

# Redes Complexas

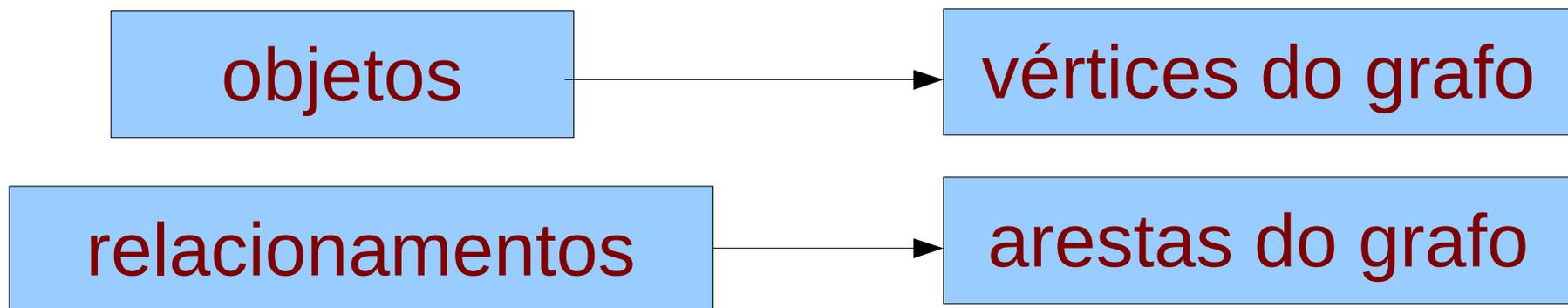
## Aula 2

### **Aula de hoje**

- Representando redes
- Falando sobre redes
- Grau, distância, clusterização
- Características de redes reais

# Rede

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



**Exemplos?**

# Objetos e Relacionamentos

## ■ Sobre objetos

- idênticos, diferentes tipos, atributos (rótulos)
- ex. nacionalidade, gênero, idade, etc

## ■ Sobre relacionamentos

- simétricos, assimétricos, diferentes intensidades (pesos), negativos
- múltiplos relacionamentos na mesma rede
- ex. amizade, colaboração, interação, confiança, ...

**Rede captura estrutura  
do relacionamento**

# Classe de Redes

- **Redes sociais:** relacionamento entre pessoas ou grupo de pessoas
- **Redes de informação** (de conhecimento): codificam associação entre informação
- **Redes tecnológicas:** construída pelo homem geralmente para distribuir *commodities*
- **Redes biológicas:** codifica relacionamentos em sistemas biológicos

## Classificação para referência

- *Twitter:* rede social ou rede de informação?

# Como Falar sobre Redes?



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

**Descrever estrutura da rede!**

## ■ **Matriz de adjacência, $A$**

### ■ Matriz $n \times n$ ( $n$ é número de vértices)

■  $A_{ij} = 1$  , se existe aresta entre vértices  $i$  e  $j$

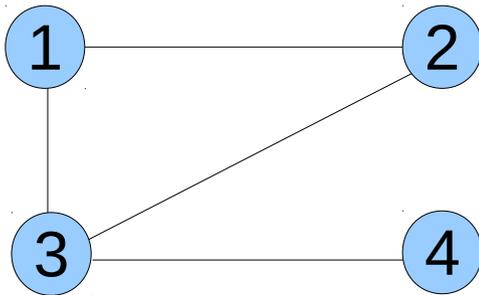
■  $A_{ij} = 0$  , caso contrário

■ 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

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

→ ?

# Matriz de Adjacência



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

**Precisamos de resumos da estrutura!**

- Resumos oferecem intuição sobre a estrutura

# Características de Redes

- Características estruturais são resumos da estrutura
  - ex: tamanho, densidade, graus, distâncias, clusterização, centralidade, homofilia, etc
- Quais características devem ser avaliadas? Quais são importantes?
- Depende do propósito!
- Como gens que formam o DNA, diferentes resumos possuem implicações diferentes



# Característica Importante



- O que faz uma característica ser importante?

