

# Redes Complexas

## Aula 3

### Aula passada

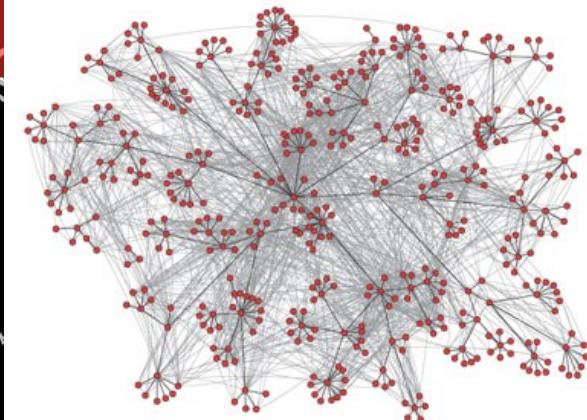
- Caracterizando redes grandes
- Grau
- Distâncias
- Clusterização

### Aula de hoje

- Características de redes reais
- Centralidade de vértices
- *Betweeness, Closeness*

# 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

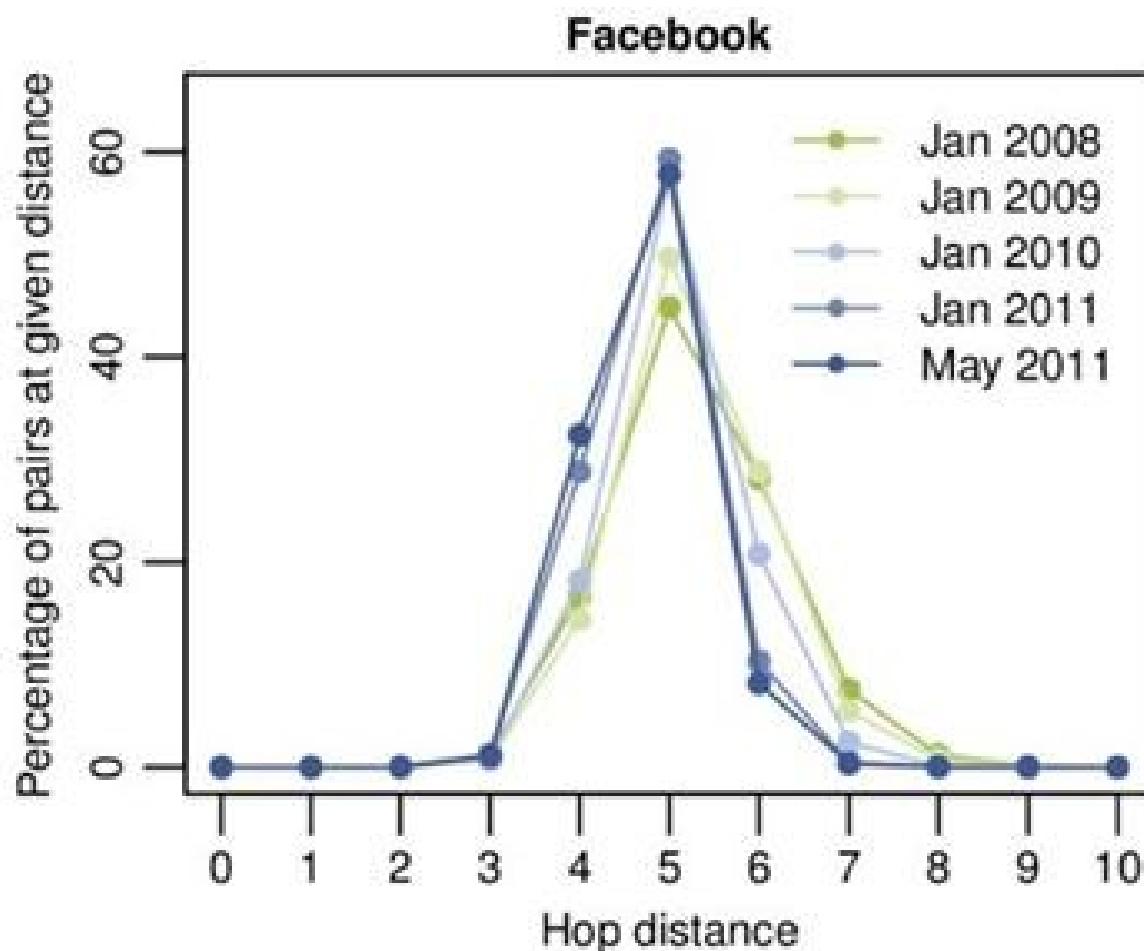
- “It's a small world, after all”
- Distância média **muito** pequena, diâmetro também
- Mesmo para redes muito grandes
  - Web (parte) -  $10^8$  vértices
  - Rede de colaboração -  $10^6$  nodes
  - Facebook -  $10^9$  nodes
    - “seis graus de separação”
  - e muitas outras!



