$

exemplo: o outdegree para nó 3 é 2, certo?

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

$$\sum_{j=1}^{n} A_{3j}$$

- Indegree = $\sum_{j=1}^{n} A_{ij}$

exemplo: o indegree para nó 3 é 1,

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

$$\sum_{i=1}^{n} A_{i3}$$

Matriz de Adjacência

- Problema em descrever a estrutura com matriz de adjacência?
- Possui toda a informação mas pouco intuitiva
- **Matriz é o DNA da rede!**
  - como descrever uma pessoa? Através do seu DNA?

Precisamos de resumos da estrutura!

- Resumos que ajudem a entender a estrutura de forma intuitiva

Características de Redes

- Resumo da estrutura da rede
  - caracterísiticas estruturais
  - ex: tamanho, densidade, graus, distâncias, clusterização, centralidade, homofilia, etc
- Dão ideia geral da estrutura da rede

- Quais características devem ser avaliadas? Quais são importantes?
- Depende do que você deseja estudar!
- Como gens que formam o DNA, estamos apenas começando a enteder seu significado

Característica Importante

- O que faz uma característica ser importante?

1) Prever (determina) comportamento geral de algum processo

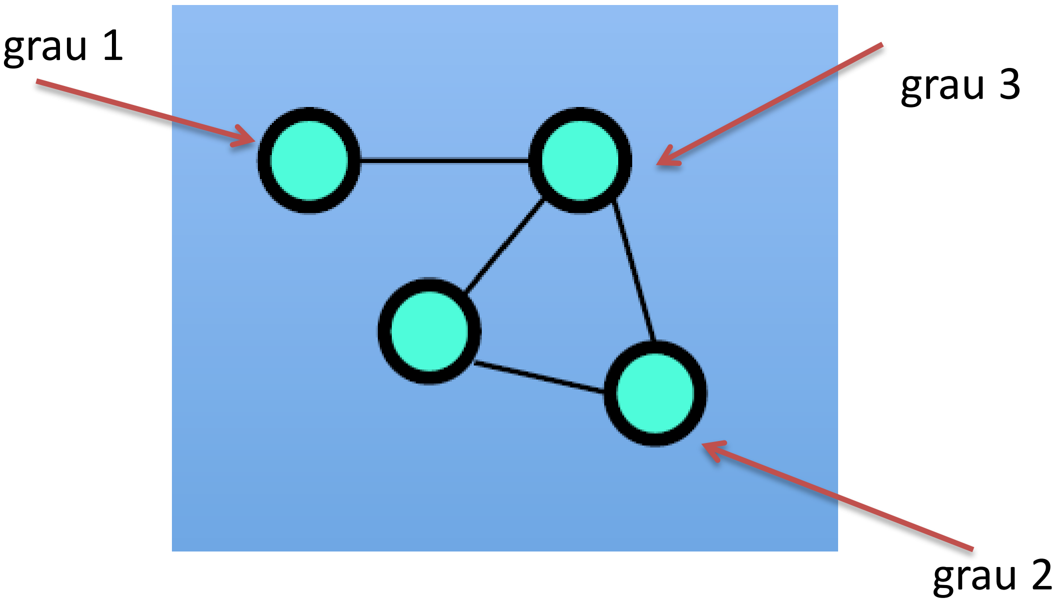
  - independente de outras características

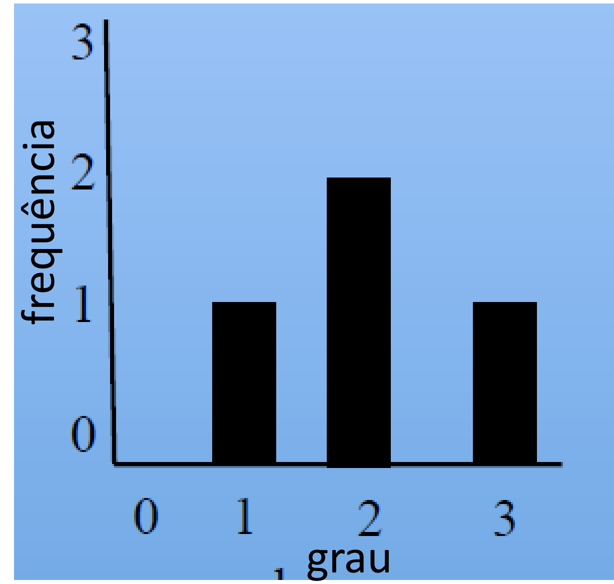
2) Influência sobre diferentes processos

- Exemplo: distribuição de grau
  - determina comportamento de passeios aleatórios
  - influencia outros processos (epidemia)
- Não conhecemos muitas características fundamentalmente importantes
  - a serem descobertas (como gens importantes)

Como caracterizar uma redes?

Grau do vértice: número de arestas ou ligações de um vértice ou nó.





- Distribuição (empírica) do grau dos vértices

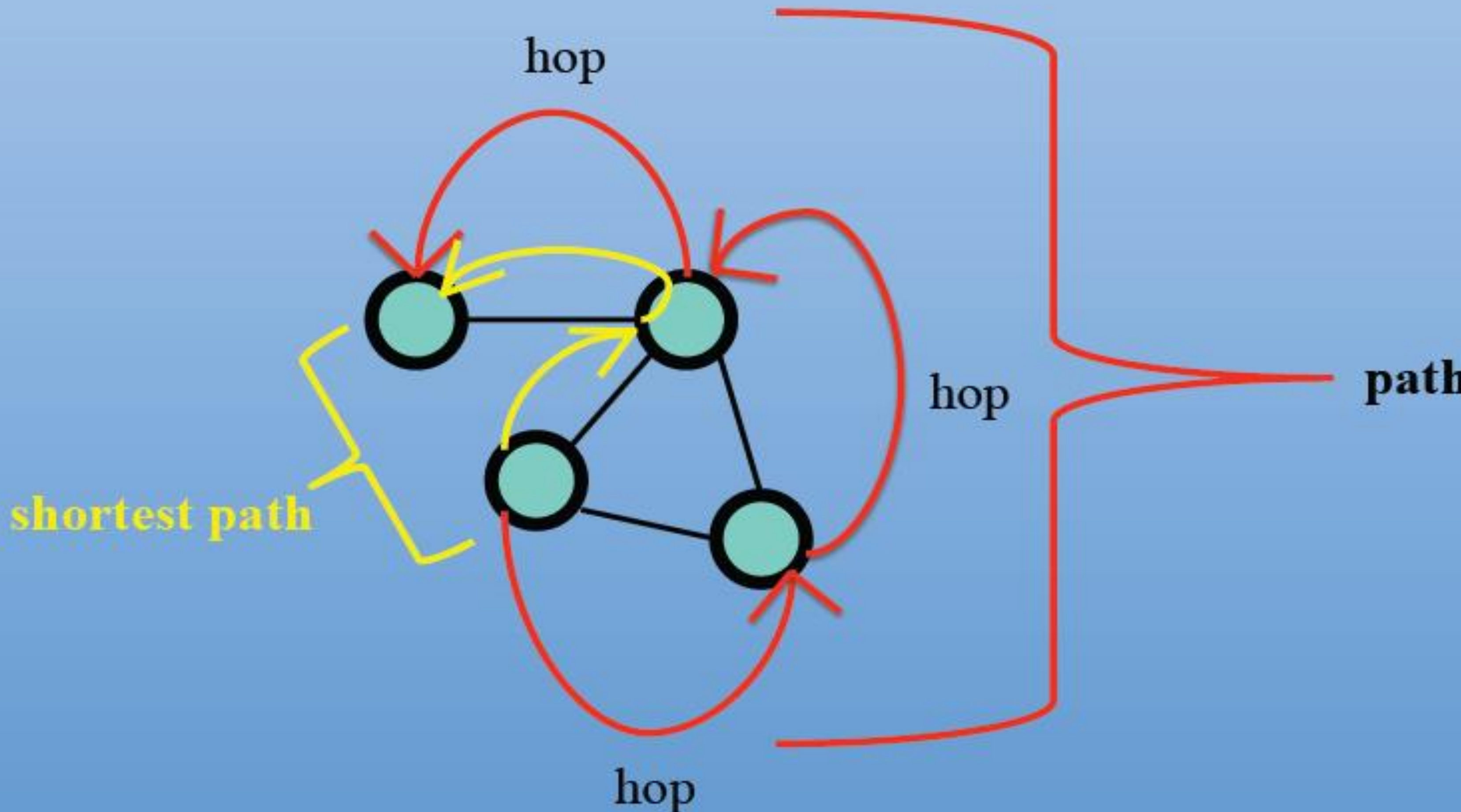
$$f_D(k) = P[D = k] = \frac{\text{Número de vértices com grau k}}{\text{Número total de vértices}}$$

Variável aleatória que denota grau de um vértice

Vértices e Arestas

- Número de vértices de um grafo
  - n = |V| ← cardinalidade do conjunto
- Número de arestas de um grafo
  - m = |E|
- Dado G = (V, E), qual é o maior número de arestas de G?
  - número de pares não ordenados em um conjunto de n = |V| objetos → $\binom{n}{2} = \frac{n(n-1)}{2} \leq n^2$
- Densidade: fração de arestas que o grafo possui $\rho = \frac{2m}{n(n-1)}$

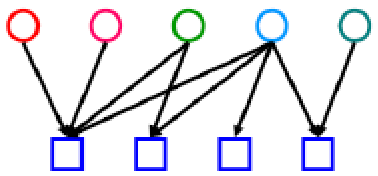
Distance and Paths in Networks



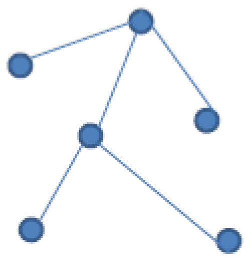
***Distance L* between nodes *A* and *B*:**

Number of hops in shortest path between *A* and *B*

Digrafos

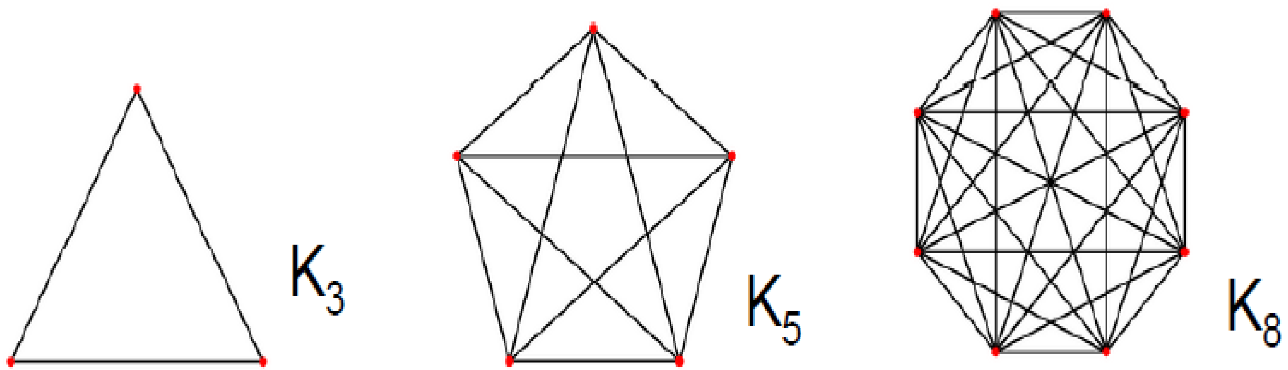


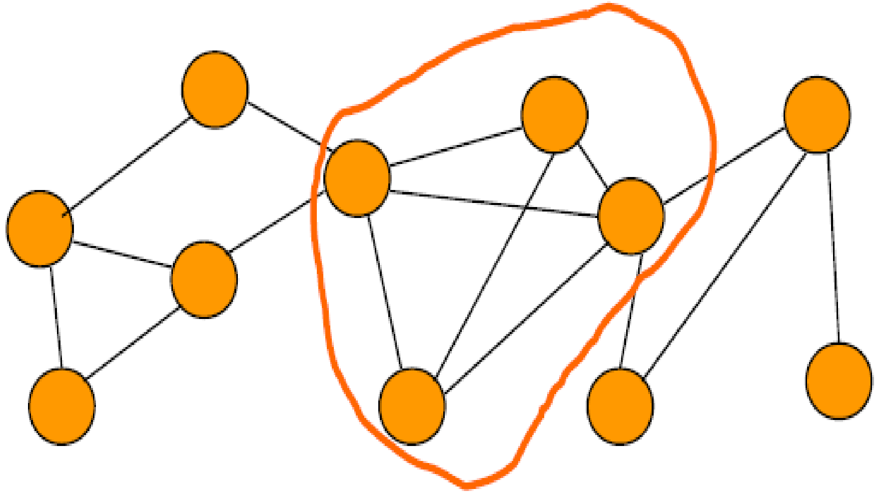
Árvores



Cliques e grafos completos

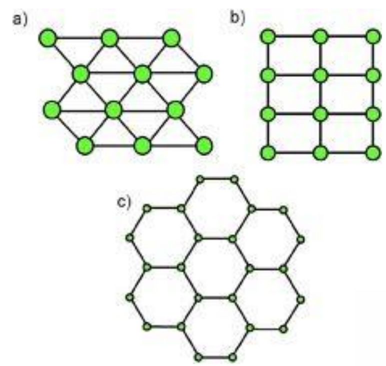
- K_n é um grafo completo (clique) com K vértices
  - Cada vértice é conectado a todos outros vértices.



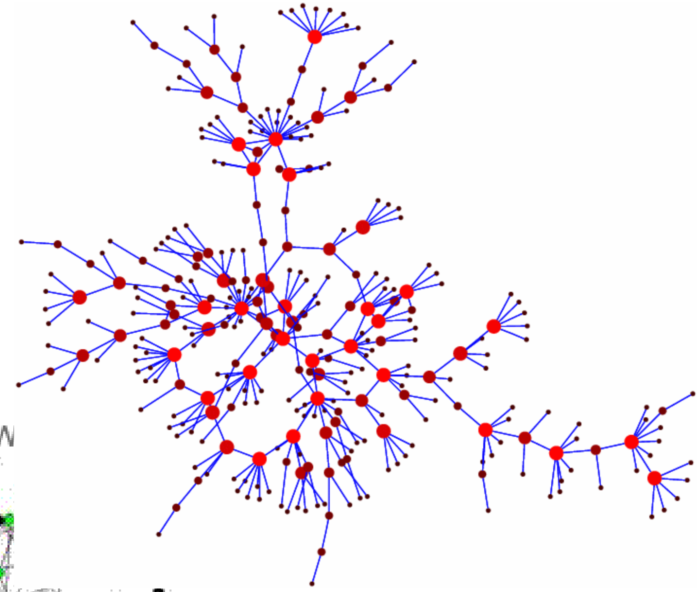


Tipos de Redes

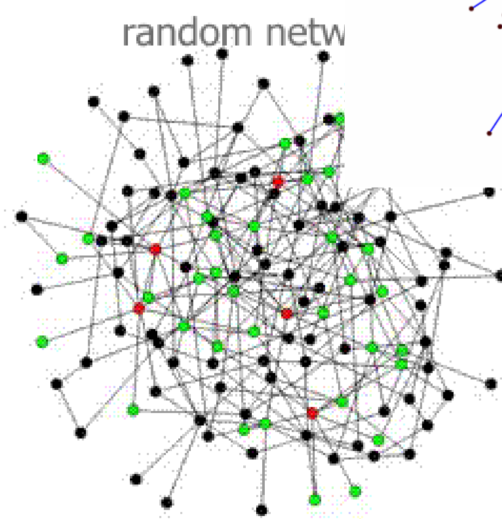
Redes Regulares



Redes Complexas



Redes Aleatória



Exemplos de redes complexas

Exemplos:

- Internet e World Wide Web (WWW)
- Redes sociais de amizade ou profissional
- Redes de relacionamentos entre empresas
- Redes neurais do cérebro
- Redes celulares, metabólicas
- Teias alimentares
- Redes de distribuição: vasos sanguíneos e rotas postais
- Redes de citações entre artigos
- Redes de colaboração de pesquisadores, etc.

Redes Complexas

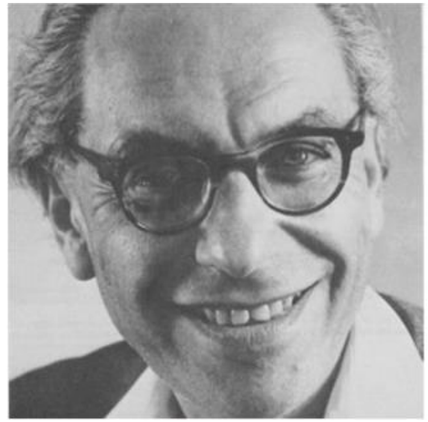
- Descobrimento de estruturas complexas
  - Em todos os lugares!
- Como e porque tais estruturas surgem?
- Qual importância da estrutura?
- Funcionalidade como função da estrutura?

Estrutura é tudo!

Primeiro vamos entender a rede aleatória

Primeiro vamos entender a rede aleatória

Pál Erdös
(1913-1996)

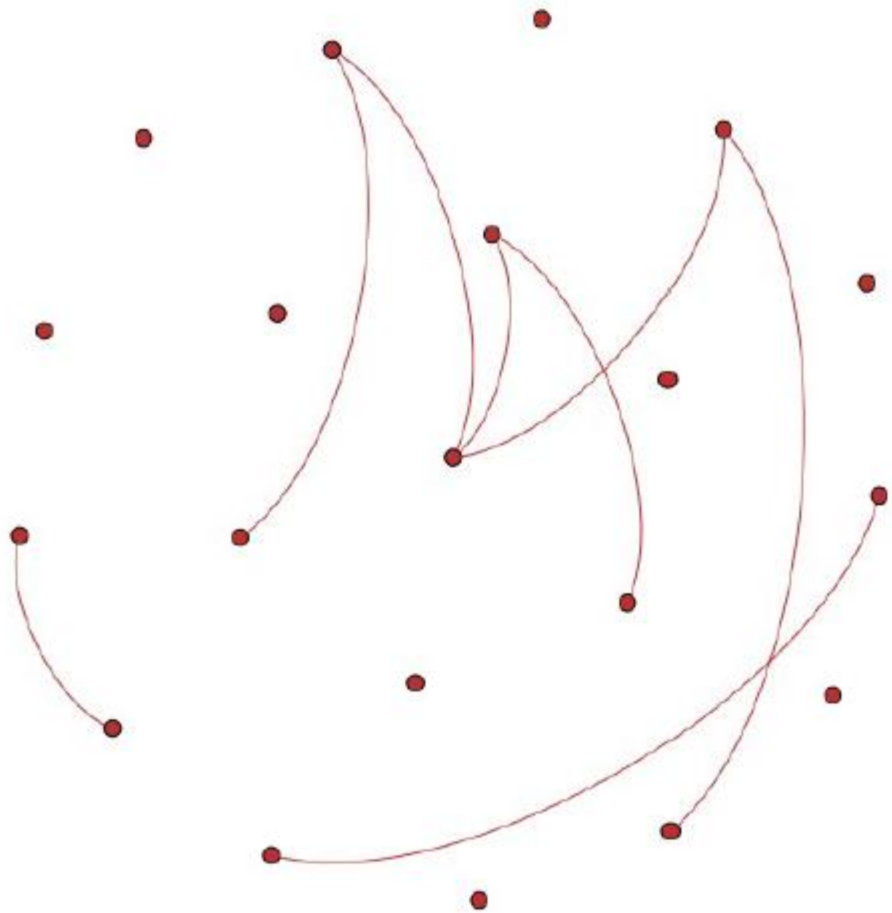


Alfréd Rényi
(1921-1970)

Erdös-Rényi model (1960)

Definição:
Um grafo aleatório é um grafo com N vértices, onde cada par de vértice é conectado com probabilidade **p**.

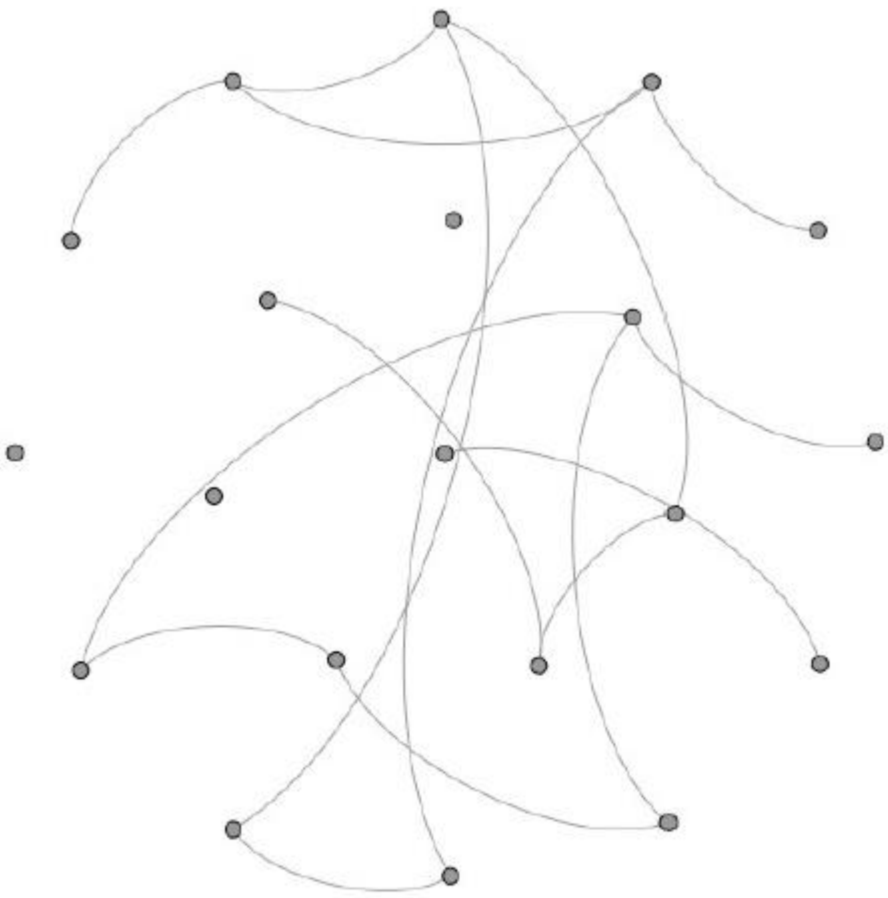
Notação G(N, p).