- 1) Prever (determinar) comportamento geral de algum processo

- independente de outras características

- 2) Influência sobre diferentes processos

- Exemplo: distribuição de grau

- determina e influencia comportamento de passeios aleatórios e epidemias

- Não conhecemos muitas características fundamentalmente importantes

# 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  $\longrightarrow \binom{n}{2} = \frac{n(n-1)}{2} \leq n^2$

- Densidade: fração de arestas que o grafo possui

$$\rho = \frac{m}{\binom{n}{2}} = \frac{2m}{n(n-1)}$$

# Grau

- Número de arestas (relacionamentos) incidente em um vértice
  - $g(u)$  : grau do vértice  $u$



- Como falar sobre o grau dos vértices da rede?
  - como falar de uma característica sobre um conjunto de objetos?
- Média, desvio padrão: um número
- Distribuição empírica: todos os graus – comportamento geral

# Grau Médio

- Grau médio do grafo considerando todos seus vértices

$$\bar{g} = 1/n \sum_{u \in V} g(u) \quad \longleftarrow \text{ Grau do vértice } u$$

- Também pode ser calculado diretamente

$$\sum_{u \in V} g(u) = 2m$$

- Cada aresta tem duas pontas!

- Logo temos  $\bar{g} = 2m/n$

# Distribuição do Grau

- Distribuição empírica do grau dos vértices
  - frequência relativa do grau

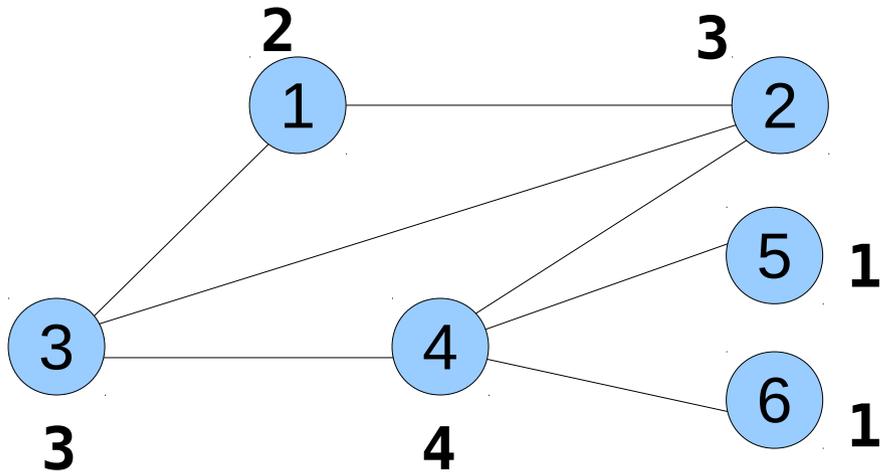
$$f_k = \frac{\text{Número de vértices com grau } k}{\text{Número total de vértices}}$$

fração de vértices com grau  $k$

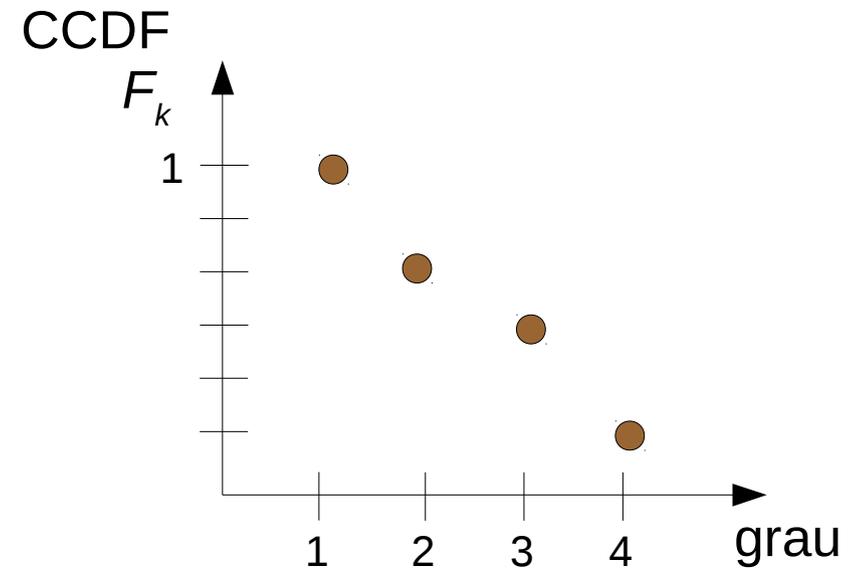
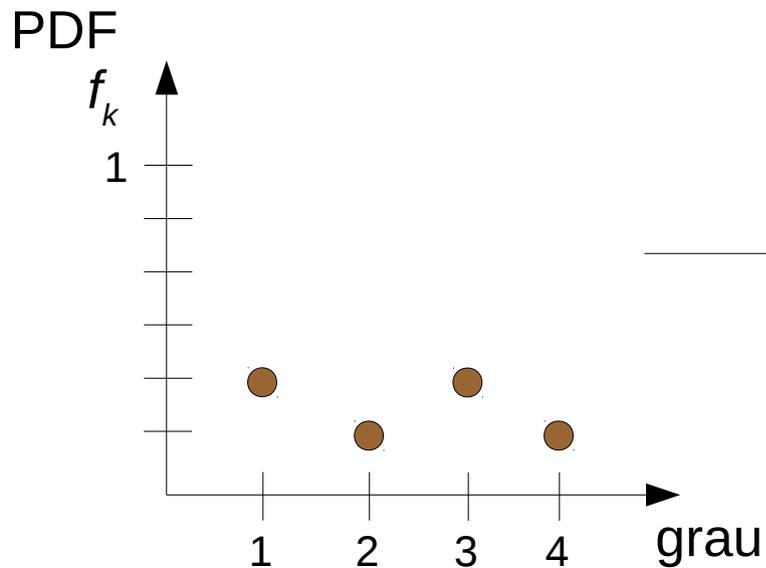
- CCDF empírica (*Complementary Cumulative Distribution Function*)
  - fração de vértices com grau  $\geq k$