**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)		grau médio		densidade	
	fb		fb		fb
2007	4.46 ( $\pm 0.04$ )	2007	99.50	2007	7.679E-06
2008	5.28 ( $\pm 0.03$ )	2008	76.15	2008	1.359E-06
2009	5.26 ( $\pm 0.03$ )	2009	88.68	2009	6.377E-07
2010	5.06 ( $\pm 0.01$ )	2010	113.00	2010	3.400E-07
2011	4.81 ( $\pm 0.04$ )	2011	169.03	2011	3.006E-07
current	4.74 ( $\pm 0.02$ )	current	190.44	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!**

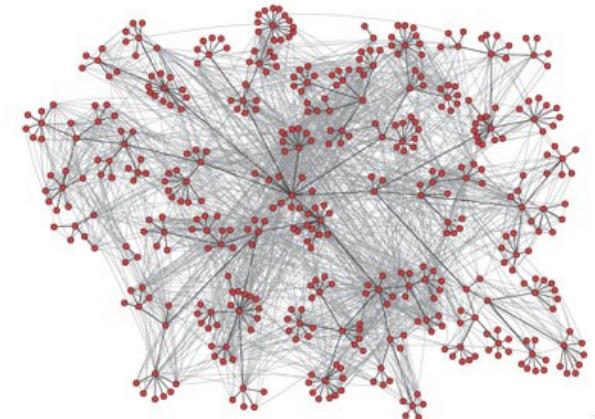
- |                           |                           |
|---------------------------|---------------------------|
| ■ AS Graph - $10^4$ nodes | ■ Citações - $10^6$ nodes |
| ■ Grau médio: <b>5.9</b>  | ■ Grau médio: <b>8.6</b>  |
| ■ Grau máx: <b>~2100</b>  | ■ Grau máx: <b>~9000</b>  |

**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**
- Rede social, rede neural, etc



**Quase sempre quase completamente conectada!**

# Curiosidade

- **Fato I:** Muitas redes possuem propriedades estruturais peculiares (não esperadas)
  - grau, distâncias, clusterização, connectividade, 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, mas nada muito definitivo
- E você pode ajudar a explicar!

# Centralidade



## Como medir a *importância* de um vértice?

- Utilizando apenas a estrutura
- Relativo a outros vértices
  - Métricas locais
    - dependem apenas da vizinhança do vértice  
(Ex. grau, random walk)
  - Métrica globais
    - Dependem do grafo inteiro  
(Ex. Closeness, pagerank)
- Grau
- Betweeness
- Closeness
- Autovetor
- Random walks
- etc.

# 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

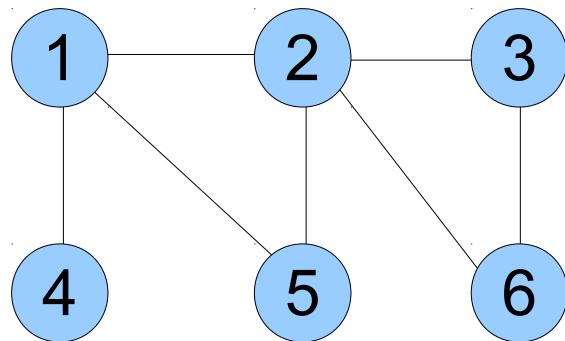
- Grau do vértice ou grau do vértice normalizado
  - valor entre 0 e 1

$$C_v = \frac{d_v}{n-1}$$

- Grafo direcionado, grau de entrada/saída
  - Duas centralidades em grau por vértice

# Centralidade de Betweenness

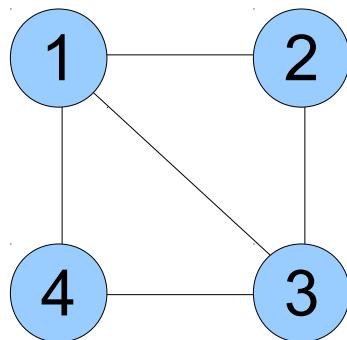
- Mede o quanto no “meio do caminho” um vértice está
- Considerar todos os caminhos mínimos do grafo
- Número de caminhos mínimos que passam pelo vértice
- Exemplo



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

# Centralidade de Betweenness

- **Problema:** Como definir métrica quando mais de um caminho mínimo existe entre um par origem/destino?
  - empate no custo do caminho mínimo
- Exemplo



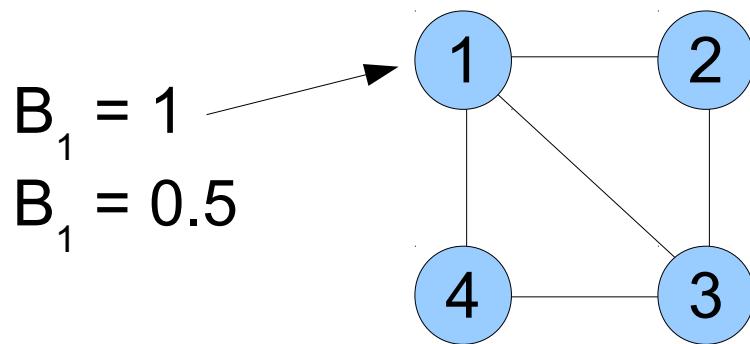
- Caminho mínimo entre 2 e 4?
- 2,1,4 ou 2,3,4?
- Centralidade do vértice 1 e 3?