$n=20$
$m=8$
$\rho = 0{,}042$

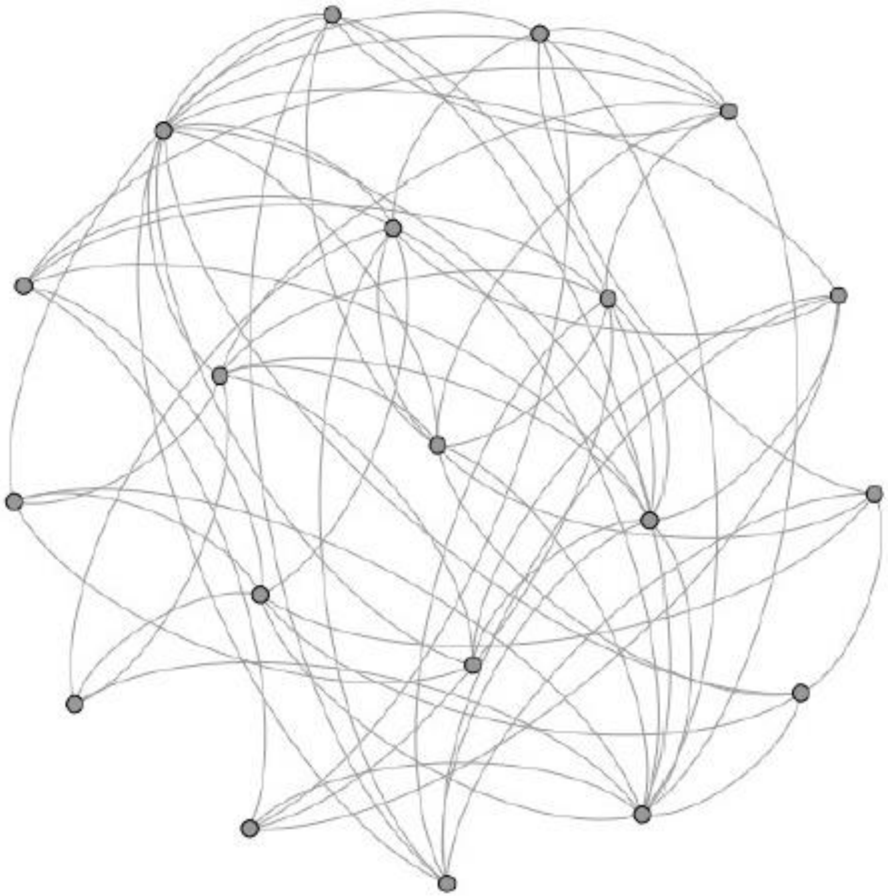
grau médio=0,8

$\rho = 2m/n(n-1) \leftarrow$ Densidade



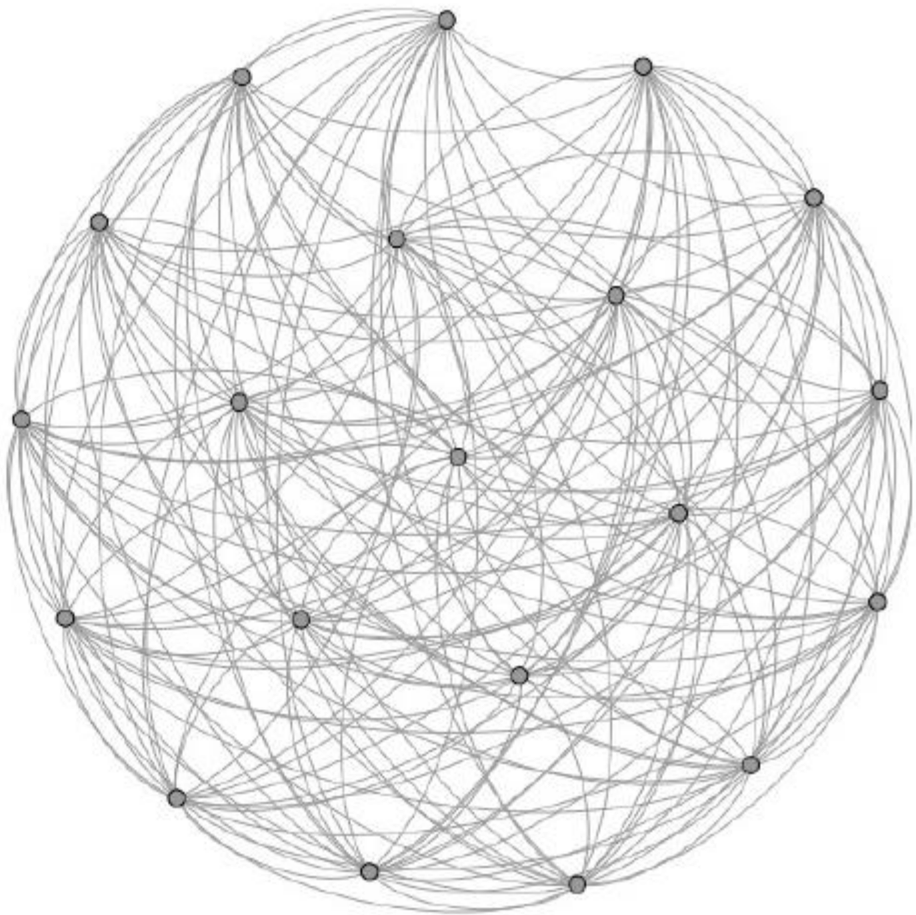
n=20
m=16
ρ = 0,084

grau médio=1,6



n=20
m=70
ρ = 0,368

grau médio=7



n=20
m=190
ρ = 1

grau médio=19

Algumas Propriedades Compartilhadas pelas Redes Complexas

1. Propriedade Mundo Pequeno (Small World)

2. Formação de Aglomerado e Comunidades

3. Robustez a falhas aleatórias

4. Vunerabilidade a ataques aos hubs

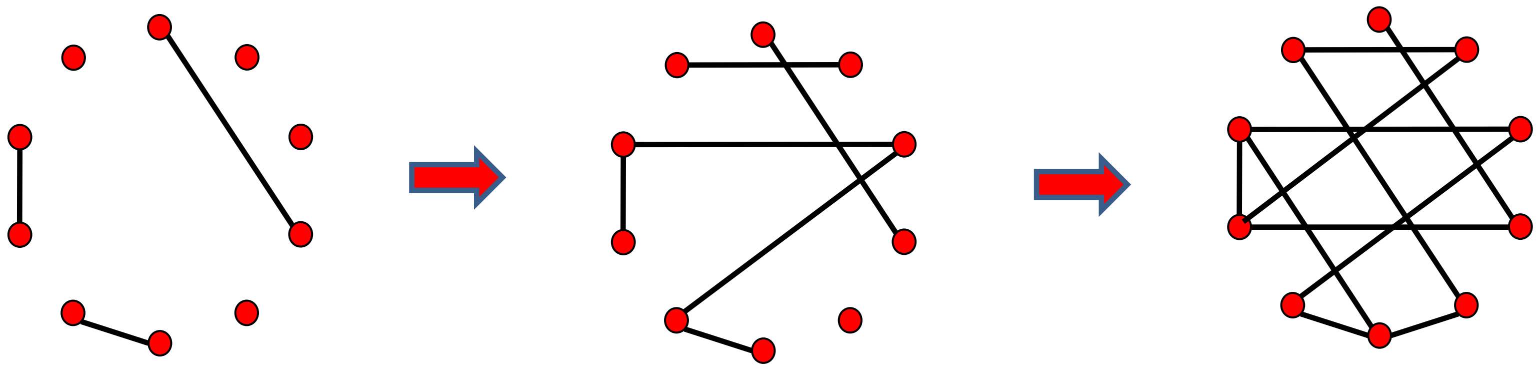
5. Vunerabilidade a falhas em cascata

Redes Aleatórias

Distribuição da conectividade segue uma binomial

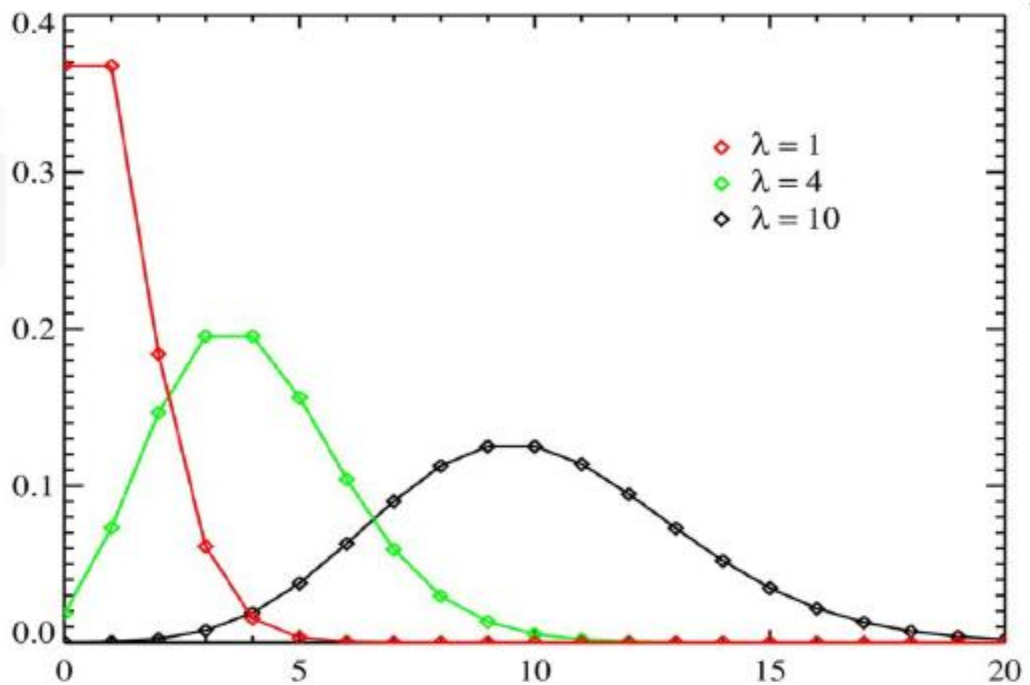
$$P(k) = \binom{N-1}{k} p^k (1-p)^{(N-1)-k}$$

Redes complexas evoluem a partir da conexão de nós inicialmente isolados



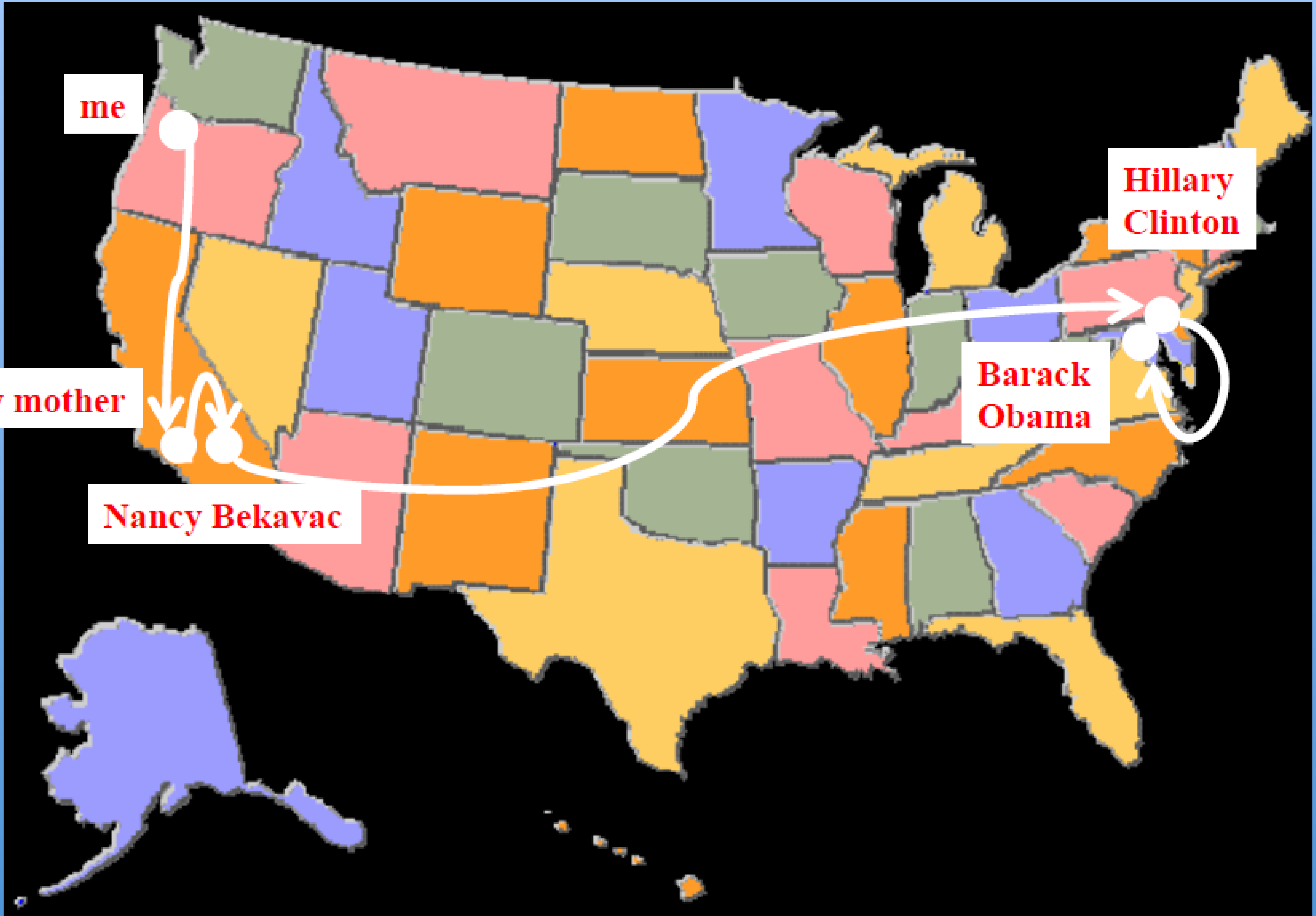
A componente gigante contém a maior parte dos nós!

Se for medido o número de arestas para cada vértice de um grafo aleatório, e o resultado desenhado em um histograma, a **distribuição de graus** seguirá uma **Poisson**.

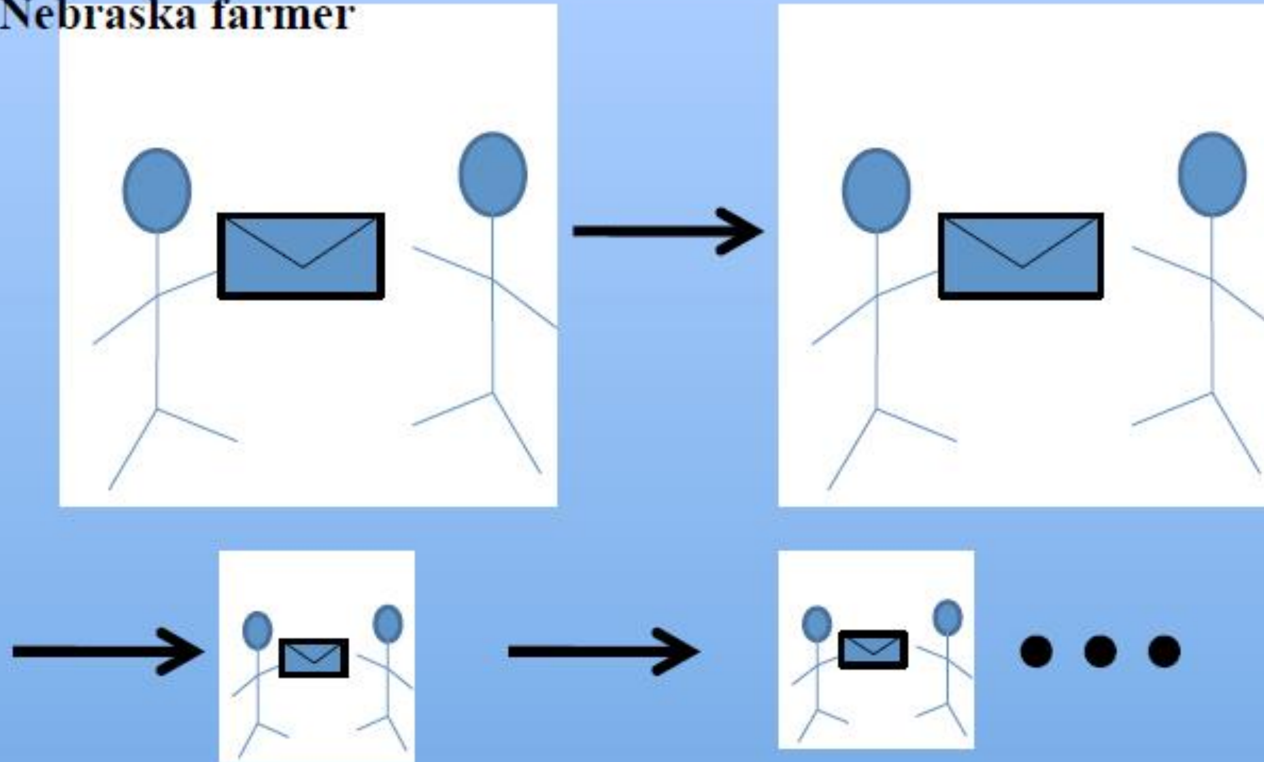


Esse tipo de grafo é conhecido também como grafo aleatório de Poisson

Small-World Networks



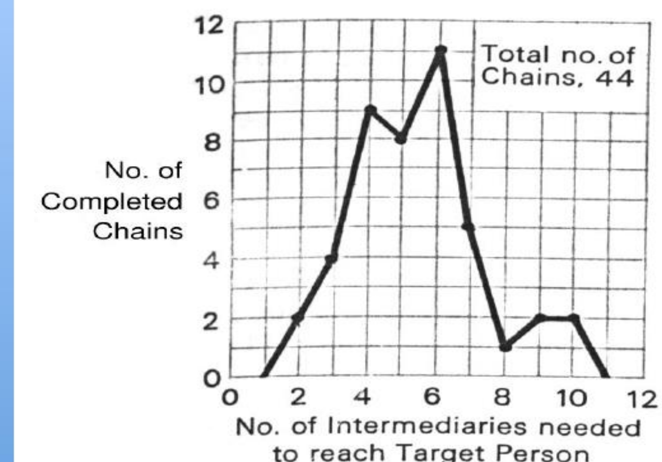
Nebraska farmer



Stanley Milgram

On average: "six degrees of separation"

From S. Milgram, The small-world problem. *Psychology Today*, 1967

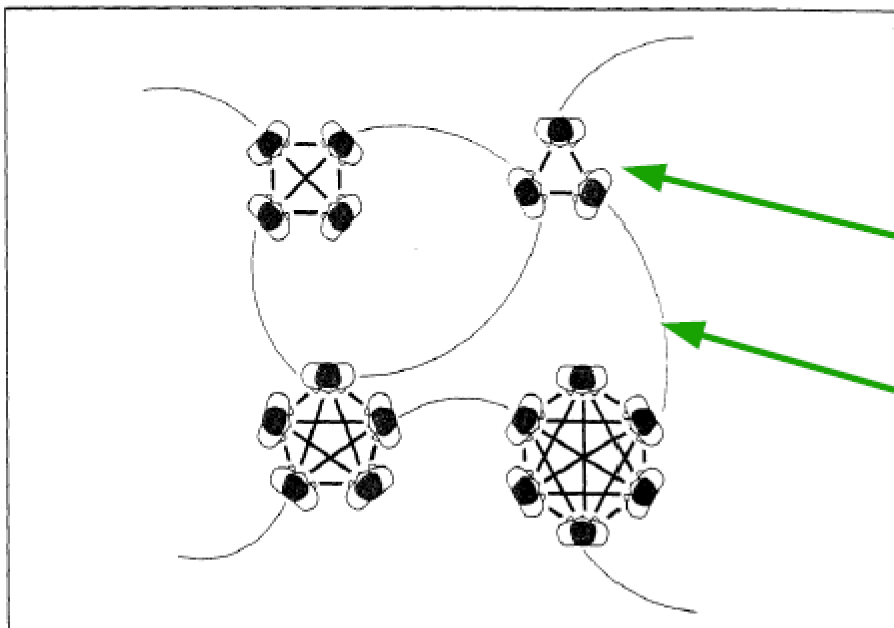


In the Nebraska Study the chains varied from two to 10 intermediate acquaintances with the median at five.

Propriedade Mundo Pequeno

- Em uma rede aleatória, as arestas são criadas aleatoriamente.

- No mundo real, as relações de amizade **não são aleatórias**.
  Mark Granovetter (1970, Harvard) estudo a existência de circulos de amigos



Conexões fortes: Agrupamentos densamente conectados.

Conexões fracas: Pontes para o mundo externo.

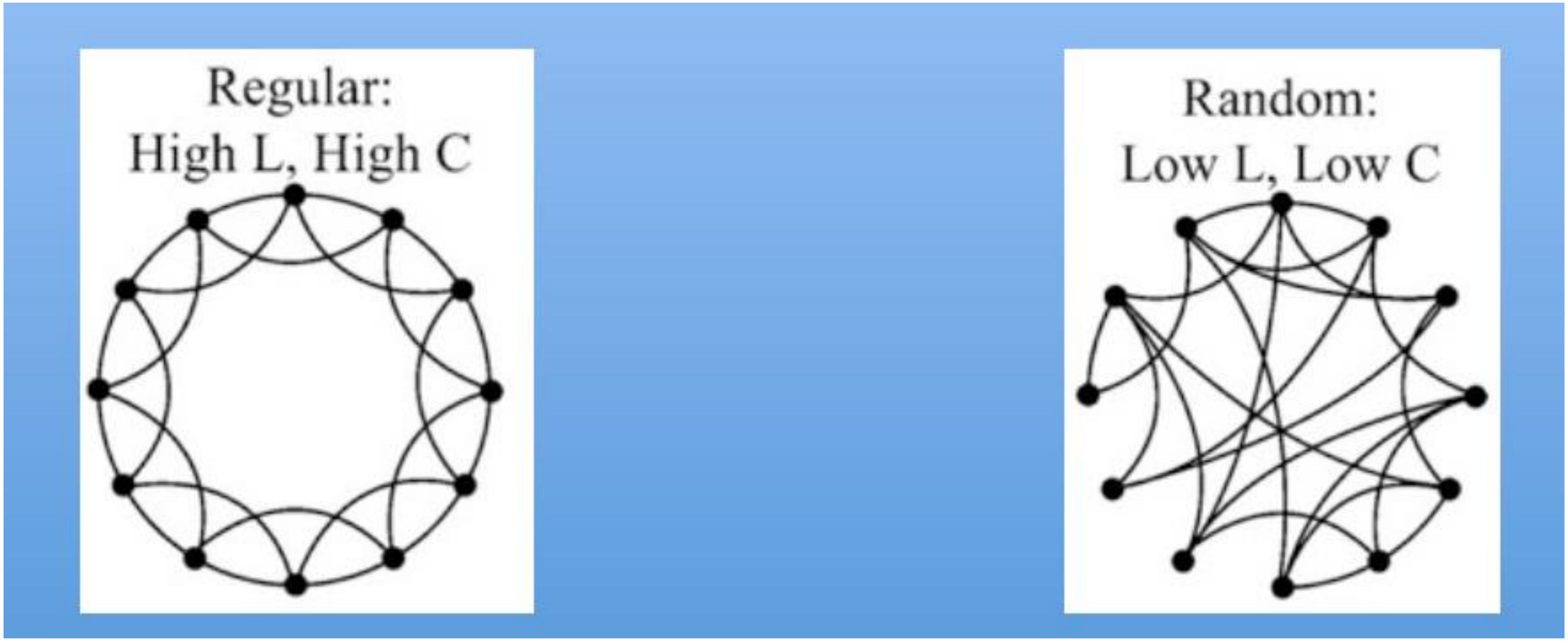
Propriedade Mundo Pequeno

- Modelo de Watts-Strogatz: sugere que algumas pessoas possuem amigos e parentes vivendo não somente próximos a eles, como também alguns **vivendo em lugares mais distantes** e remotos.
- Estas **conexões mais distantes oferecem atalhos entre as pessoas**, diminuindo significativamente a separação média entre elas.
- A presença de alguns **poucos atalhos** é geralmente **suficiente para manter um mundo pequeno**.