$$F_k = 1 - \sum_{i=0}^{k-1} f_i$$

# Exemplo de Distribuição

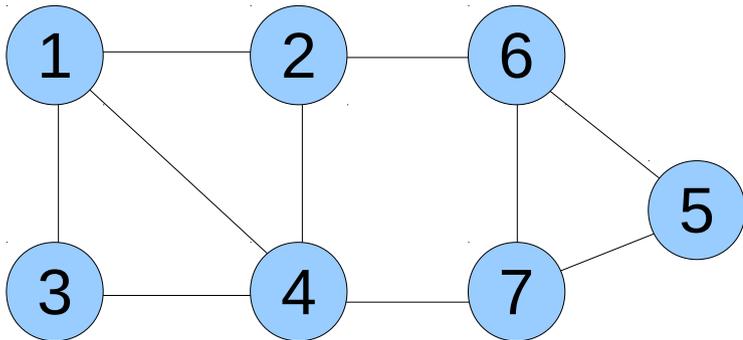


□ Distribuição empírica de grau?



# Distância

- Comprimento do **menor** caminho entre dois vértices
- Função  $d(u,v)$ , onde  $u$  e  $v$  são vértices
  - não definido quando não há caminho



- Exemplo

- $d(6, 3) = ?$

- $d(7, 1) = ?$

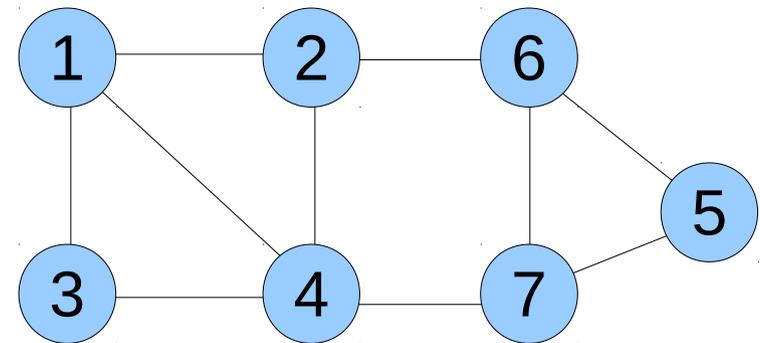
- **Importante:** muitos caminhos podem realizar a distância entre dois vértices