# Centralidade de Betweenness

- Duas abordagens

- Cada caminho mínimo conta 1 vez
  - “Carga” dividida pelos caminhos mínimos (cada caminho mínimo conta  $1/k$  para a métrica, para  $k$  caminhos)

- Exemplo



- Para muitas redes, diferença é pequena
  - Mas nem sempre!

# 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

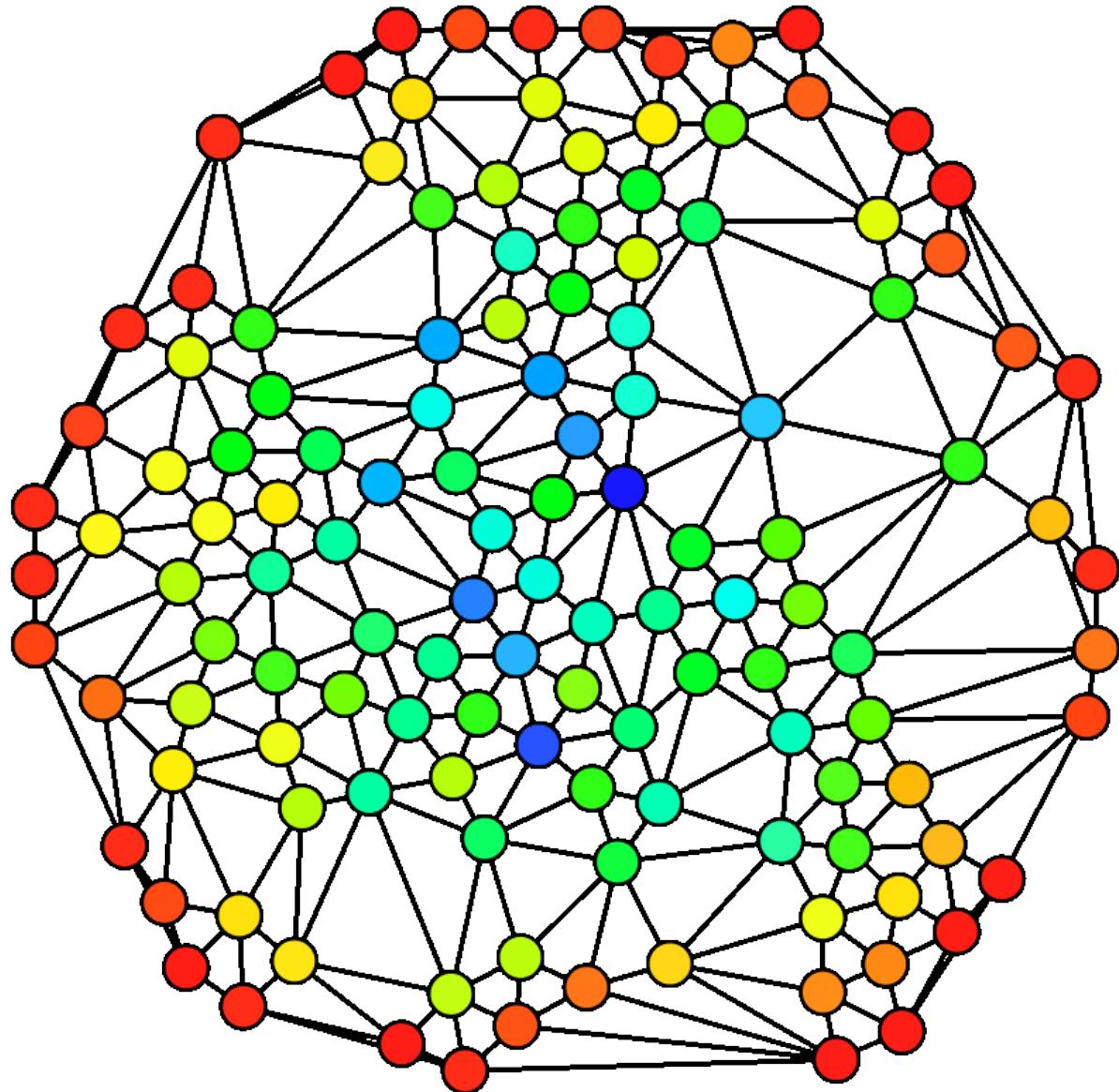
- ou

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

Número de caminhos  
mais curtos entre s e t

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

# Exemplo



- Cores indicam betweeness
- Vermelho = 0, azul = máximo
- Ilustra vértices mais centrais

# Centralidade de Closeness

- Utiliza conceito de distância
  - com ou sem pesos
- Distância média entre vértice e o resto do grafo
  - capturar o quanto central é o vértice

$$C_v = \frac{\sum_{t \in V - \{v\}} d(v, t)}{n - 1}$$

- “Velocidade” com a qual informação se propaga de um vértice para o resto da rede

# Centralidade de Closeness

- Maior distância a qualquer outro vértice do grafo
  - eccentricidade (*eccentricity*) do vértice

$$C_v = \max_{t \in V - \{v\}} d(v, t)$$