Modelo de Watts e Strogatz

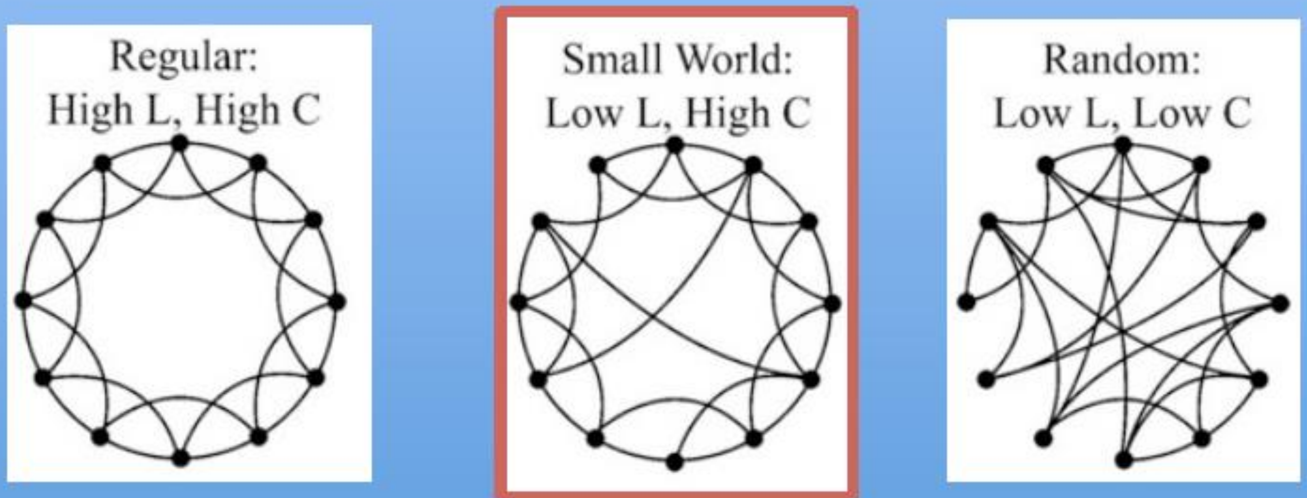
. Redes de atores, power grid, redes neuronais (c. elegans)

São redes com coeficiente de cluster alto (C), mas distância (L) curta.



Modelo de Watts e Strogatz

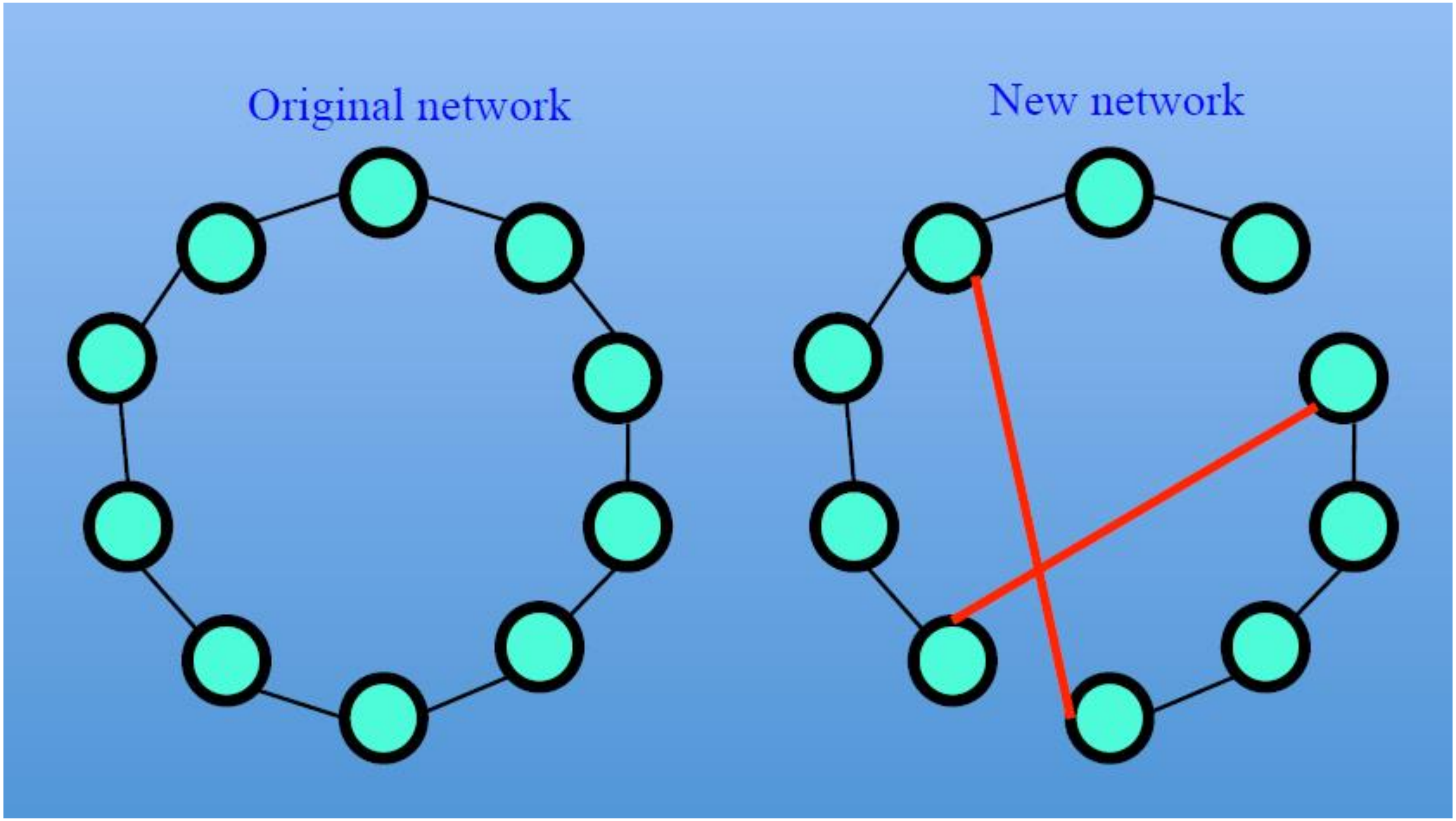
Network Structure



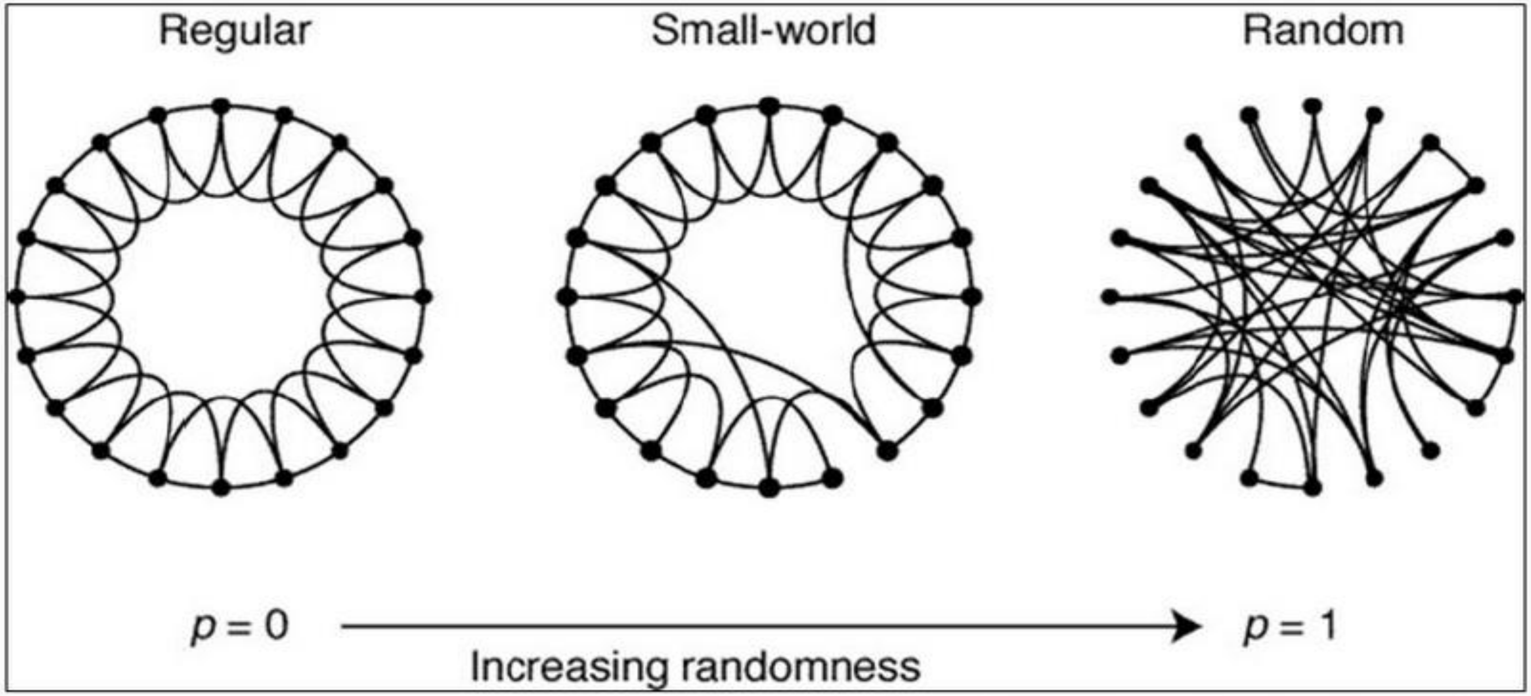
Small-World Network: Has low average distance between nodes, but high clustering

Modelo de Watts e Strogatz

Para construir uma rede pequeno mundo segundo modelo Watts –Strogatz Parte-se de um anel com vértices possuindo dois vizinhos. Com probabilidade p, tire uma conexão existente de um nó qualquer e conecte com outro nó.



Modelo de Watts e Strogatz



D. J. Watts and S. H. Strogatz, Collective dynamics of “smallworld” networks, Nature, 393 (1998), pp. 440–442.

Regular network
- high average distance
- high clustering

Small-world network
- small average distance
- high clustering

Random network
- small average distance
- low clustering

Resumindo...

- Tanto o modelo de Watts-Strogatz como o modelo aleatório de Erdos-Renyi **representam uma sociedade igualitária**, **onde os vértices são muito parecidos entre si**.

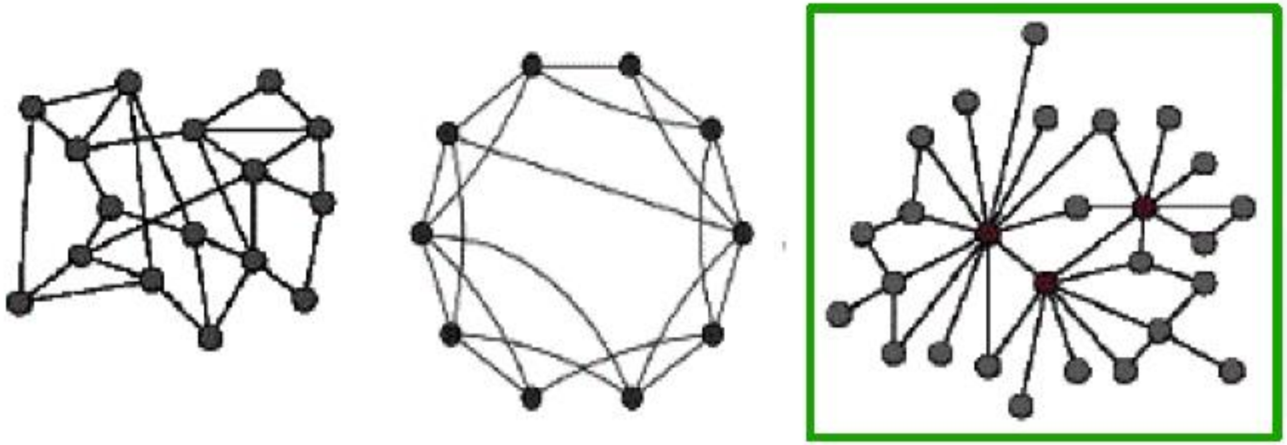
- **Modelo aleatório de Erdos-Renyi:**
  escala para os graus dos vértices, ou seja, eles se concentram em torno de uma média.

- **Modelo de Watts-Strogatz:**
  inicia de um grafo regular (ou anel/reticulado), onde todos os vértices possuem mesmo grau.

Contudo...

Consequentemente, os dois modelos (Erdos-Renyi e Watts-Strogatz) **proíbem a presença de vértices com um grau muito acima da média.**

Vértices com um grau (número de conexões) muito acima da média são chamados de “hubs” (conectores).



Curiosidade

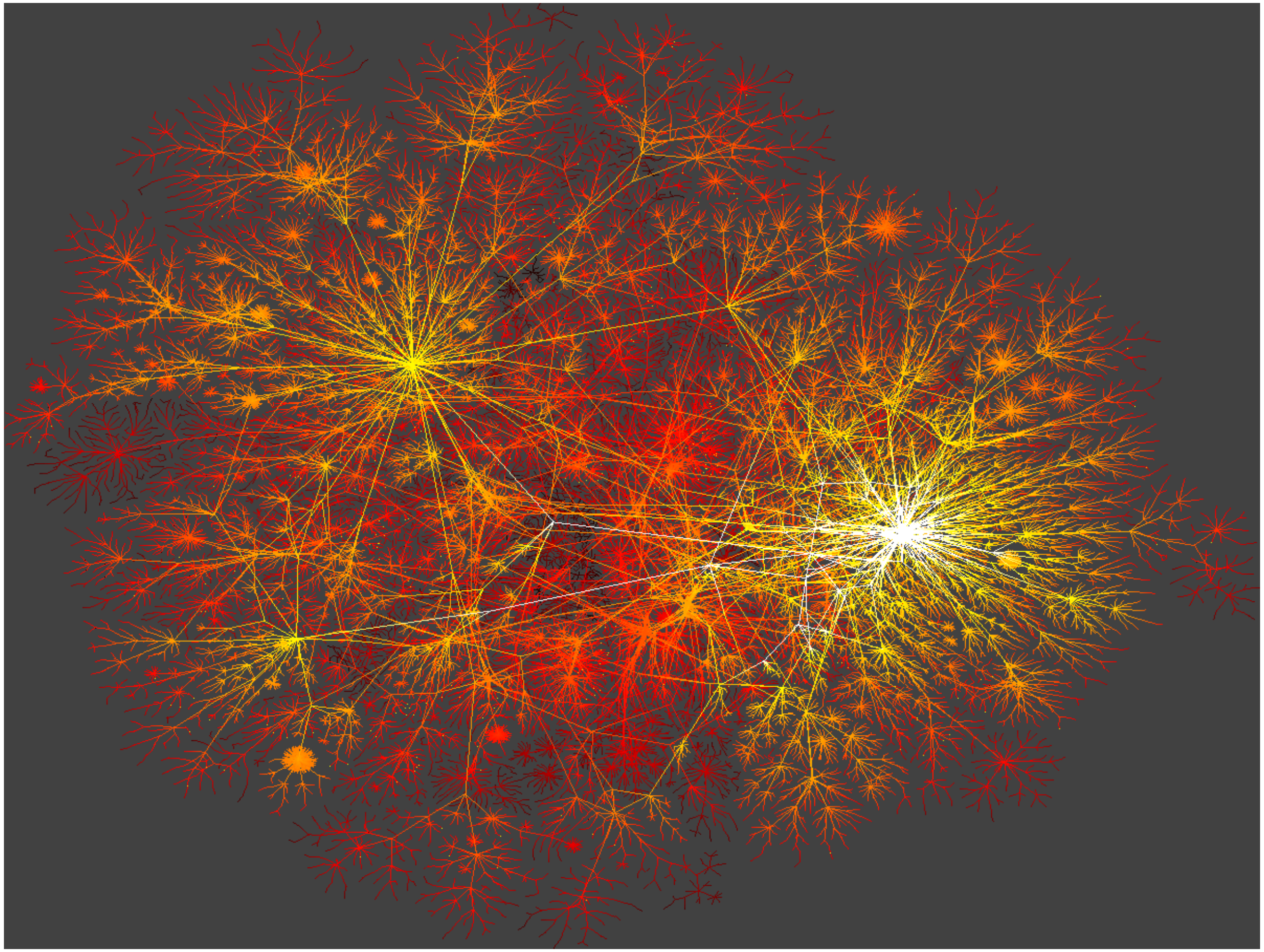
- **Fato I:** Muitas redes possuem propriedades estruturais peculiares (não esperadas)
  - grau, distâncias, clusterização, connectividade, etc
- **Fato II:** Muitas redes differentes possuem propriedades estruturais semelhantes
  - Web, Facebook, AS Graph, Neural network, etc

- "Million dollar question"

Por que?

- Algumas respostas na literatura, mas nada muito definitivo

As Redes Complexas



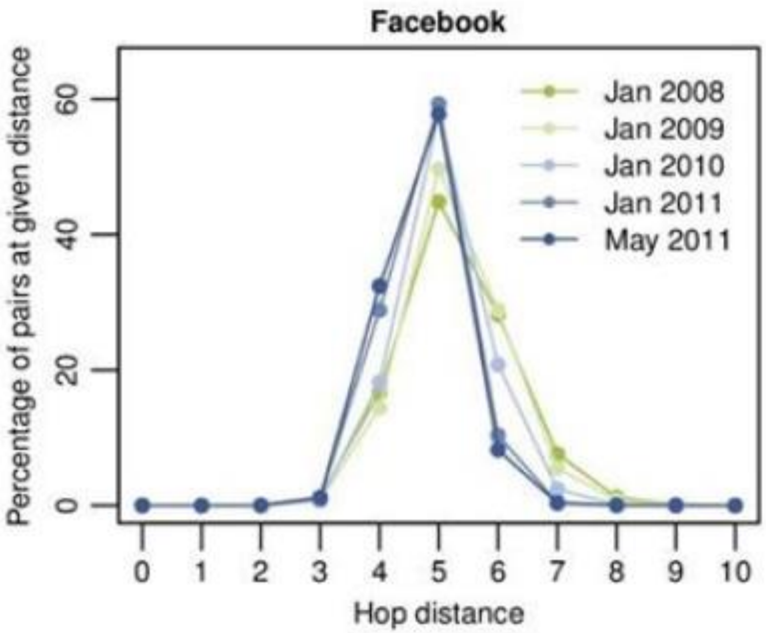
Mundo Pequeno

- *"It's a small world, after all"*
- Distância média **muito** pequena, diâmetro também
- Mesmo para redes muito grandes
  - Web (parte) – 10^8 vértices ← **7.5**
  - Rede de colaboração – 10^6 nodes ← **4.9**
  - Facebook - 10^9 nodes ← **4.5**
    - "seis graus de separação"
  - e muitas outras!

Ordens de grandeza menor!
Aparentemente da ordem de *log n*

Exemplo do Facebook

- Distribuição da distância ao longo do tempo (truncada em 10)



vértices (arestas)

	fb
2007	13.0 M (644.6 M)
2008	56.0 M (2.1 G)
2009	139.1 M (6.2 G)
2010	332.3 M (18.8 G)
2011	562.4 M (47.5 G)
current	721.1 M (68.7 G)

- Rede cresce (50 x), mas distância média diminui!
- current = 12/2011

Anatomy of Facebook, Lars Backstrom et al. 2011

Meus amigos também são amigos

- A relacionado com B e C faz com que B e C se relacionem mais provavelmente
- Rede possui transitividade – caminhos de comprimento dois viram triângulos
  - métrica: coeficiente de clusterização
  - densidade: chance de dois vértices ao acaso estarem relacionados

	clusterização	densidade
AS graph – 10^4 nodes	**0.39**	**0.00056**
Facebook – 10^9 nodes	**0.14**	**0.00000026**
Biology coauthorship	**0.67**	**0.00001**

Ordens de magnitude maior!

Normalidade Ausente

- Grau dos vértices é muito desigual
- Muitos com grau pequeno, poucos com grau muito grande

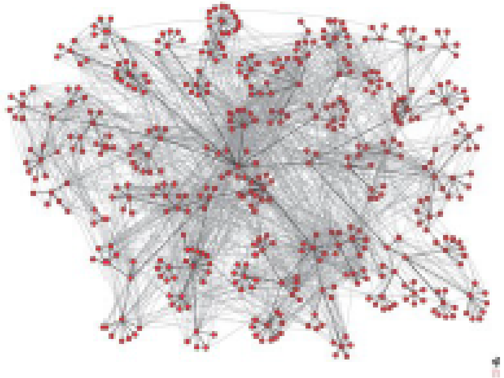
Parecido com distribuição da renda no Brasil!