# Distância Média e Diâmetro

- Distância média do grafo

- média entre todos os pares de vértices

$$\bar{d} = \frac{\sum_{u, v \in V} d(u, v)}{\binom{n}{2}}$$



- **Excentricidade:** maior distância de um vértice a todos os outros

$$e(u) = \max_{v \in V} d(u, v)$$

- **Diâmetro:** maior distância entre dois vértices da rede

$$r = \max_{u, v \in V} d(u, v)$$

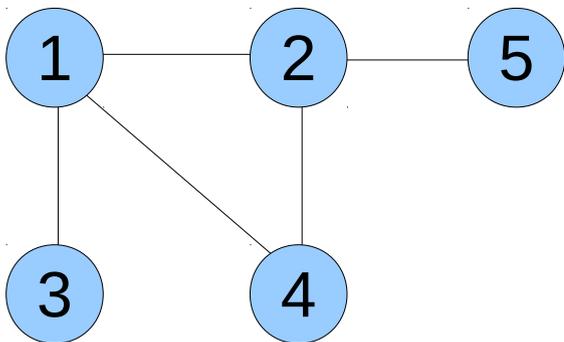
# Distribuição da Distância

- Distribuição empírica da distância entre os vértices do grafo
  - frequência relativa das distâncias

$$f_D(d) = \frac{n(d)}{\binom{n}{2}}$$

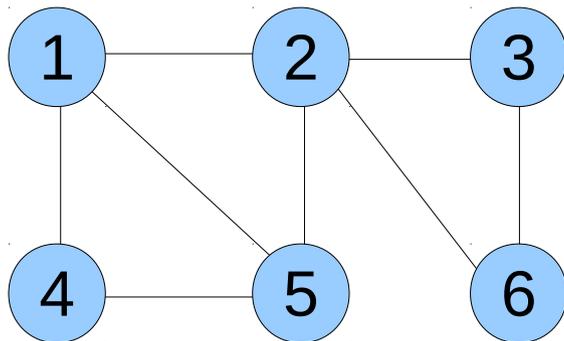
← Número de pares não ordenados de vértices que tem distância  $d$

## ■ Exemplo



# Clusterização

- Medida da propensão de triângulos se formarem na rede
  - triângulo: ciclo de comprimento três
  - quase-triângulo: caminho de comprimento dois



- Duas métricas na literatura
  - métrica local: cada vértice tem o seu
  - métrica global: único valor para a rede

# Clusterização – I

- Fração de aresta entre vizinhos
  - prob. de dois vizinhos também serem vizinhos
- Definida para cada vértice da rede

$$C_i = \frac{E_i}{\binom{d_i}{2}}$$

$E_i$  : # de arestas entre os vizinhos de  $i$

$d_i$  : grau do vértice  $i$

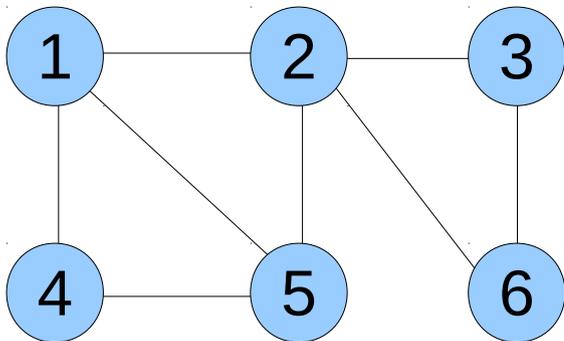
$$C_i = \frac{2 E_i}{d_i(d_i - 1)}$$

- Clusterização do grafo
  - média da clusterização dos vértices

$$C = 1/n \sum_{v \in V} C_i$$

# Calculando Clusterização

## ■ Exemplo



## ■ Quanto vale C?

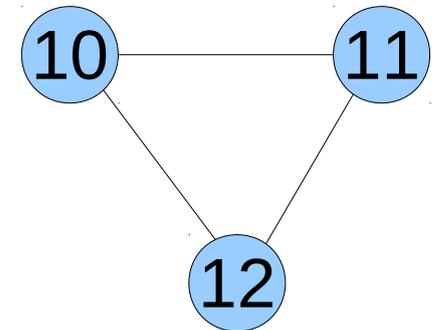
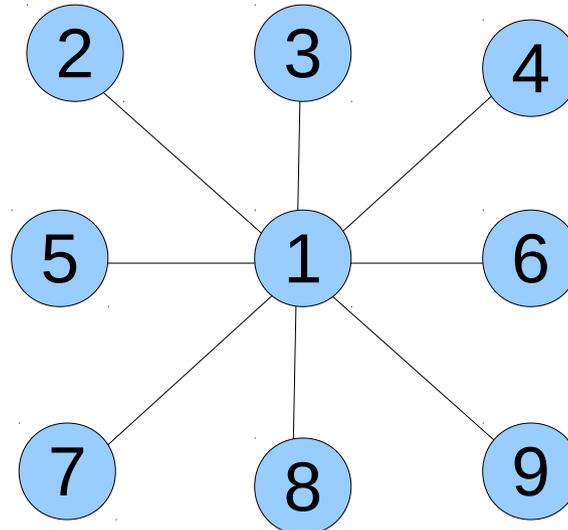
$$C_1 = \frac{E_1}{\binom{d_1}{2}} = \frac{2}{\binom{3}{2}} = \frac{2}{3}$$