- AS Graph – 10^4 nodes
  - Grau médio: **5.9**
  - Grau máx: ~**2100**
- Citações – 10^6 nodes
  - Grau médio: **8.6**
  - Grau máx: **~9000**

Distribuição de grau possui cauda pesada
Abrange diversas orders de grandeza

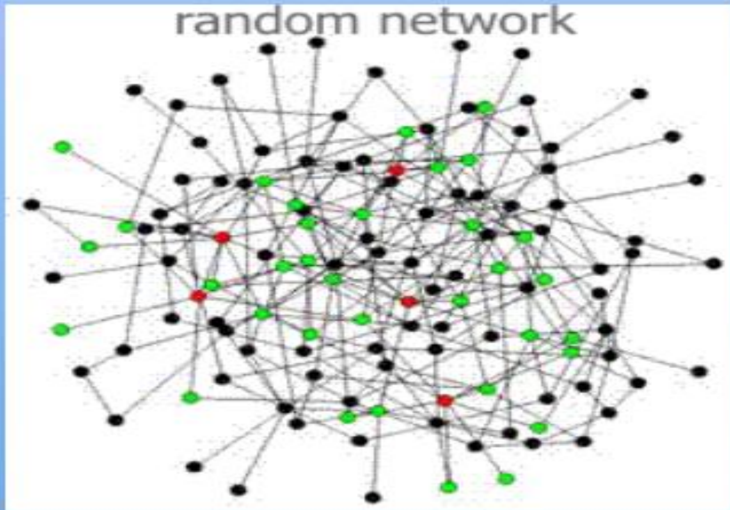
- Muito diferente de distribuição normal

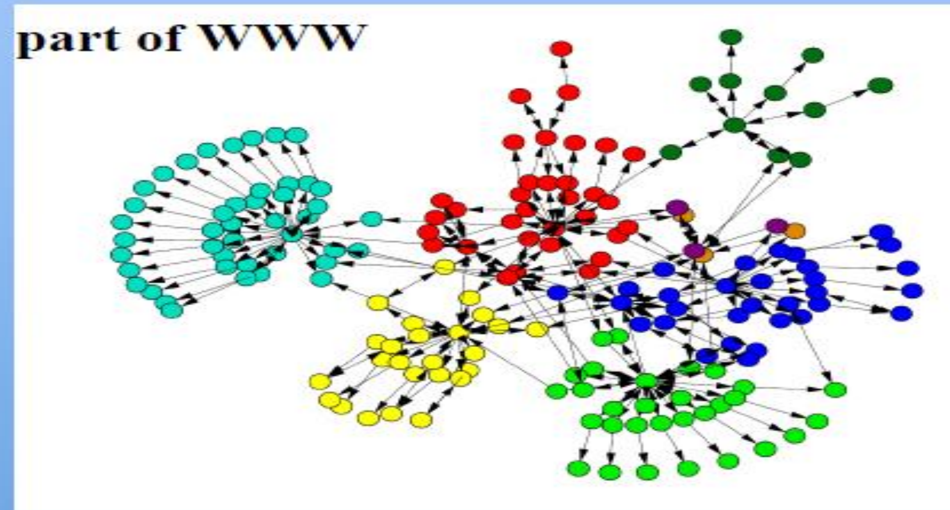
Tudo Conectado

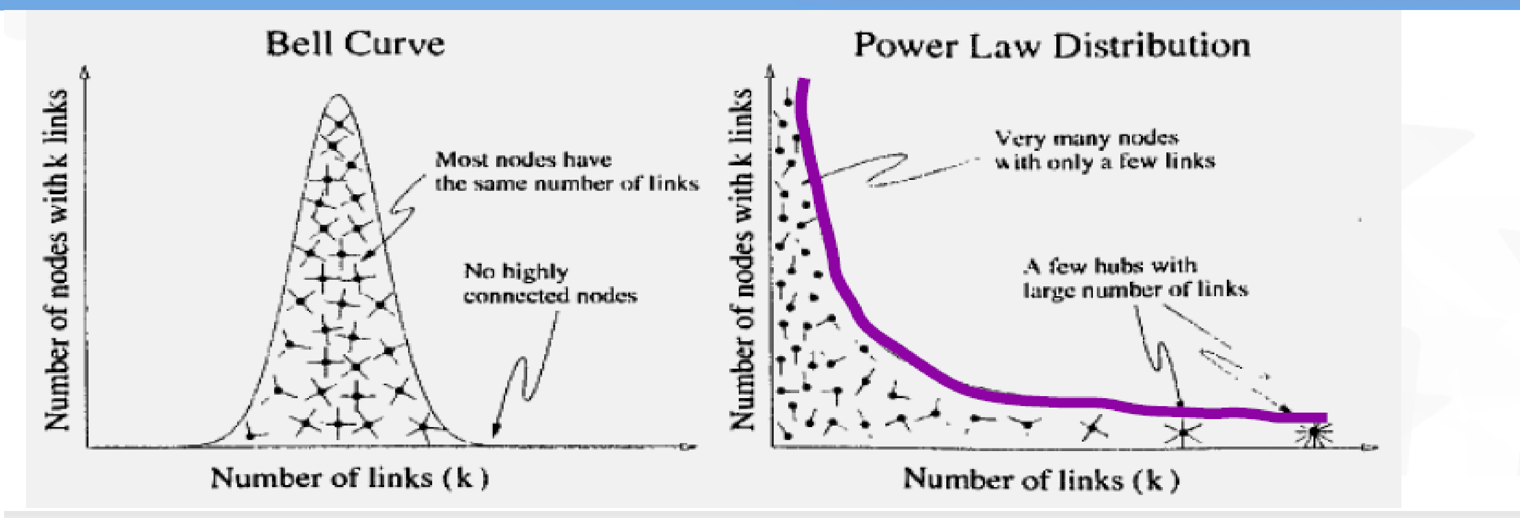


- Maior componente conexa possui **quase todos** os vértices
  - CC gigante
- Outras componentes muito pequenas
  - Muitas outras componentes
- Rede de sinônimos – 23K vértices
  - Maior componente: ~**22K**
- Rede social, rede neural, etc

Quase sempre quase completamente conectada!







Inúmeras redes naturais ou sociais não são descritas pelos modelos de Erdos-Renyi ou Watts-Strogatz. Elas possuem uma distribuição de grau que segue uma lei de potência

$$P(x) = ax^{\gamma}$$

Algumas Propriedades Compartilhadas pelas Redes Complexas

1. Propriedade Mundo Pequeno (Small World)

2. Formação de Aglomerado e Comunidades

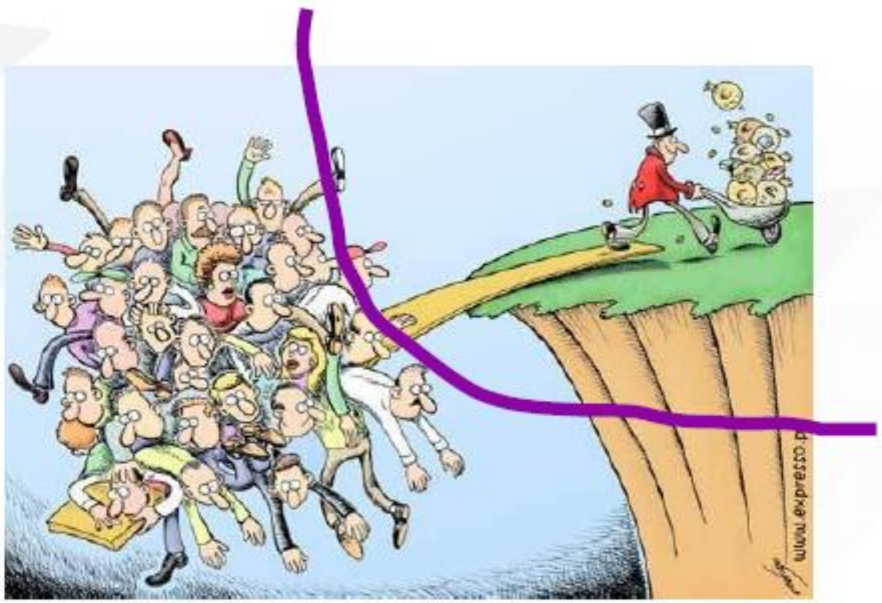
3. Robustez a falhas aleatórias

4. Vunerabilidade a ataques aos hubs

5. Vunerabilidade a falhas em cascata

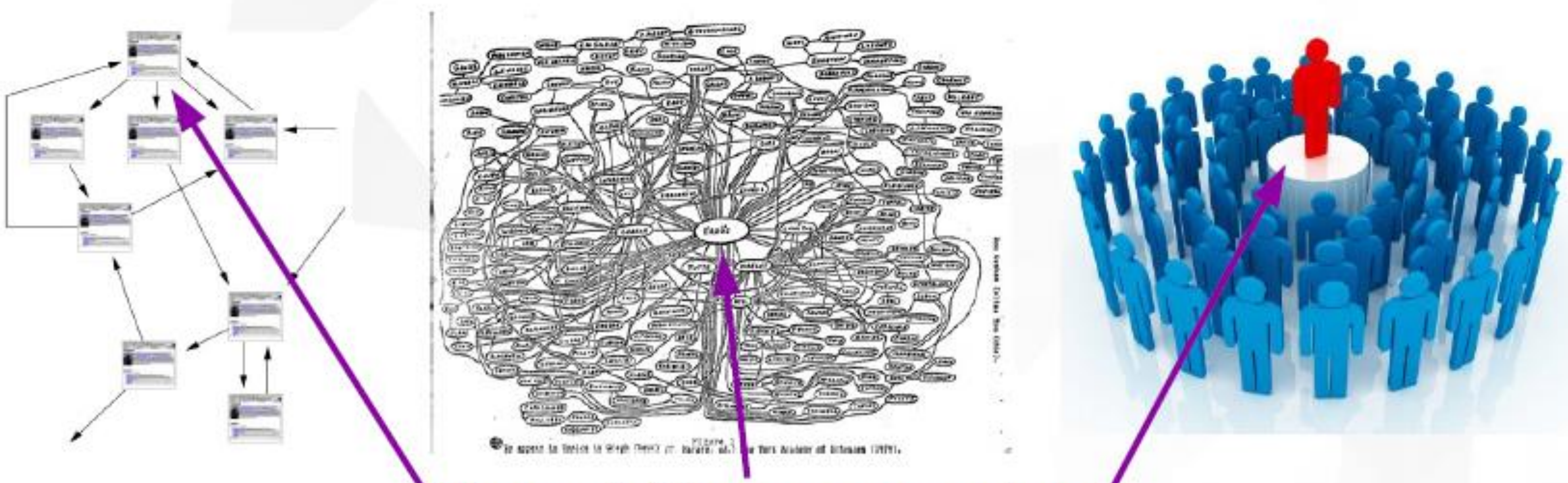
- **Não existe um pico na curva de distribuição de graus**, ou um valor esperado, ou um vértice característico, ou uma média.
- "Desigualdade" e ausência de "democracia":

  Maioria dos vértices possuem poucas conexões, enquanto alguns poucos possuem muitas conexões (hubs).



Uma das características das redes no mundo real é o **crescimento** delas:

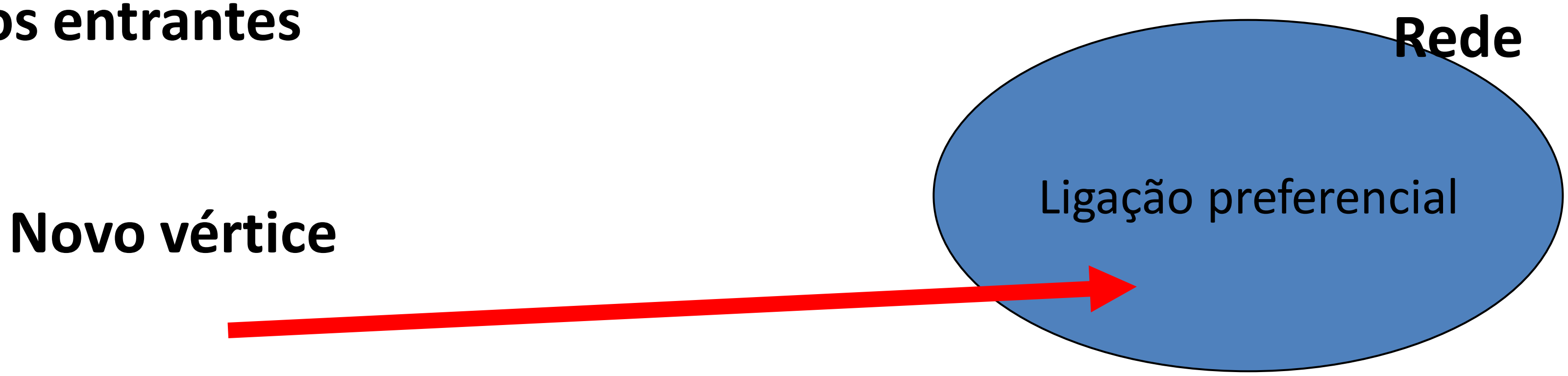
- **Amizade:** novos amigos são inseridos na rede.
- **Web:** a cada dia, novas páginas são adicionadas.
- **Hollywood:** novos filmes são produzidos.
- **Colaboração científica:** novos artigos são publicados.



Uma característica é a presença de hubs

Modelo Albert-Barabasi

Ele parte do suposto que a rede evolui a partir de novos entrantes



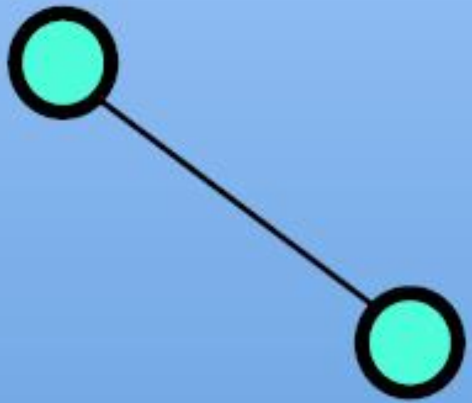
O novo entrante tem preferência por nós mais conectados

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

a) A probabilidade acima é linear em k
b) Ki mede o grau do nó i

Preferential Attachment Model (Netlogo)

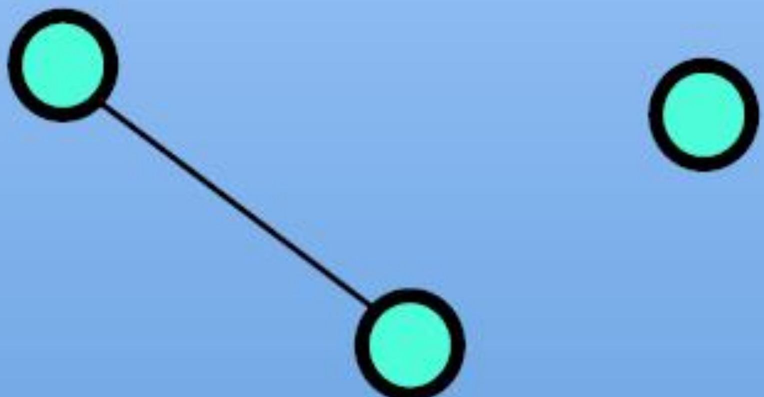
Start with two linked nodes.



At each time step, create a new node.

Preferential Attachment Model (Netlogo)

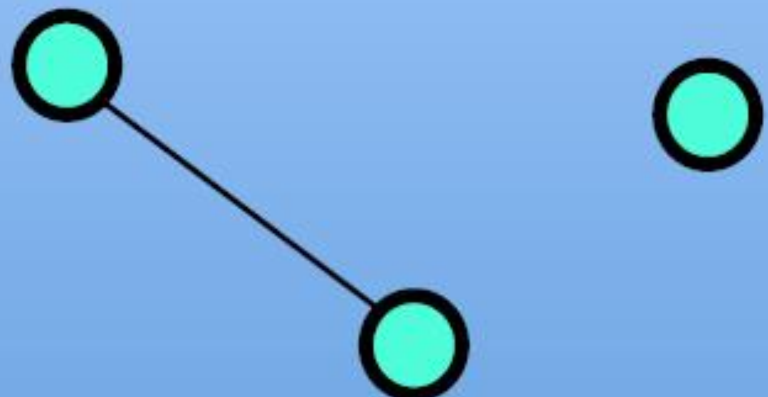
Start with two linked nodes.



At each time step, create a new node.

Preferential Attachment Model (Netlogo)

Start with two linked nodes.

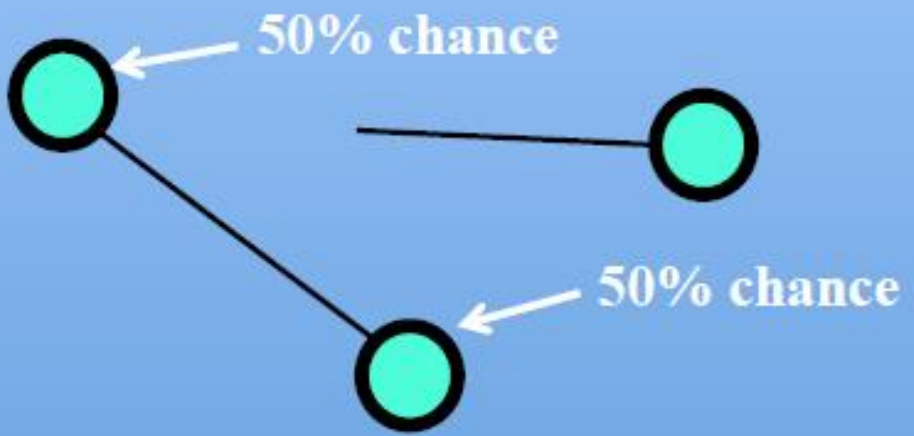


At each time step, create a new node.

This new node randomly chooses an existing node to link to, **but with a bias**: the more links an existing node has, the more likely it is to be chosen.

Preferential Attachment Model (Netlogo)

Start with two linked nodes.



At each time step, create a new node.

This new node randomly chooses an existing node to link to, **but with a bias**: the more links an existing node has, the more likely it is to be chosen.

Preferential Attachment Model (Netlogo)

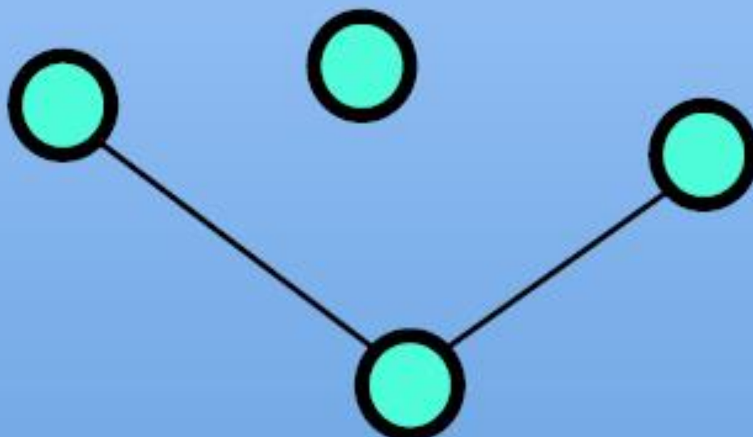
Start with two linked nodes.

At each time step, create a new node.

This new node randomly chooses an existing node to link to, **but with a bias**: the more links an existing node has, the more likely it is to be chosen.

Preferential Attachment Model (Netlogo)

Start with two linked nodes.

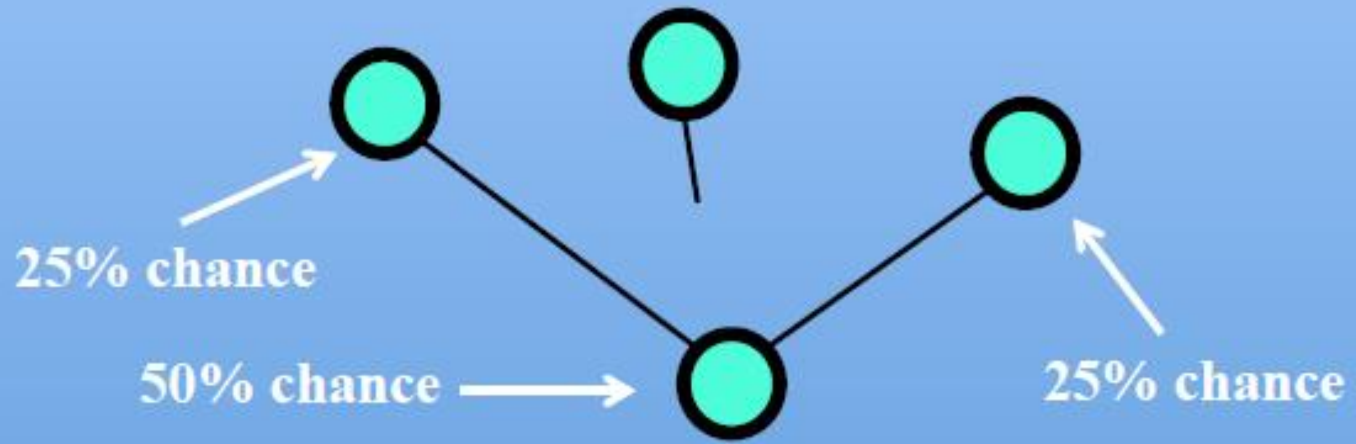


At each time step, create a new node.

This new node randomly chooses an existing node to link to, **but with a bias**: the more links an existing node has, the more likely it is to be chosen.

Preferential Attachment Model (Netlogo)

Start with two linked nodes.

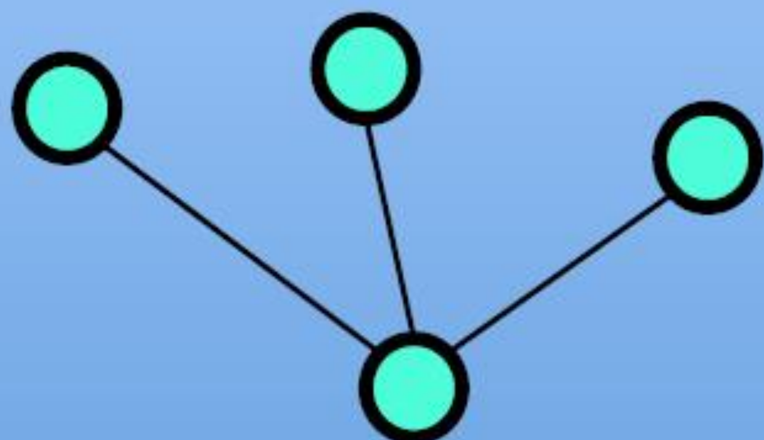


At each time step, create a new node.

This new node randomly chooses an existing node to link to, **but with a bias**: the more links an existing node has, the more likely it is to be chosen.

Preferential Attachment Model (Netlogo)

Start with two linked nodes.



At each time step, create a new node.

This new node randomly chooses an existing node to link to, **but with a bias**: the more links an existing node has, the more likely it is to be chosen.

Modelo A (Barabasi)

Rico-fica-mais-rico: crescimento induz vértices mais antigos se tornarem mais ricos.

- A cada passo, **todos os vértices mais conectatos possuem maior probabilidade de ser escolhido**.
- Mas ao final da construção, não temos uma igualdade de escolha para todos os vértices.
- Os **vértices mais ricos serão os mais antigos**, pois tiveram um maior tempo para coletar links.
- Os **vértices mais pobres serão os mais recentes**, pois tiveram menos tempo de coletar links.

- A lei de Potência é **invariante a escala**.

- Se pegarmos uma amostra da nossa distribuição ela terá o mesmo formato da distribuição completa.

Can Latecomers Make It? Fitness Model

SF model: $k(t) \sim t^{1/2}$ ⟶ (first mover advantage)

Real systems: nodes compete for links -- ***fitness***

Fitness Model: fitness (η)

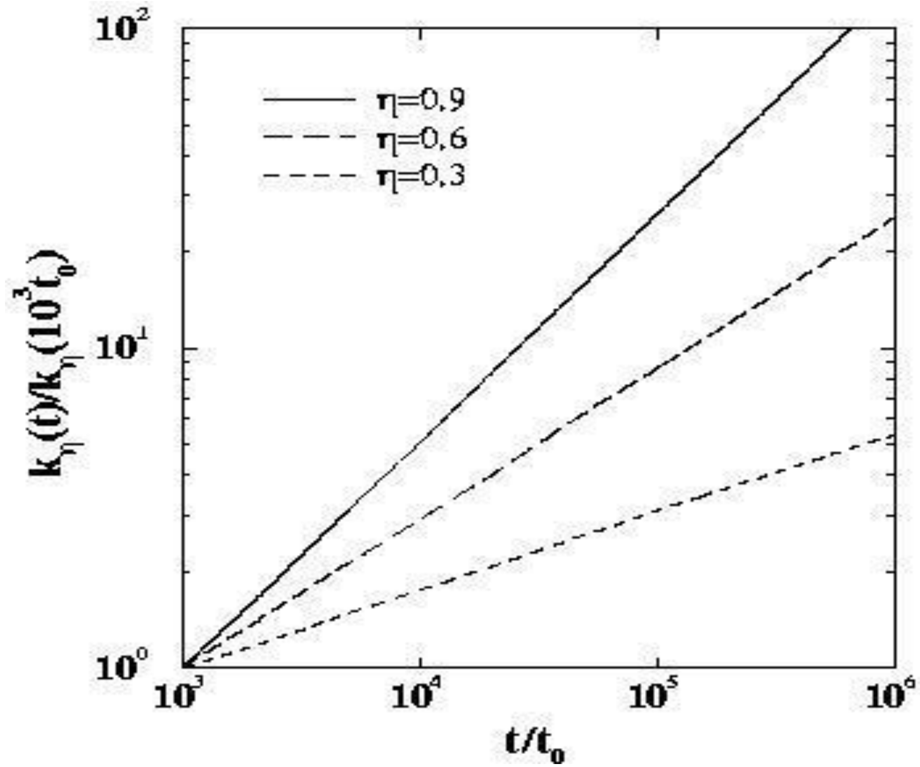
$$\Pi(k_i) \cong \frac{\eta_i k_i}{\sum_j \eta_j k_j}$$

$$\boxed{k(\eta,t) \sim t^{\beta(\eta)}}$$

where

$\beta(\eta)=\eta/C$

$$\int d\eta\, \rho(\eta) \frac{1}{C/\eta - 1} = 1$$



SUMMARY: PROPERTIES OF THE BA MODEL

- **Nr. of nodes:** $N = t$
- **Nr. of links:** $L = mt$
- **Average degree:** $\langle k \rangle = \frac{2L}{N} \rightarrow 2m$
- **Degree dynamics** $k_i(t) = m\left(\frac{t}{t_i}\right)^{\beta}$ $\beta = \frac{1}{2}$ β: dynamical exponent
- **Degree distribution:** $P(k) \sim k^{-\gamma}$ $\gamma = 3$ γ: degree exponent
- **Average Path Length:** $l \approx \frac{\ln N}{\ln \ln N}$
- **Clustering Coefficient:** $C \sim \frac{(\ln N)^2}{N}$

The network grows, but the degree distribution is stationary.

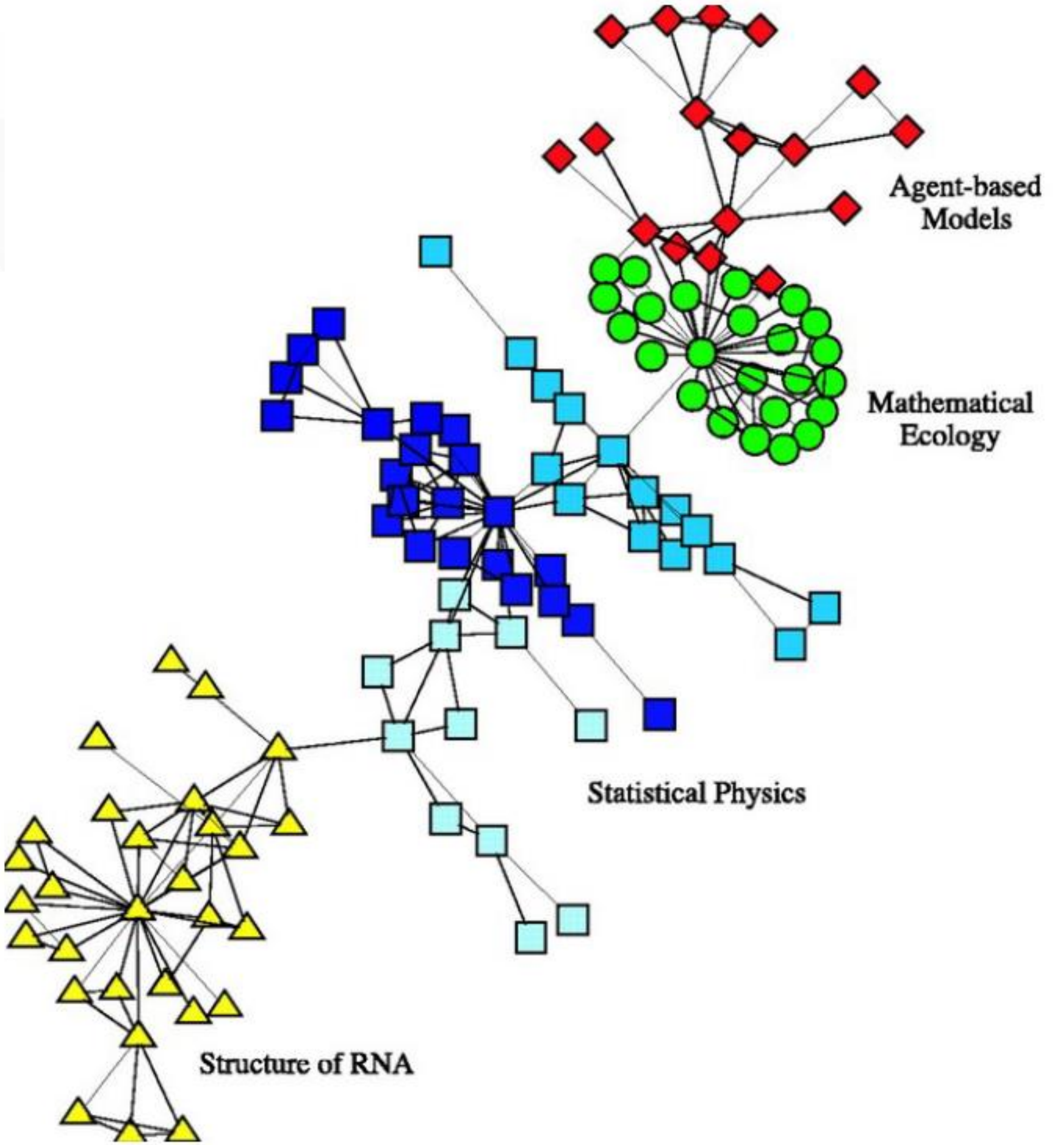
	Pathlenght	Clustering	Degree Distr.
Regular network	$l \gg N^{1/D}$	$C \sim const$	$P(k)=\delta(k-k_d)$
Erdos-Renyi	$l_{rand} \approx \frac{\log N}{\log\langle k\rangle}$	$C_{rand} = p = \frac{\langle k\rangle}{N}$	$P(k) = e^{-<k>}\frac{<k>^k}{k!}$
Watts-Strogatz	$l_{rand} \approx \frac{\log N}{\log\langle k\rangle}$	$C \sim const$	*Exponential*
Barabasi-Albert	$l \approx \frac{\ln N}{\ln\ln N}$	$C \sim \frac{(\ln N)^2}{N}$	$P(k) \sim k^{-\gamma}$

Network Science: Evolving Network Models *February 2012*

Visualização de redes grandes

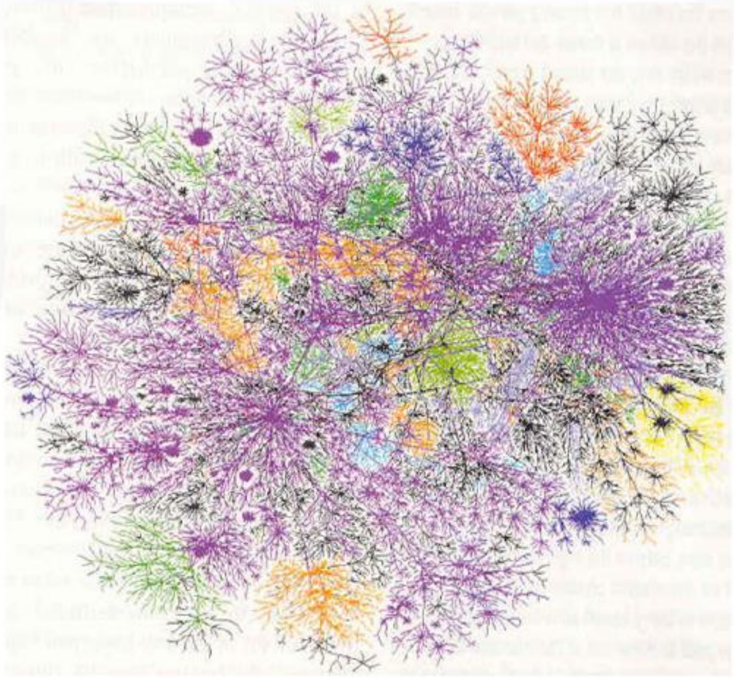
- Em uma rede pequena é possível obter uma interpretação visual.

  (Exemplo 'pequeno' de colaboração entre cientistas.)

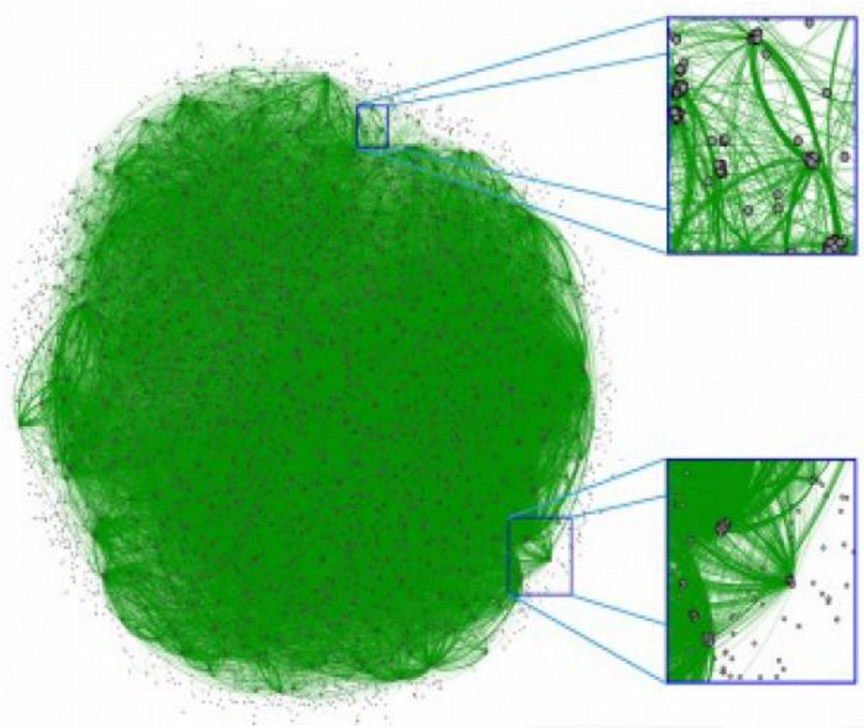


Entretanto para grandes redes

- … é fundamental usar recursos computacionais.



Internet



Pesquisadores de Ciências Exatas

Métricas de Redes

Quais são os nós estruturalmente importantes?

Quais são os nós relevantes para o fluxo de informação?

Centralidade

Como medir a *importância* de um vértice?

- Utilizando apenas a estrutura
- Relativo a outros vértices

- Grau
- Betweeness
- Closeness
- Autovetor
- Random walks
- etc.

- Métricas locais
  - dependem apenas da vizinhança do vértice (Ex. grau, random walk)
- Métrica globais
  - Dependem do grafo inteiro (Ex. Closeness, pagerank)

Centralidade

Qual é a melhor métrica de importância?

- Como determinar a qualidade do ranqueamento produzido?

Impossível sem referência externa!

- Referência externa para avaliar ranqueamento
  - empírica ou processual
- Referência depende do contexto e do objetivo do ranqueamento

Não existe a melhor métrica!

Centralidade de Grau

A **centralidade de Grau** de um vértice é medida pelo número de contatos diretos que ele possui .

- Uma pessoa que se encontra em uma posição que permite o contato direto com muitos outros **é vista pelos demais como um canal maior de informações**, razão pela qual dizemos ser mais central.

$$d_k = \sum_{j=1}^{n} a_{kj} \ ,$$

Centralidade Relativa de Grau

Seja G um grafo qualquer com **n** vértices.

A centralidade Relativa de grau de um vértice k será dada por:

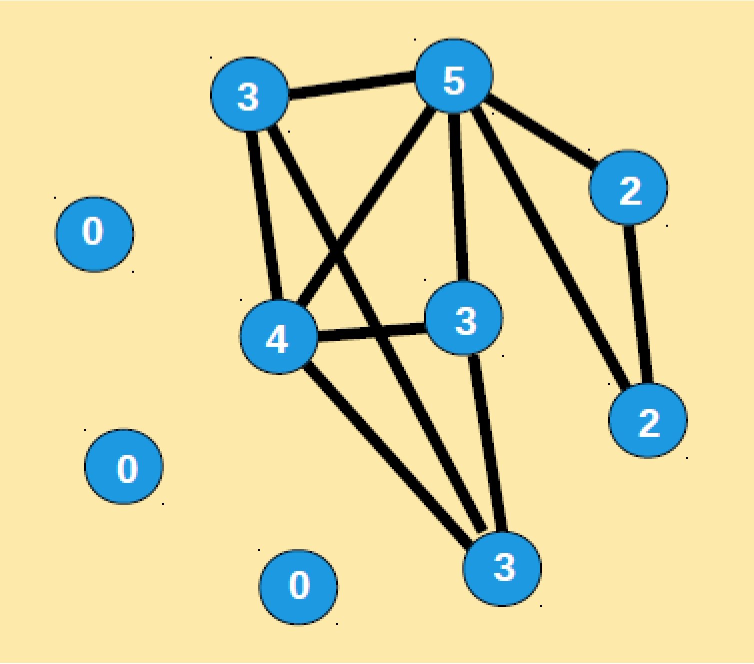
$$c'_D(v_k) = \frac{d_k}{n-1}$$

Essa medida reflete a proporção dos vértices adjacentes a k em relação à ordem do grafo

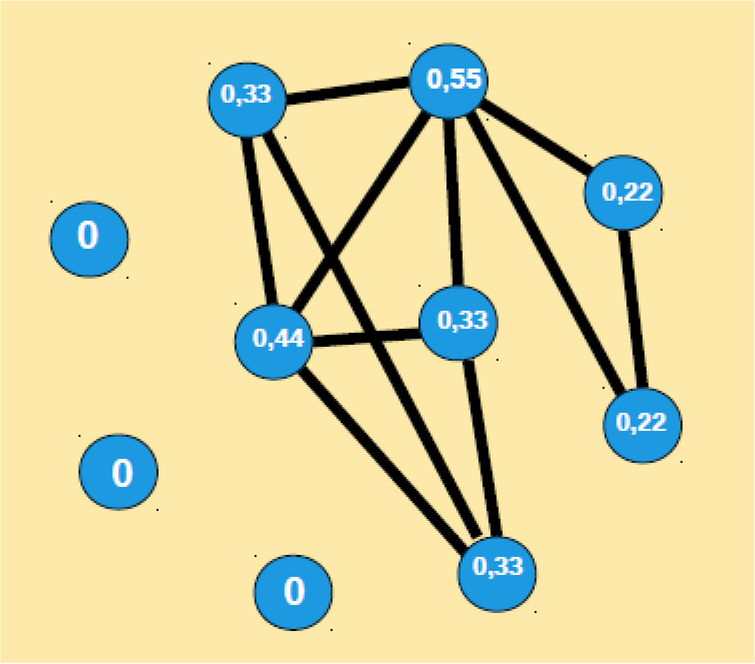
Porque (n-1) no denominador?

Centralidade Relativa de Grau

Centralidade de grau



Centralidade Relativa de Grau



Centralidade de Proximidade

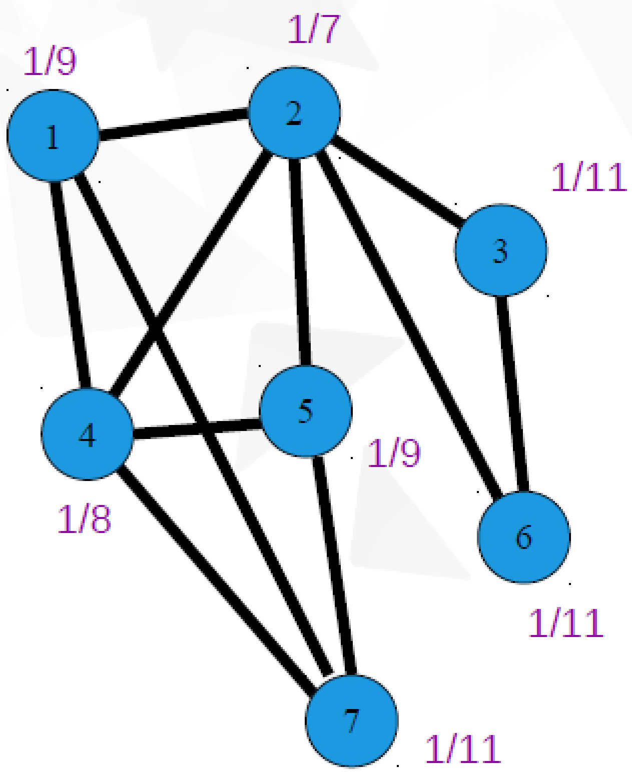
Medida baseada na soma das distâncias de um vértice em relação em relação aos demais vértices do grafo.

$$\text{Centralidade de Proximidade do vértice k } = \frac{1}{\sum_{j=1}^{n} dist(v_j, v_k)}.$$

- "Velocidade" com a qual informação se propara de um vértice para o resto da rede

Centralidade de Proximidade

Maior componente conexa



1-2: 1
1-3: 2
1-4: 1
1-5: 2
1-6: 2
1-7: 1

2-1: 1
2-3: 1
2-4: 1
2-5: 1
2-6: 1
2-7: 2

3-1: 2
3-2: 1
3-4: 2
3-5: 2
3-6: 1
3-7: 3

4-1: 1
4-2: 1
4-3: 2
4-5: 1
4-6: 2
4-7: 1

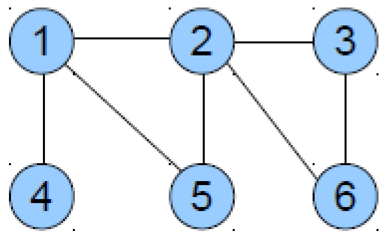
5-1: 2
5-2: 1
5-3: 2
5-4: 1
5-6: 2
5-7: 1

6-1: 2
6-2: 1
6-3: 1
6-4: 2
6-5: 2
6-7: 3

7-1: 1
7-2: 2
7-3: 3
7-4: 1
7-5: 1
7-6: 3

Centralidade de Betweeness

- Mede o quanto no “meio do caminho” um vértice está
- Considerar todos os caminhos mínimos do grafo
- Número de caminhos mínimos que passam pelo vértice
- Exemplo



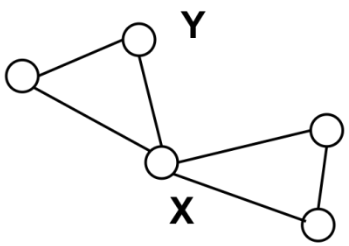
- Grafo completo, K_n?
- Grafo estrela, com *n* folhas?

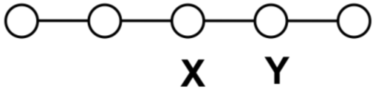
Betweenness ou intermediação

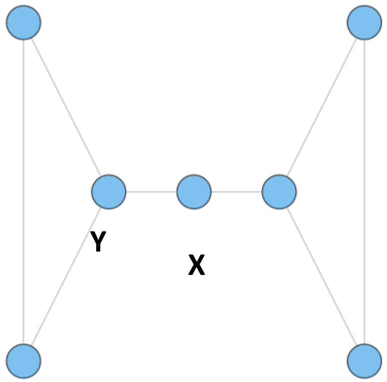
Mede o quanto o nó está no meio do caminho de outros nós. Para isso, considera-se todos os caminhos mínimos que passam pelo nó bem como todos os caminhos mínimos do grafo.

Idéia: quantos indivíduos necessitam de um certo nó para atingir qualquer outro nó com o menor número de arestas?

Quem tem um papel mais importante na intermidiação, X or Y?







Calculando Betweeness

- Mais precisamente

$$C_v = \sum_{s,t \in V;\, s,t \neq v} \frac{\sigma_v(s,t)}{\sigma(s,t)}$$

Número de caminhos mais curtos entre s e t que passam por v

Número de caminhos mais curtos entre s e t

- ou

$$C_v = \sum_{s,t \in V} \sigma_v(s,t)$$

- Pode ser normalizada pelo número total de pares origem/destino (sem contar v)
  - métrica entre 0 e 1

Centralidade do Vetor Próprio (eigenvector)

A centralidade do vetor próprio é dada pela atribuição de pontuações a cada vetor tendo por base o primeiro vetor próprio da matriz de adjacência1. Neste sentido, pontuações elevadas representam indivíduos que estão relacionados com outros, que por sua vez estão interligados a outros indivíduos. Desta forma, representa uma medida de poder e de estatuto do indivíduo na rede. Estes indivíduos são os líderes da rede.

network. The defining equation of an eigenvector is

$$\lambda v = A v$$

where A is the adjacency matrix of the graph, λ is a constant (the eigenvalue), and v is the eigenvector. The equation lends itself to the interpretation that a node that has a

PageRank

O pagerank é uma versão da métrica de centralidade do autovetor. Dado a matrix de adjacência A, a métrica do autovetor é dada por x^t:

$$x_i{}^{t+1} = \sum_{j=1}^{n} A_{ij}\, x_i{}^t$$

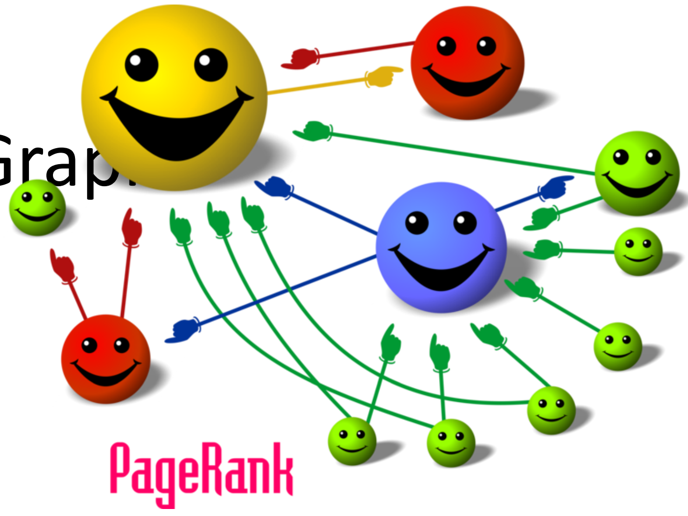
Enquanto para o pagerank é

$$x_i{}^{t+1} = \alpha \sum_{j=1}^{n} A_{ij}\, \frac{x_i{}^t}{k_{out}} + (1 - \alpha)$$

O termo contante α foi adicionado para contabilizar nós que tem indegree zero e outdegree não nulo. O score x_i aparece dividido por k_out para evitar que o score i suba muito devido ao grande número de conexões que eventualmente podem chegam a ele.