$$C_2 = \frac{E_2}{\binom{d_2}{2}} = \frac{2}{\binom{4}{2}} = \frac{1}{3}$$

# Clusterização – I

- Métrica não é a fração de arestas entre vizinhos do grafo (como um todo)
  - média dos vértices
- Favorece vértices com menos vizinhos
  - denominador é  $\sim d_i^2$
- Valor da clusterização para grau 0 e 1?
- Exemplo

$$C_2 = ?$$



# Clusterização – II

- Fração entre o número de triângulos e o número de pseudo-triângulos no grafo como um todo

- pseudo-triângulo: tripla conectada

$$C' = \frac{3 \times \text{número de triângulos}}{\text{número de triplas conectadas}}$$

Cada triângulo dá origem a três triplas conectadas

- Tripla é não-ordenada (“1,2,3” igual a “3,2,1”)
- Métrica global
  - não é média dos vértices
- Não tem problemas com vértices de grau 0 e 1
  - mais bem comportada a anomalias
- Valor pode ser bem diferente da outra definição

# Quatro Importantes Características

- ❑ Observada em diversas redes reais
  - a partir do final da década de 90



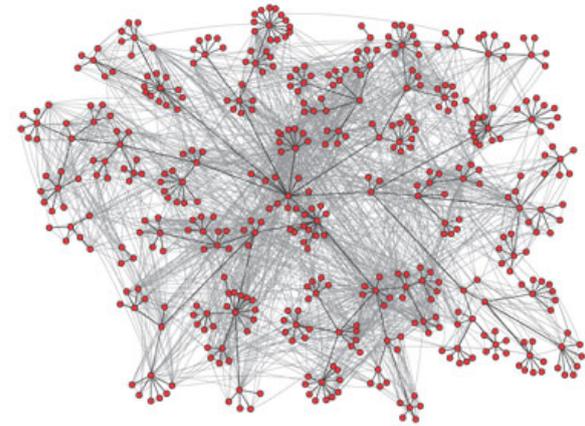
■ Mundo pequeno



■ Meus amigos são amigos



■ Normalidade ausente



■ Tudo Conectado

# 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
  - Rede de colaboração –  $10^6$  nodes
  - Facebook -  $10^9$  nodes
    - “seis graus de separação”
  - e muitas outras!