Descrição

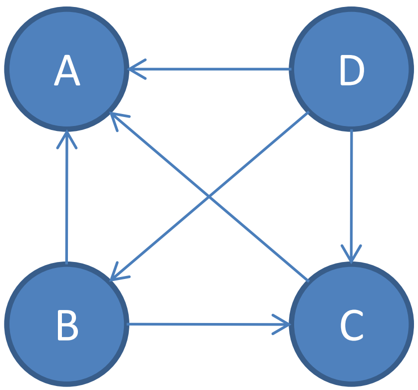
- Para a Google: Páginas (nós) que recebem maior importância (PageRank) devem ter maior chance de estarem entre os primeiros resultados
- É considerada a importância do nó que aponta para outro na hora de calcular a importância do segundo
- O PageRank age sobre um WebGraph
  - Nós são as páginas
  - As arestas são os links



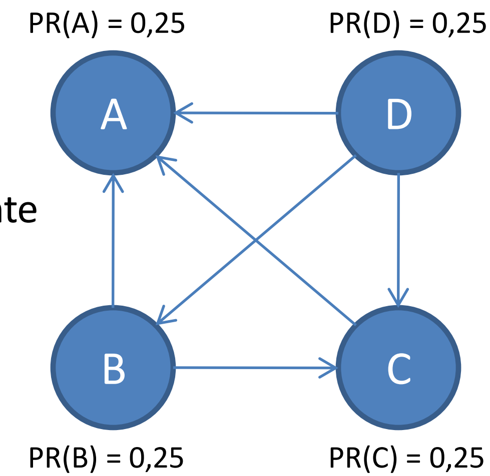
Algoritmo

- O PageRank representa a probabilidade de uma pessoa clicando aleatoriamente chegar a uma página em particular
  - Normalmente, assume-se uma probabilidade inicial para todos os nós, para então ter esse valor refinado em algumas iterações pelo conjunto de nós

Algoritmo Simplificado



Page Rank inicial igualmente dividido PR = 0,25



$$PR(A) = PR(B) + PR(C) + PR(D)$$ = 0,75

$$PR(A) = \frac{PR(B)}{2} + \frac{PR(C)}{1} + \frac{PR(D)}{3}$$ = 0,4583

Temos que levar em consideração que eles não apontam só para A

Generalizando, seja L de um nó a quantidade de arestas saindo que ele tem.

$$PR(A) = \frac{PR(B)}{L(B)} + \frac{PR(C)}{L(C)} + \frac{PR(D)}{L(D)}$$

$$PR(u) = \sum_{v \in B_u} \frac{PR(v)}{L(v)}$$, onde B_u é o conjunto de nós que apontam para u.

Centralidade de Eficiência

Em Pesquisa Operacional, alguns problemas de localização consistem em se determinar um local de modo que **minimize** o tempo máximo de viagem entre o mesmo e todas as demais localizações.

Estes problemas possuem diversas aplicações práticas, como por exemplo, a instalação de um hospital, cujo objetivo é minimizar o tempo máximo de atendimento de uma ambulância a uma possível emergência.

É neste sentido que HAGE e HARARY, em 1995, propuseram uma medida chamada **centralidade de eficiência**, baseada no conceito de **excentricidade de um vértice**.

Centralidade de Eficiência

Seja G um grafo conexo com **n** vértices.

A **Centralidade de Eficiência** do vértice **k** será definida por:

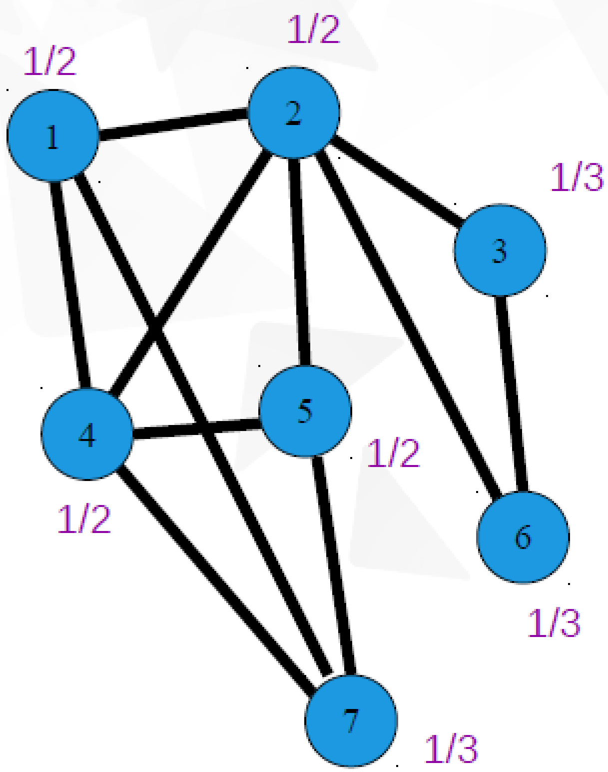
$$\frac{1}{e(v_k)}$$

onde $e(v_k) = \max\{dist(v_j, v_k) : v_j \in V\}$

Esta medida indica que um vértice é mais eficiente quanto menor for a sua excentricidade.

Centralidade de Eficiência

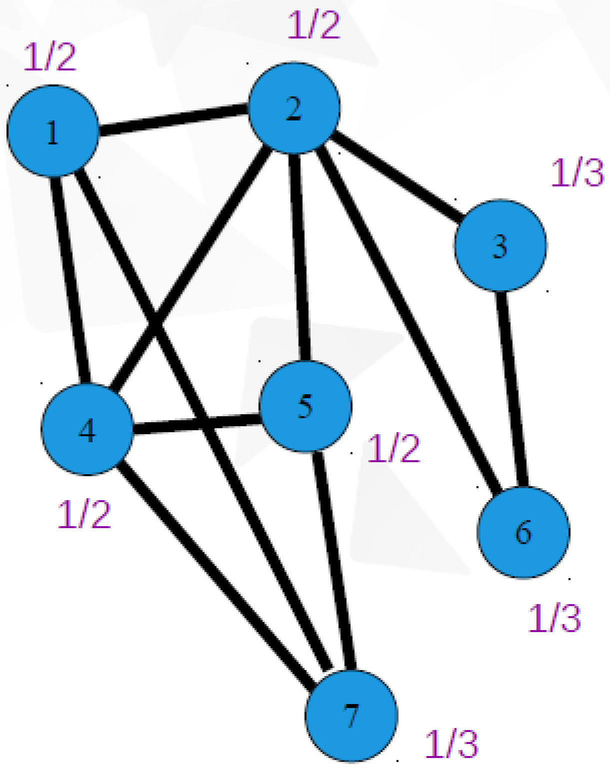
Maior componente conexa

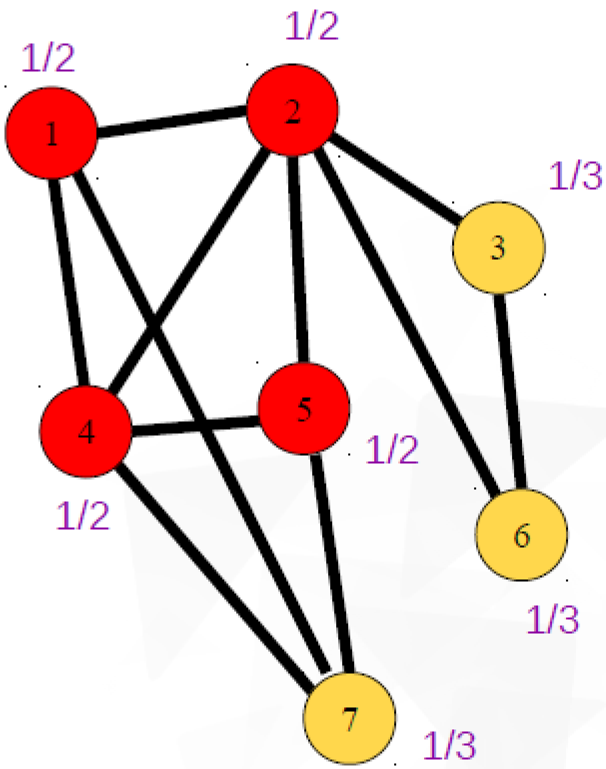


1-2: 1	2-1: 1	3-1: 2	4-1: 1
1-3: 2	2-3: 1	3-2: 1	4-2: 1
1-4: 1	2-4: 1	3-4: 2	4-3: 1
1-5: 2	2-5: 1	3-5: 2	4-5: 1
1-6: 2	2-6: 1	3-6: 1	4-6: 2
1-7: 1	2-7: 2	3-7: 3	4-7: 1

5-1: 2	6-1: 2	7-1: 1
5-2: 1	6-2: 1	7-2: 2
5-3: 2	6-3: 1	7-3: 3
5-4: 1	6-4: 2	7-4: 1
5-6: 2	6-5: 2	7-5: 1
5-7: 1	6-7: 3	7-6: 3

Centralidade de Eficiência

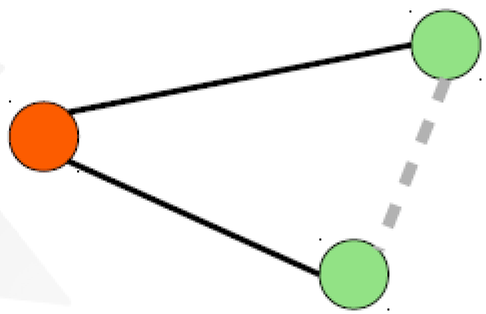




Coeficiente de Clusterização

- Uma característica importante de um grafo é sua correlação das arestas ao redor de um vértice.

- Considere um vértice que está relacionado a outro dois.



Quais as chances destes dois também estarem relacionados?

O Coeficiente de Clusterização captura esta ideia

Coeficiente de Clusterização Local

- O Coef. de clusterização do vértice i é **a fração de arestas que os vizinhos de i possuem entre si** e o máximo entre arestas que eles poderiam possuir entre si.

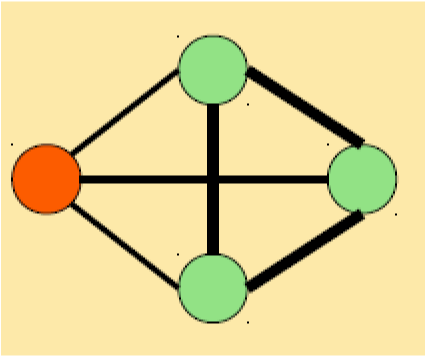
- Dado que o grau do vértice i é di, o maior número de arestas entre seus vizinho é **c(di,2)**.

  Ou seja, todos os pares de vizinhos de i possuem aresta entre si.

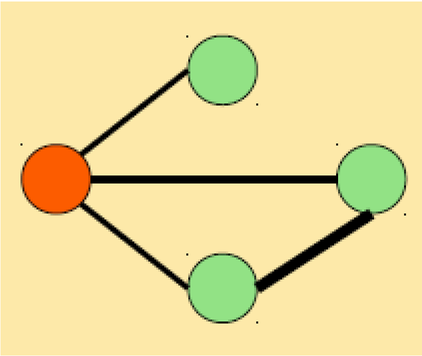
- Seja Ei o número efetivo de arestas entre os vizinhos do vértice i:

Coeficiente de Clustericação = Ei / c(di,2)

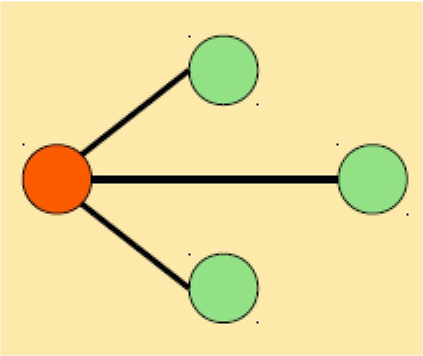
Coeficiente de Clusterização Local



CC = 3/3 = 1

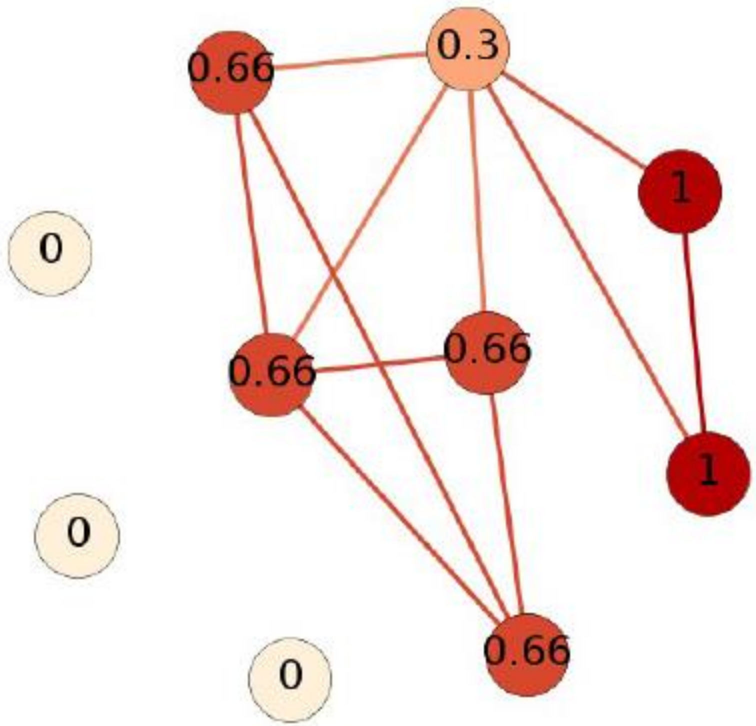


CC = 1/3 = 0,33



CC = 0/3 = 0

Coeficiente de Clusterização do Grafo



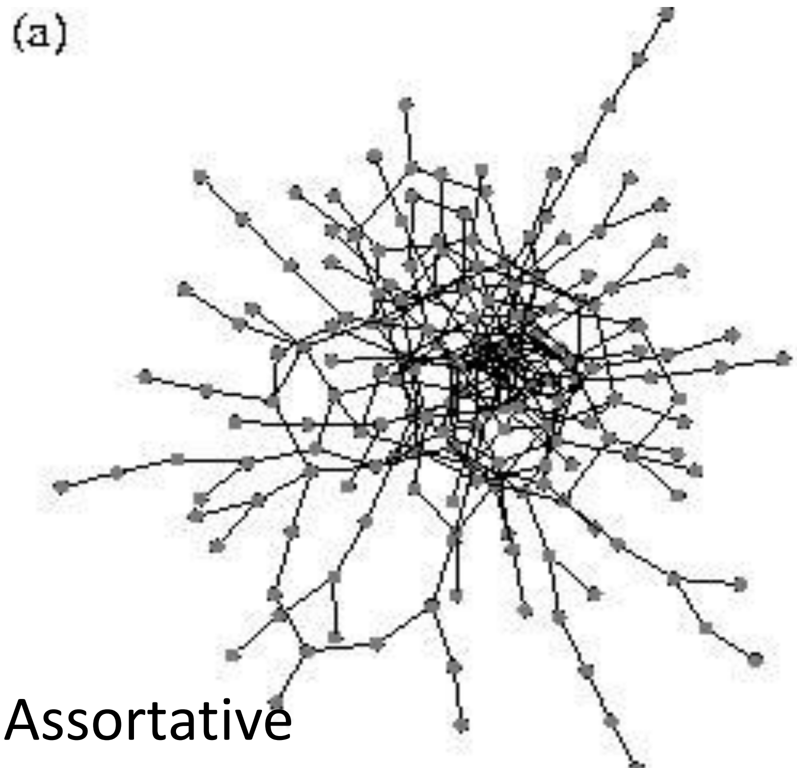
CC = 2Ei/(di(di-1))

O CC não está definido para vértices com grau zero ou um.
Comumente o CC para esses casos é zero.

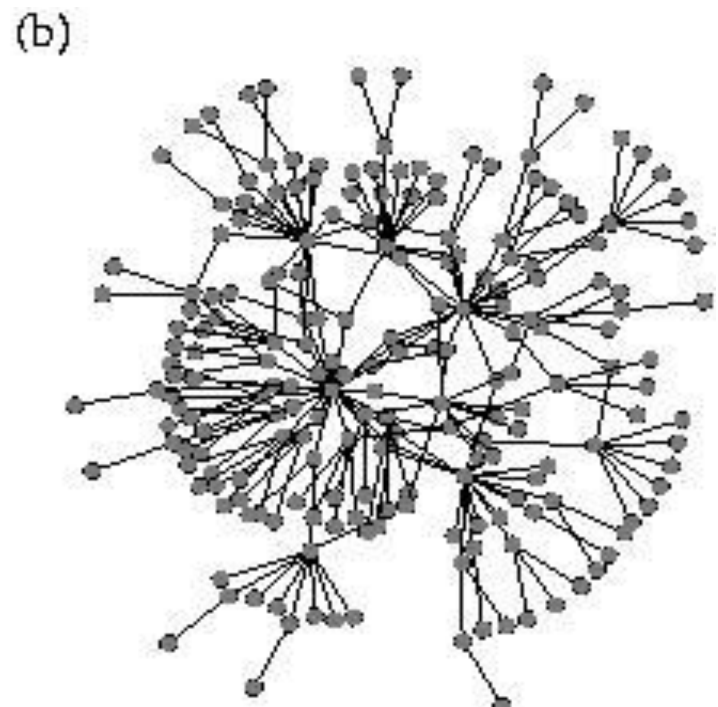
O CC do grafo é a média aritmética dos CC de cada vértice:
1/10*(4,97) = 0,497

Assortatividade (homofilia)

- Rich goes with the rich (selective linking)
  - Um ator famoso prefere emparelhar-se com algum outro ator famoso em um filme, em vez de um recém-chegado na indústria cinematográfica.



Assortative
Scale-free network



Disassortative
Scale-free network

Medida de Assortatividade

$$r = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^{n}(X_i - \bar{X})^2}\sqrt{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}}.$$

$$k_i = \sum_j a_{ij}$$

Similarly, one can define a measure of the assortativity by starting from the correlation function for degrees of vertices *i,j*

$$\langle k_i k_j \rangle - \langle k_i \rangle \langle k_j \rangle = \sum_{k_i, k_j} k_i k_j \left(P(k_i, k_j) - P(k_i) P(k_j) \right)$$

If we introduce the variance

$$\sigma^2 = \sum_k k^2 P(k) - \left(\sum_k k P(k) \right)^2$$

We have the ASSORTATIVITY COEFFICIENT r given by

$$\boxed{r = \frac{1}{\sigma^2} \sum_{k_i, k_j} k_i k_j \left(P(k_i, k_j) - P(k_i) P(k_j) \right)}$$

Medida de Assortatividade

ANND (Média grau vizinho mais próximo)

– Encontre o grau médio dos vizinhos de cada nó i com grau k

– Encontrar a correlação de Pearson (r) entre o grau de i e o grau médio de seus vizinhos

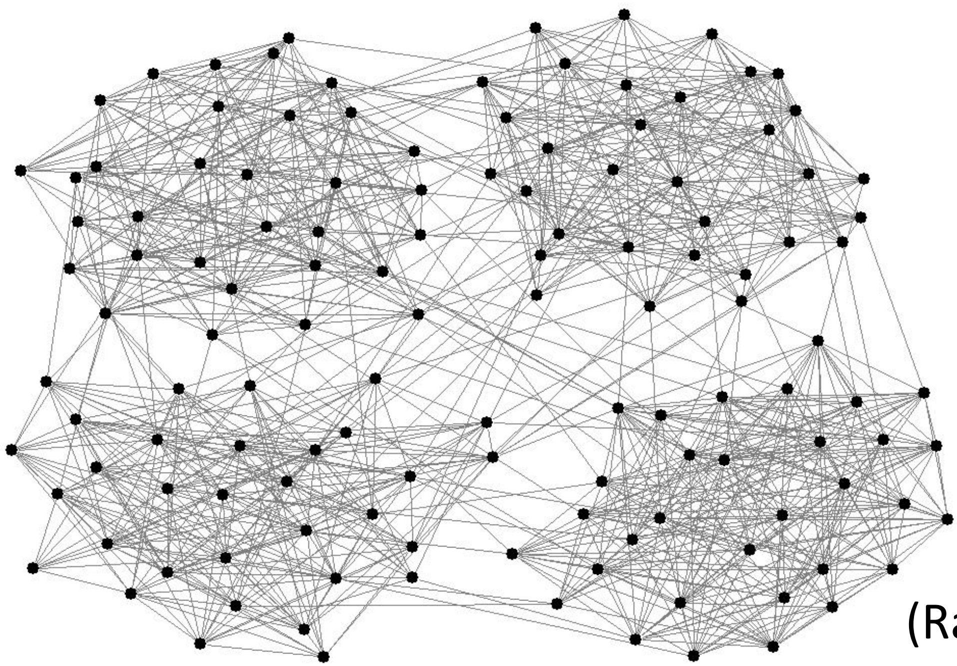
	network	type	size n	assortativity r
social	physics coauthorship	undirected	52 909	0.363
	biology coauthorship	undirected	1 520 251	0.127
	mathematics coauthorship	undirected	253 339	0.120
	film actor collaborations	undirected	449 913	0.208
	company directors	undirected	7 673	0.276
	email address books	directed	16 881	0.092
technol.	Internet	undirected	10 697	−0.189
	World-Wide Web	directed	269 504	−0.067
	software dependencies	directed	3 162	−0.016
biological	protein interactions	undirected	2 115	−0.156
	metabolic network	undirected	765	−0.240
	neural network	directed	307	−0.226
	marine food web	directed	134	−0.263
	freshwater food web	directed	92	−0.326

Community definition

- Definition of Community in a Strong Sense

  The subgraph V is a community in a strong sense if

  $k(i)_{in}(V) > k(i)_{out}(V)$, for all in V

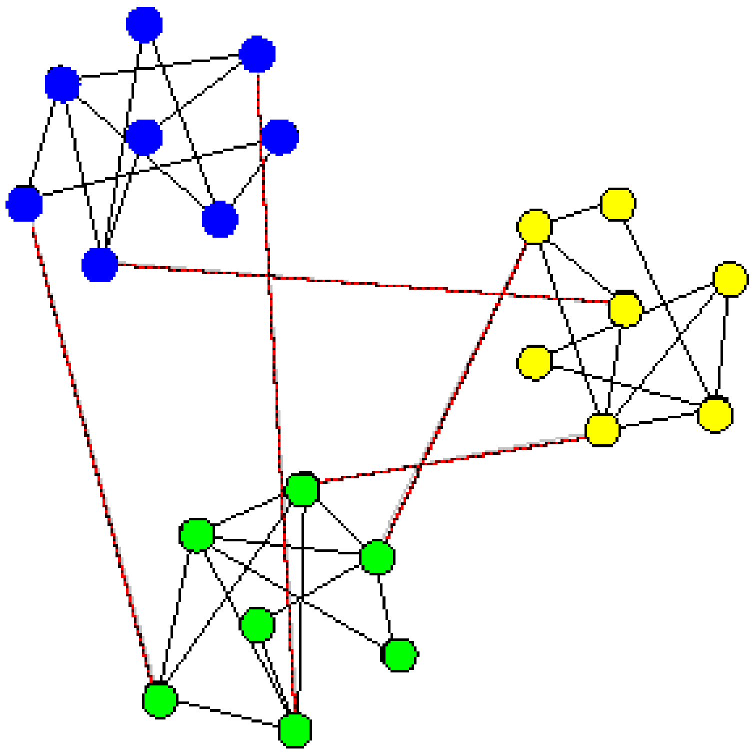


(Radicchi et al. PNAS, 2004).

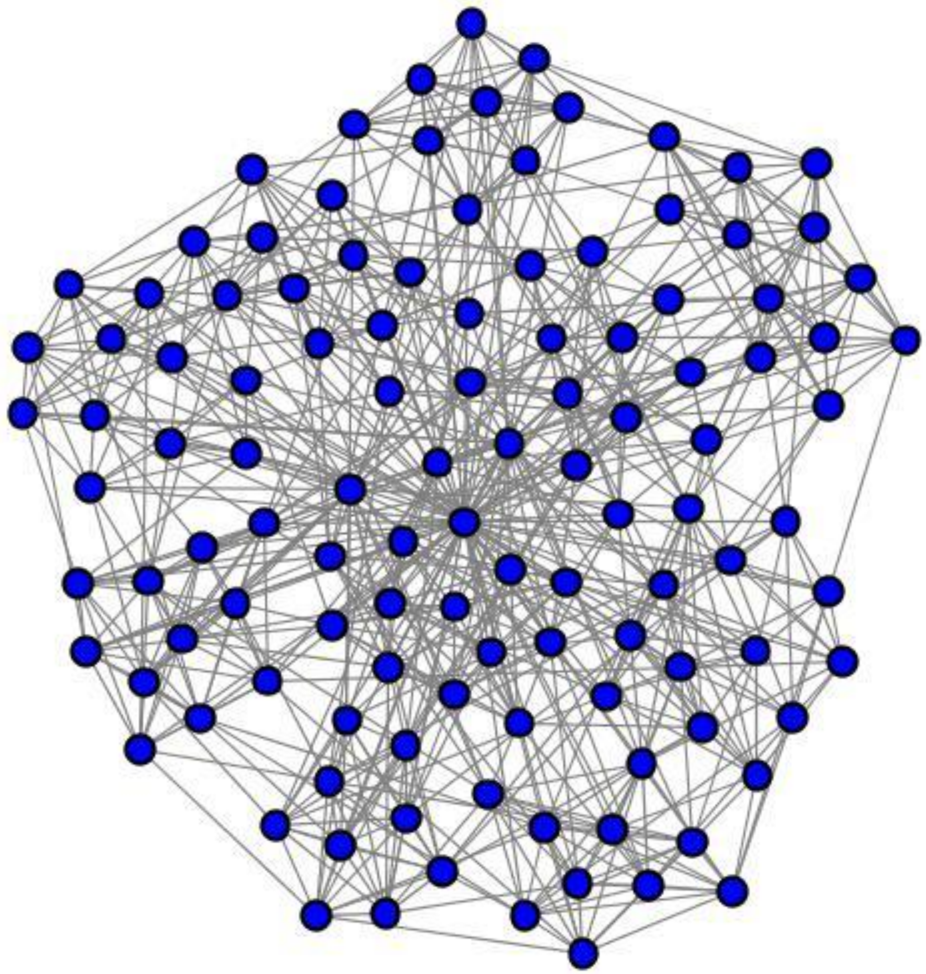
“A group of nodes of a graph which are more strongly connected to each other than with other nodes in the same graph”

Community structure

- Community structure: a group of vertices that have a high density of edges within them and a low density of edges in between groups
- Example:
  - Friendship n/w of children
  - Citation n/ws: research interest
  - World Wide Web: subject matter of pages
  - Metabolic networks: Functional units
  - Linguistic n/ws: similar linguistic categories

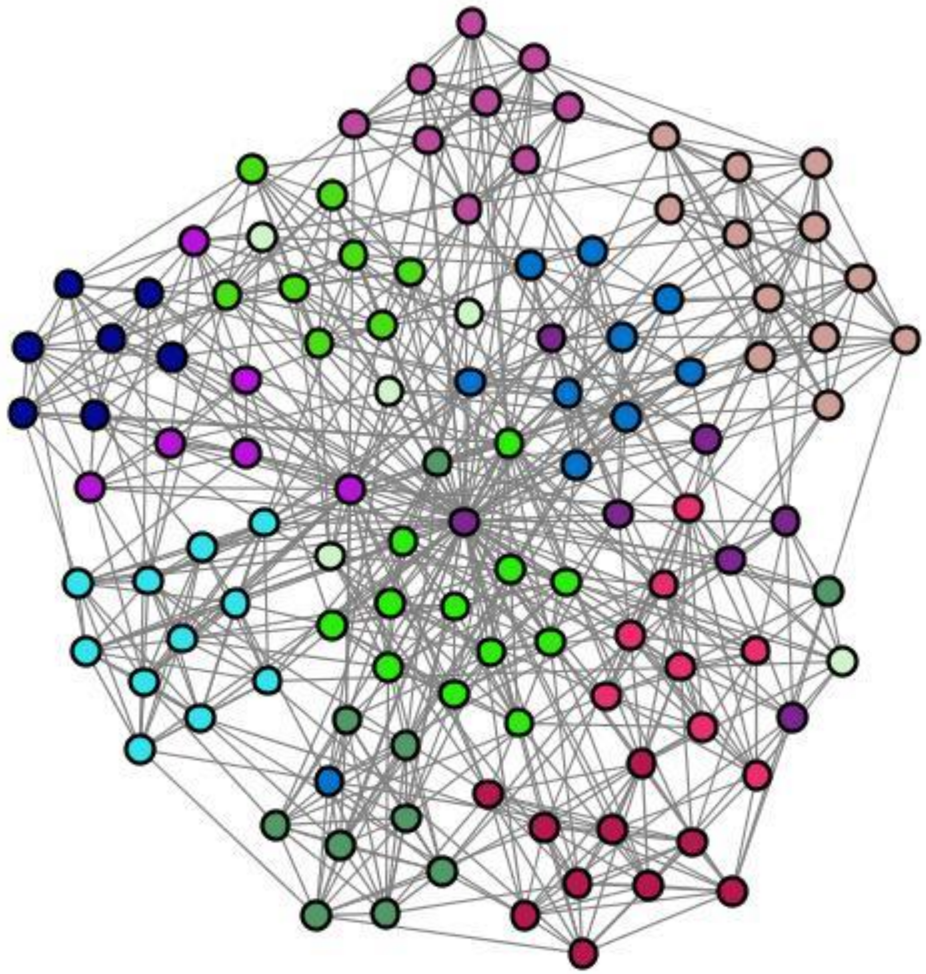


Especially where the community structure isn't apparent or the networks are large



is there community structure?

- Edges: teams that played each other



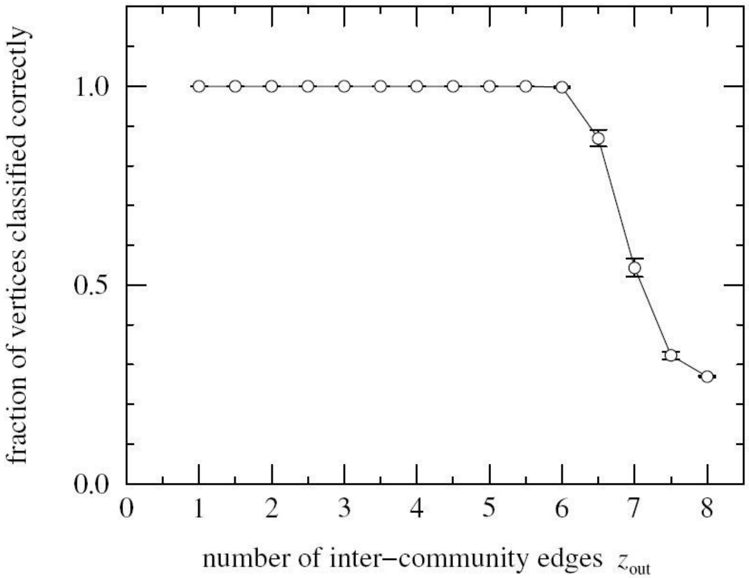
Community Identification Algorithms

- Hierarchical
- Girvan-Newman
- Radicchi et al.
- Chinese Whishpers
- Spectral Bisection

See (Newman 2004) for a comprehensive survey (you will find the ref. in the supplementary material)

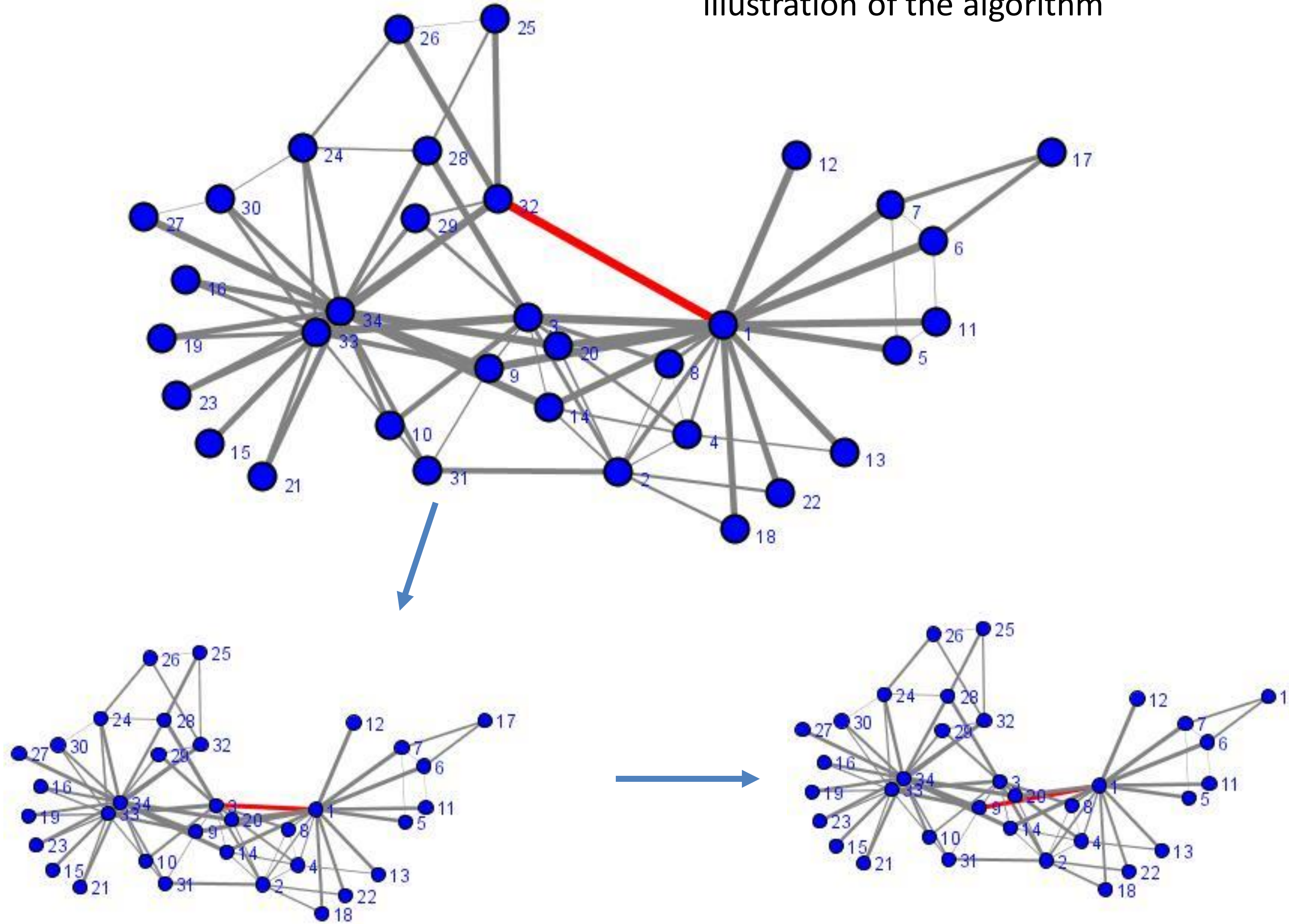
Girvan-Newman Algorithm

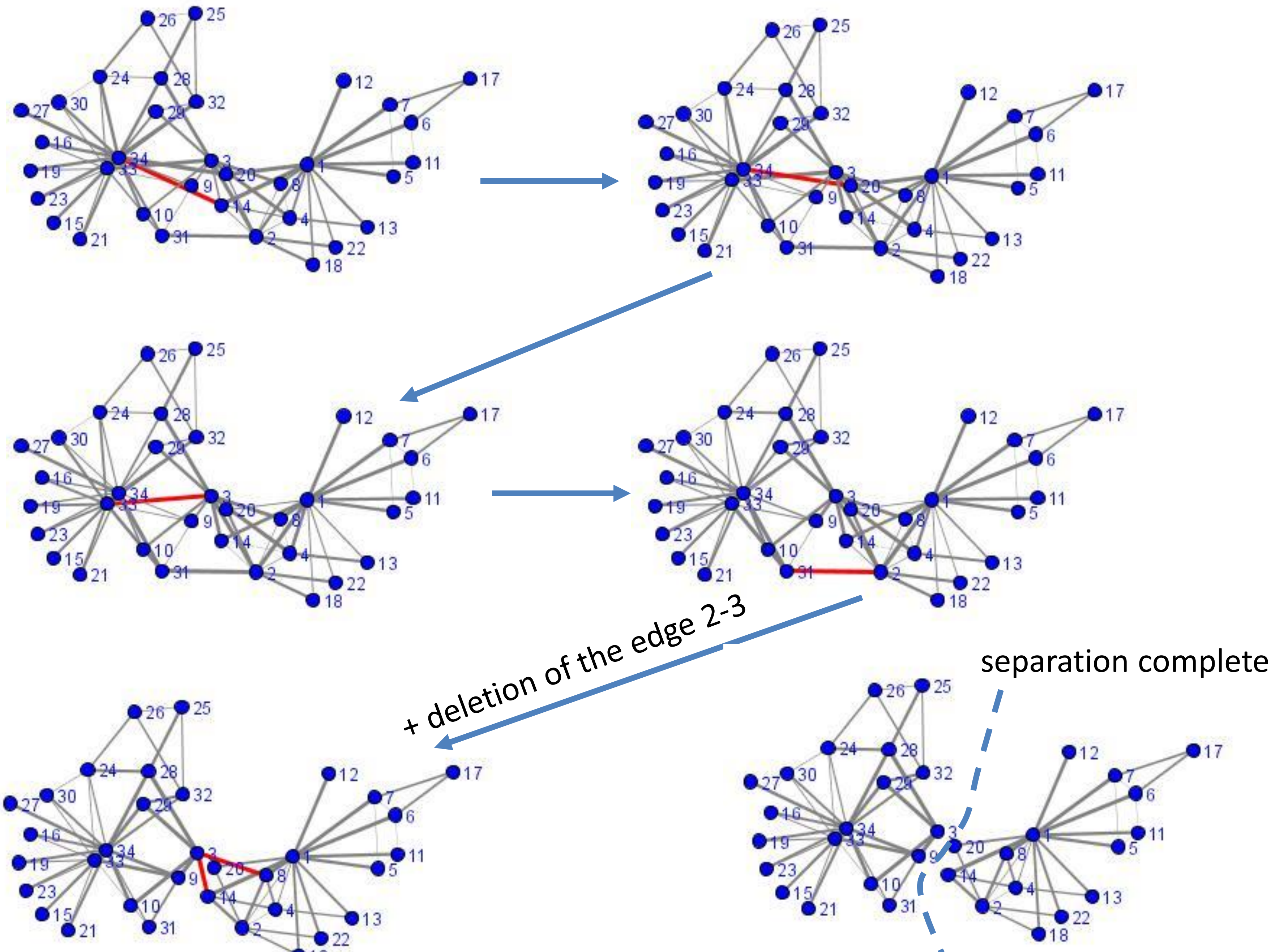
- Bisection Method
  - Calculate the betweenness for all edges in the network.
  - Remove the edge with the highest betweenness.
  - Recalculate betweennesses for all edges affected by the removal.
  - Repeat from step 2 until no edges remain.



Limitation: time of processing $O(n^3)$

(M. Girvan and M. E. J. Newman,PNAS (2002 illustration of the algorithm




+ deletion of the edge 2-3
separation complete

Modularidade

- A possible measure for evaluation of community decomposition is the so–called "Modularity"
- Given an undirected graph $G(E, V)$, where each node is assigned to one of C possible communities, the modularity of the decomposition is defined as (Newman):

$$Q = \frac{1}{2m} \sum_{i,j \in V} \left[A_{ij} - \frac{k_i k_j}{2m} \right] \delta(c_i, c_j)$$

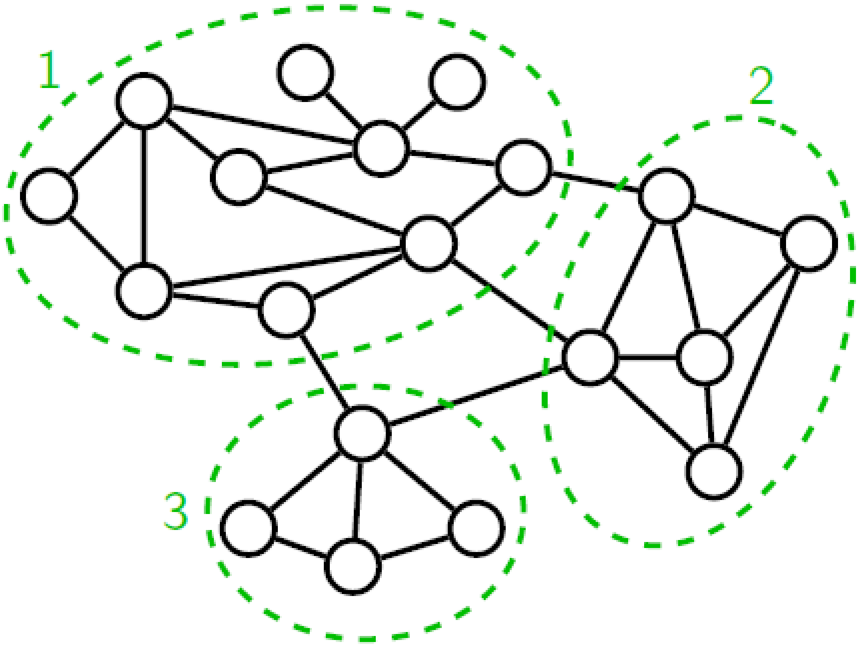
- So the extension of modularity for directed graphs can be written as:

$$Q_d = \frac{1}{m} \sum_{i,j \in V} \left[A_{ij} - \frac{k_i^{out} k_j^{in}}{m} \right] \delta(c_i, c_j)$$

where:

- A_{ij} are the elements of the adjacency matrix of $G(E, V)$
- k_i is the out-degree of node i
- $m = |E|$
- $\delta(c_i, c_j)$ is equal to 1 if i and j belong to the same community, and is equal to 0 otherwise

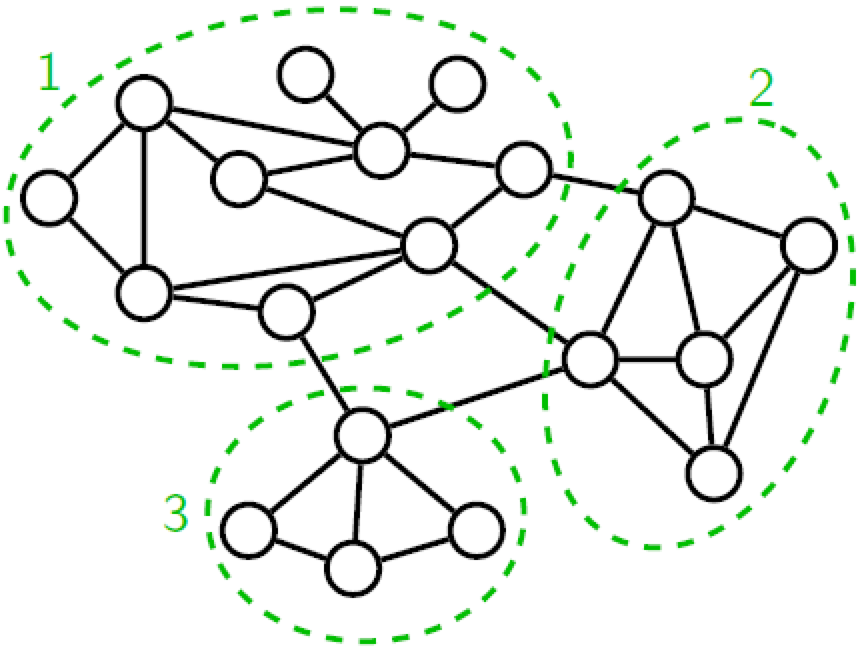
Exemplo



- Dado que identificamos c comunidades na rede, a modularidade Q diz a qualidade da identificação
- Seja E uma matriz de $c \times c$ onde cada elemento e_{ij} indica a proporção de conexões entre vértices da comunidade i e vértices da comunidade j
- Logo definimos a modularidade $Q = \sum_i (e_{ii} - (\sum_j e_{ij})^2)$