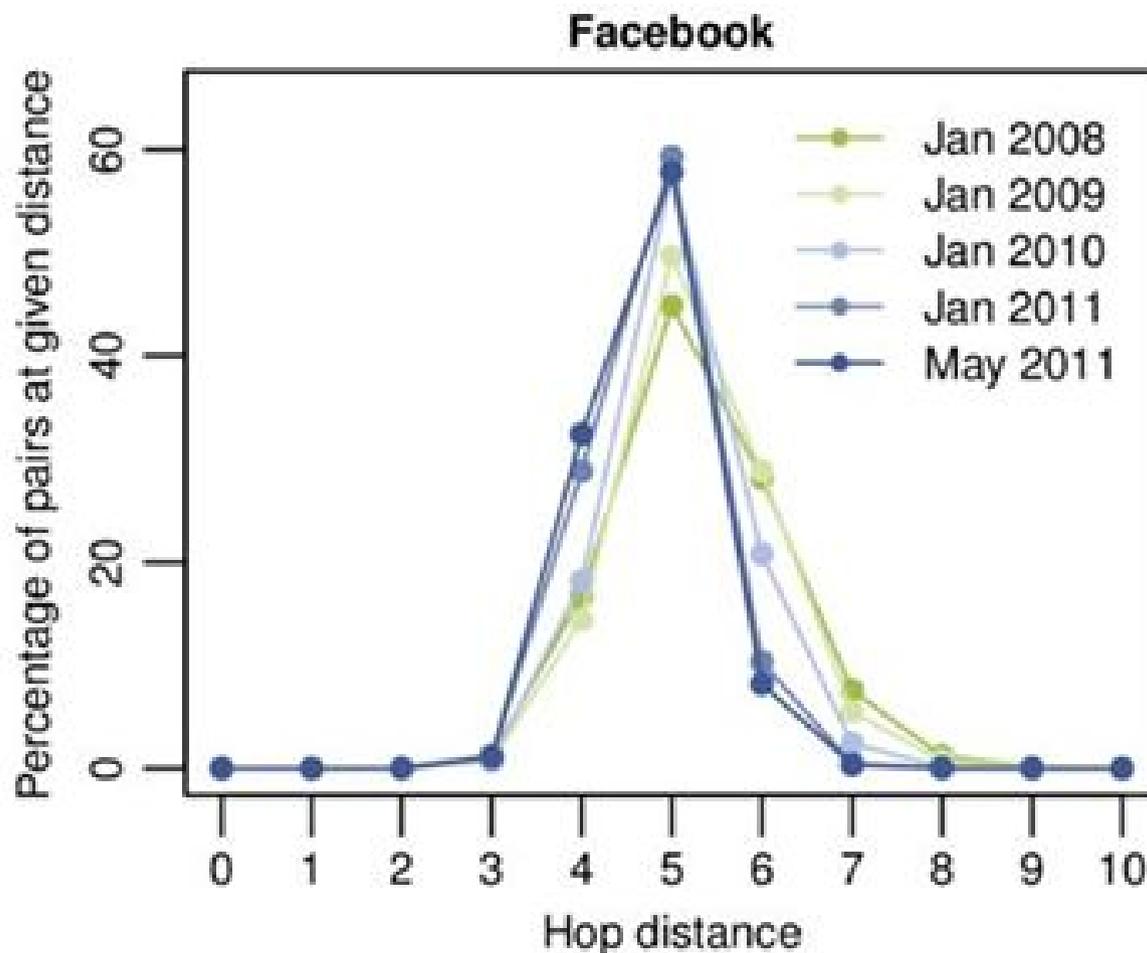


←	7.5
←	4.9
←	4.5

**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

# Exemplo do Facebook

## ■ Mudanças na rede

distância média (std)

	fb
2007	4.46 ( $\pm 0.04$ )
2008	5.28 ( $\pm 0.03$ )
2009	5.26 ( $\pm 0.03$ )
2010	5.06 ( $\pm 0.01$ )
2011	4.81 ( $\pm 0.04$ )
current	4.74 ( $\pm 0.02$ )

grau médio

	fb
2007	99.50
2008	76.15
2009	88.68
2010	113.00
2011	169.03
current	190.44

densidade

	fb
2007	7.679E-06
2008	1.359E-06
2009	6.377E-07
2010	3.400E-07
2011	3.006E-07
current	2.641E-07

- Grau médio cresce
- Rede muito, muito esparsa
- Grau médio aumenta e densidade diminui?

# 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!**

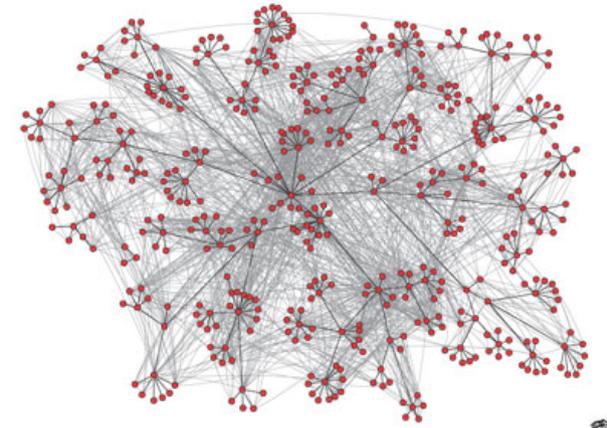
- |                           |                           |
|---------------------------|---------------------------|
| ■ AS Graph – $10^4$ nodes | ■ Citações – $10^6$ nodes |
| ■ Grau médio: <b>5.9</b>  | ■ Grau médio: <b>8.6</b>  |
| ■ Grau máx: <b>~2100</b>  | ■ Grau máx: <b>~9000</b>  |

**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**
- Também em redes sociais, web, etc



**Quase sempre quase completamente conectada!**

# Curiosidade

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



- “Million dollar question”

**Por que?**



- Algumas respostas na literatura, nada muito definitivo
- E você pode ajudar a explicar!