Comunidades

Exemplo

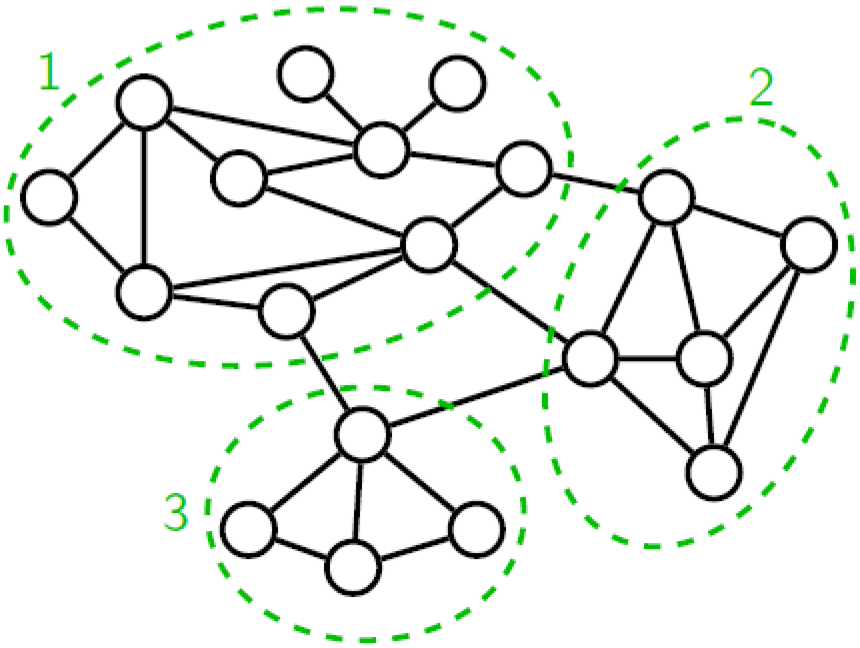


Calculamos $\boldsymbol{E} * \mathbf{31}$

	1	2	3
1	**14**	**2**	**1**
2	**2**	**8**	**1**
3	**1**	**1**	**5**

Comunidades

Exemplo

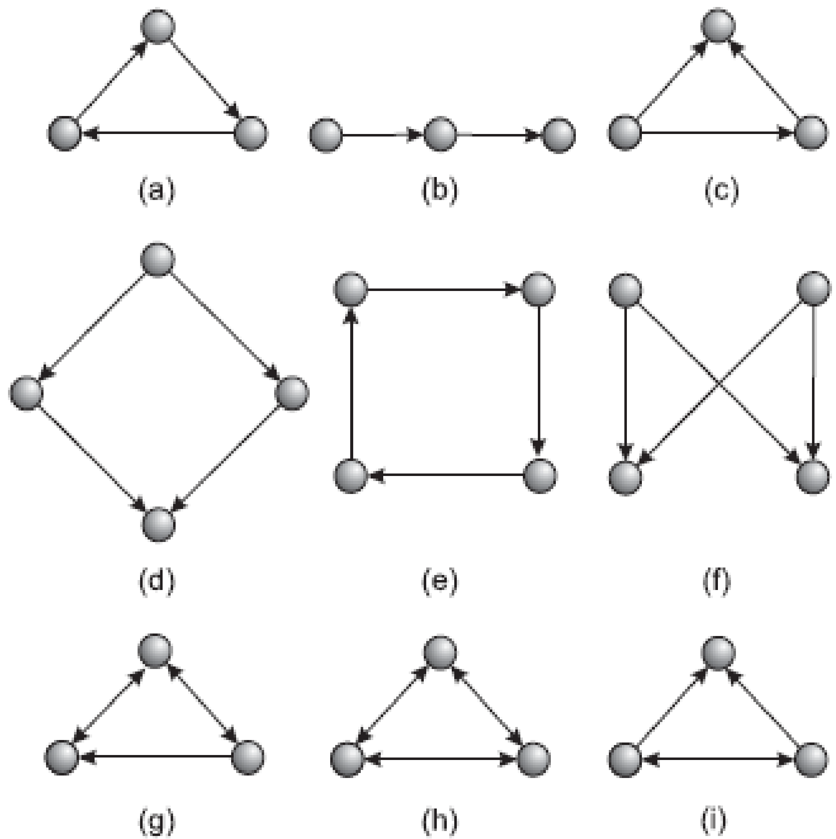


Calculamos $\boldsymbol{E * 31}$

	1	2	3
1	**14**	**2**	**1**
2	**2**	**8**	**1**
3	**1**	**1**	**5**

$$
\begin{aligned}
Q &= \left(\frac{14}{31} - \left(\frac{17}{31}\right)^2\right) + \left(\frac{8}{31} - \left(\frac{11}{31}\right)^2\right) + \left(\frac{5}{31} - \left(\frac{7}{31}\right)^2\right) \\
&= 0.39334027
\end{aligned}
$$

Motivos (Motifs)



Conjunto de nós que exibe padrões cuja presença pode estar relacionado com algum papel de natureza estrutural ou funcional

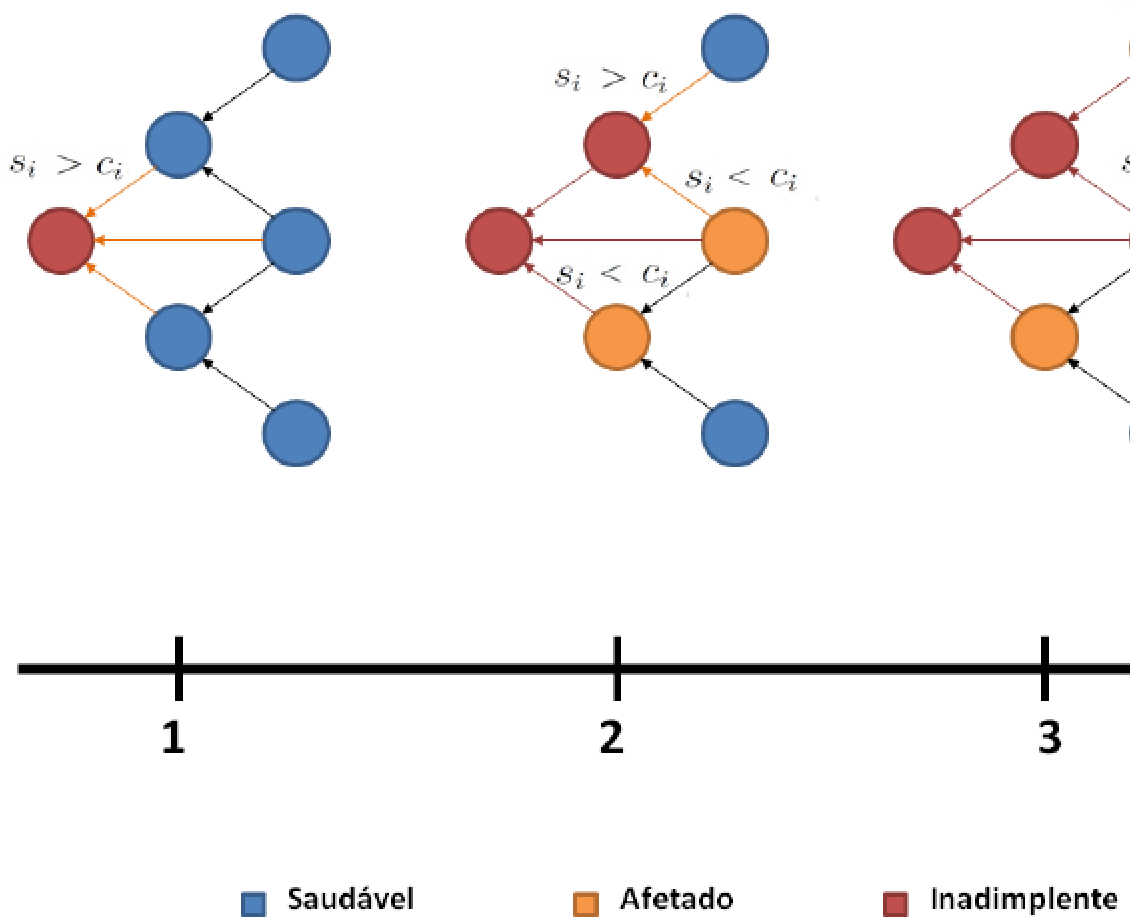
Ajuda na busca de princípios organizadores gerais

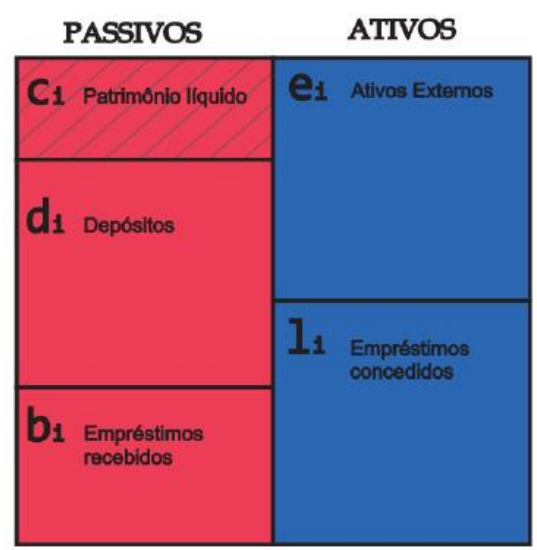
Métrica usada
Para detectar
Motifs é o z_score

Figura 3.5: Exemplos de motivos: (a) *three-vertex feedback*, (b) *three chain*, (c) *feed-forward loop*, (d) *bi-parallel*, (e) *four-vertex feedback*, (f) *bi-fan*, (g) *feedback with two mutual dyads*, (h) *fully connected triad* e (i) *uplinked mutual dyad*. Mantivemos os nomes em inglês conforme encontrados na literatura. Figura adaptada de [dFCRTB07].

Nesta seção, achar exemplos de aplicações de redes!

Modelos Dinâmicos em Redes Complexas

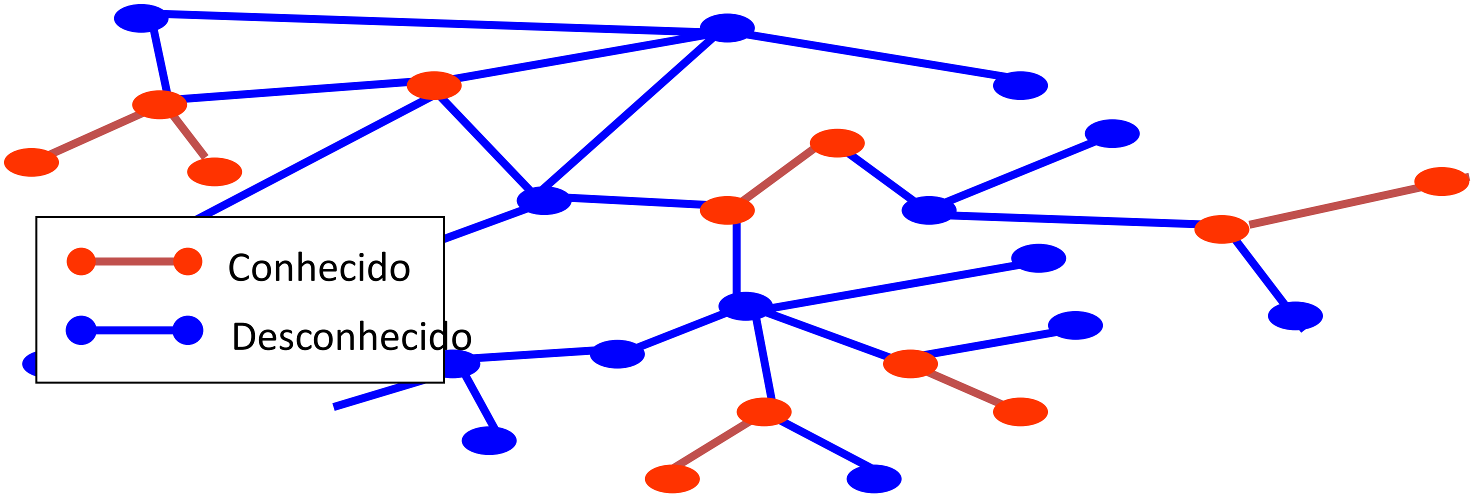




gura 5.4: *Balanço patrimonial de um ban*

Modelagem baseada em Agentes e Redes Complexas

- Nós da rede são pessoas (grupos), os links são relações (contatos)
- Quem está ligado intimamente a quem (comprimento do caminho, clustering)?
- Quem é fundamental na rede (centralidade)?
- Podemos inferir uma estrutura de rede a partir de uma quantidade muito pequena de dados (redes ocultas)?



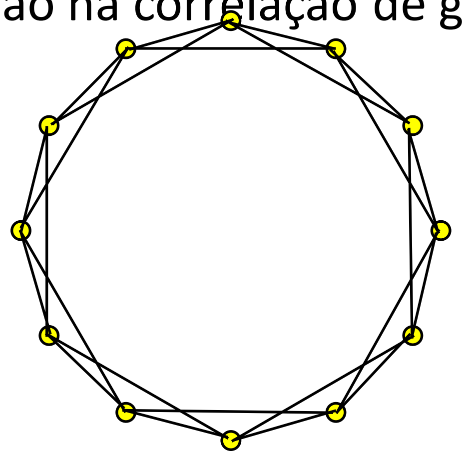
Afinal, o que são redes Complexas?

Afinal, o que são redes Complexas?

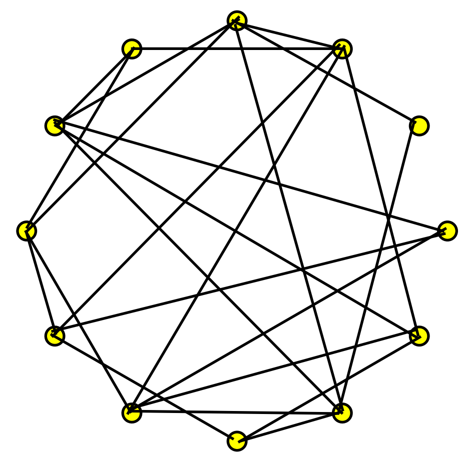
- redes regulares e redes aleatórias não são complexas

A primeira é definida pelo fato da distribuição de grau ser constante, ou seja, todos os vértices tem a mesma conectividade e apresenta uma geometria regular, com diâmetro grande. Apresenta alta clusterização.

A segunda se caracteriza por uma heterogenidade na distribuição de grau, mas ainda possui uma escala, pois a maioria dos vértices tem grau próximo a media. Além disso a distância é curta e o coeficiente de cluster é baixo. Basicamente não há correlação de grau entre os vértices.



??



ORDEM

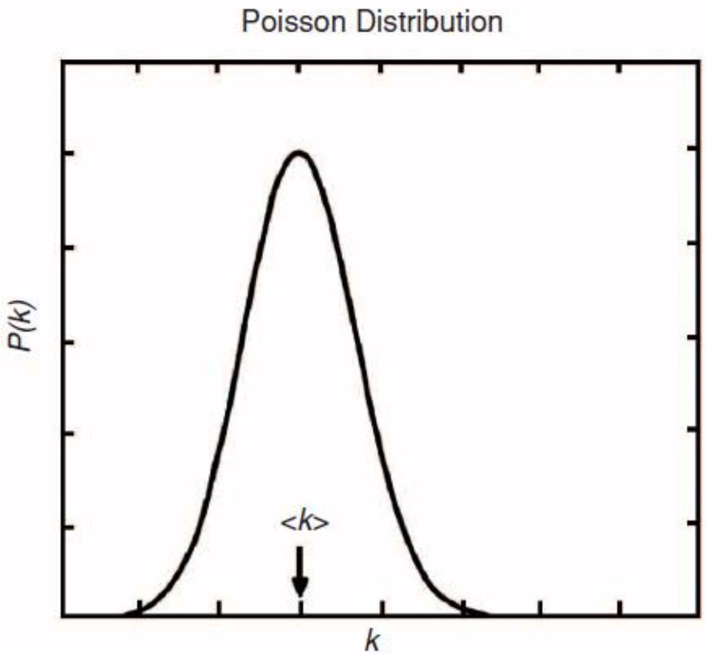
DeSorDeM

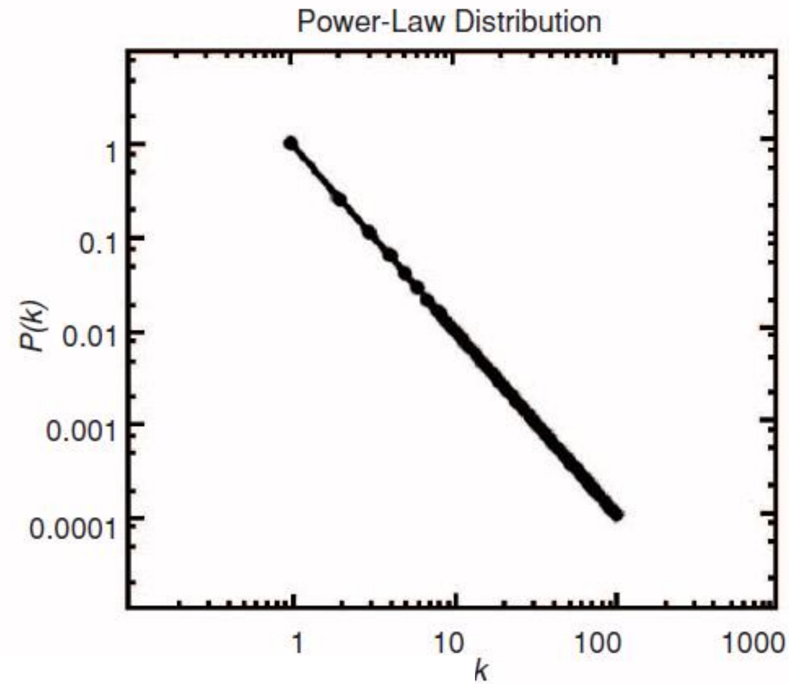
$\beta = 0$

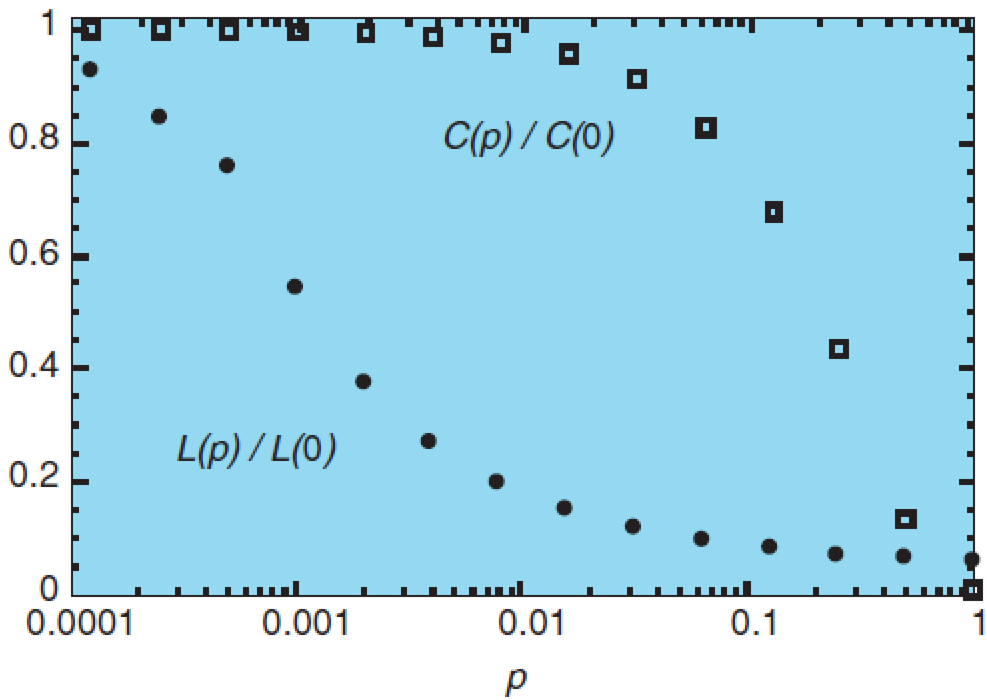
$\beta = 1$

Table 1.
Small-world pattern and scale-free property of several real networks. Each network has the number of nodes N, the clustering coefficient C, the average path length L and the degree exponent γ of the power-law degree distribution. The WWW and metabolic network are described by directed graphs.

Network	Size	Clustering coefficient	Average path length	Degree exponent
Internet, domain level [13]	32711	0.24	3.56	2.1
Internet, router level [13]	228298	0.03	9.51	2.1
WWW [14]	153127	0.11	3.1	$\gamma_{\text{in}} = 2.1 \quad \gamma_{\text{out}} = 2.45$
E-mail [15]	56969	0.03	4.95	1.81
Software [16]	1376	0.06	6.39	2.5
Electronic circuits [17]	329	0.34	3.17	2.5
Language [18]	460902	0.437	2.67	2.7
Movie actors [5, 7]	225226	0.79	3.65	2.3
Math. co-authorship [19]	70975	0.59	9.50	2.5
Food web [20, 21]	154	0.15	3.40	1.13
Metabolic system [22]	778	–	3.2	$\gamma_{\text{in}} = \gamma_{\text{out}} = 2.2$


Poisson Distribution
P(k)
<k>
k

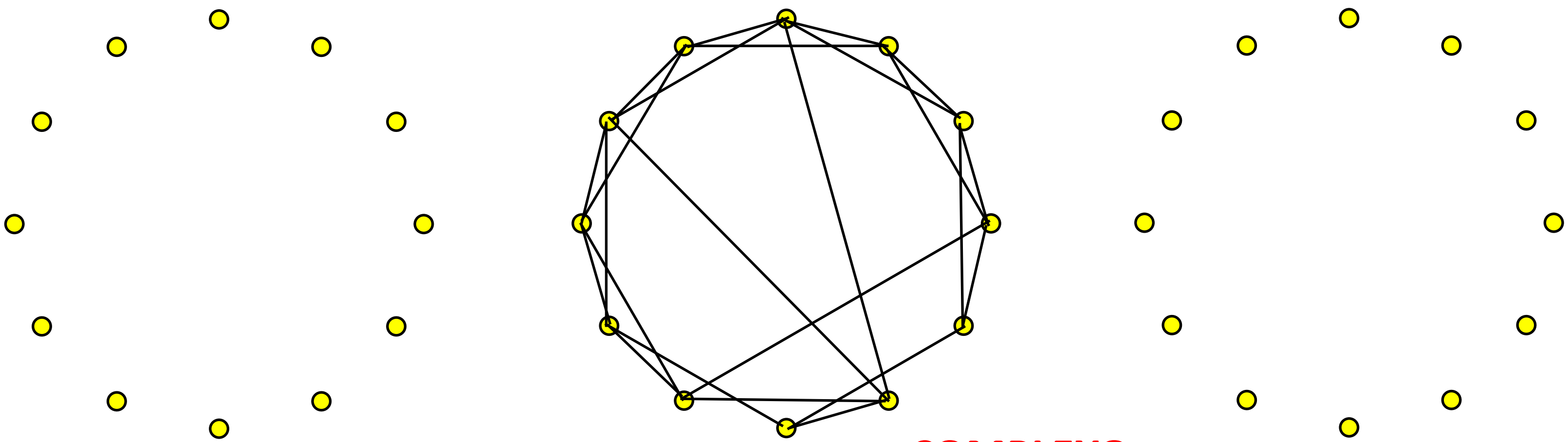

Power-Law Distribution
P(k)
1
0.1
0.01
0.001
0.0001
1
10
100
1000
k


C(p) / C(0)
L(p) / L(0)
1
0.8
0.6
0.4
0.2
0
0.0001
0.001
0.01
0.1
1
p

Afinal, o que são redes Complexas?

Redes Complexas: para identificar redes complexas há que se investigar algumas medidas

Distribuição de grau heterogênea (geralmente, mas não obrigatorimente Livre de Escala)
Clusterização significativa ou hieraquia ou estrutura de comunidades
Propriedade Pequeno Mundo
Presença de Assortatividade ou Desassortatividade



COMPLEXO

$\beta = 0.125$

Muitos elementos interagindo, entrelaçamento, interdependência e auto-organização e emergência

Materiais disponíveis na Internet em abundância!

http://www.icmc.usp.br/pessoas/francisco/resources.htm

http://www.network-science.org/

http://videolectures.net/Top/Computer_Science/Network_Analysis/