Redes Complexas
Aula 3

Roteiro

- Centralidade de vértices
- *Betweeness, Closeness*
- Centralidade de Autovetor, Katz, PageRank

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 da rede inteira (ex. closeness, Pagerank)

Centralidade

Qual é a melhor métrica para capturar importância?

- Como determinar a qualidade de um ranqueamento?

Impossível sem referência externa!

- Permite avaliar um ranqueamento
  - referência empírica ou processual
- Referência depende do contexto e do objetivo do ranqueamento

Não existe a melhor métrica!

Centralidade de Grau

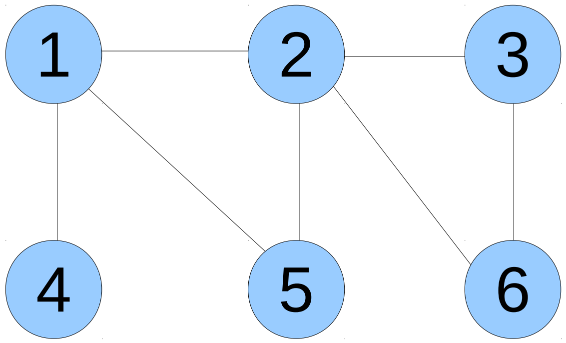
- Grau do vértice ou grau do vértice normalizado
  - valor entre 0 e 1

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

- Rede direcionada, grau de entrada/saída
  - duas centralidades em grau por vértice

Centralidade de Betweeness

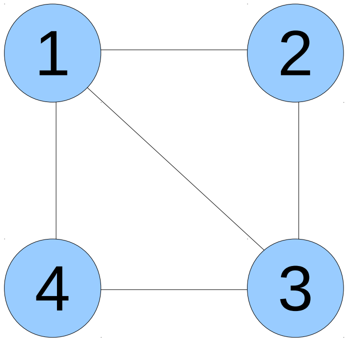
- Mede o quanto no "meio do caminho" um vértice está
- Considerar todos os caminhos mínimos da rede
- Número de caminhos mínimos que passam pelo vértice
- Exemplo



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

Centralidade de Betweeness

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



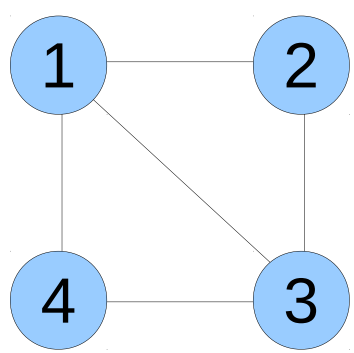
- Caminho mínimo entre 2 e 4
  - 2,1,4 ou 2,3,4?
- Centralidade do vértice 1 e 3?

Centralidade de Betweeness

- 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

$B_1 = 1$

$B_1 = 0.5$



- 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 mínimos entre s e t que passam por v

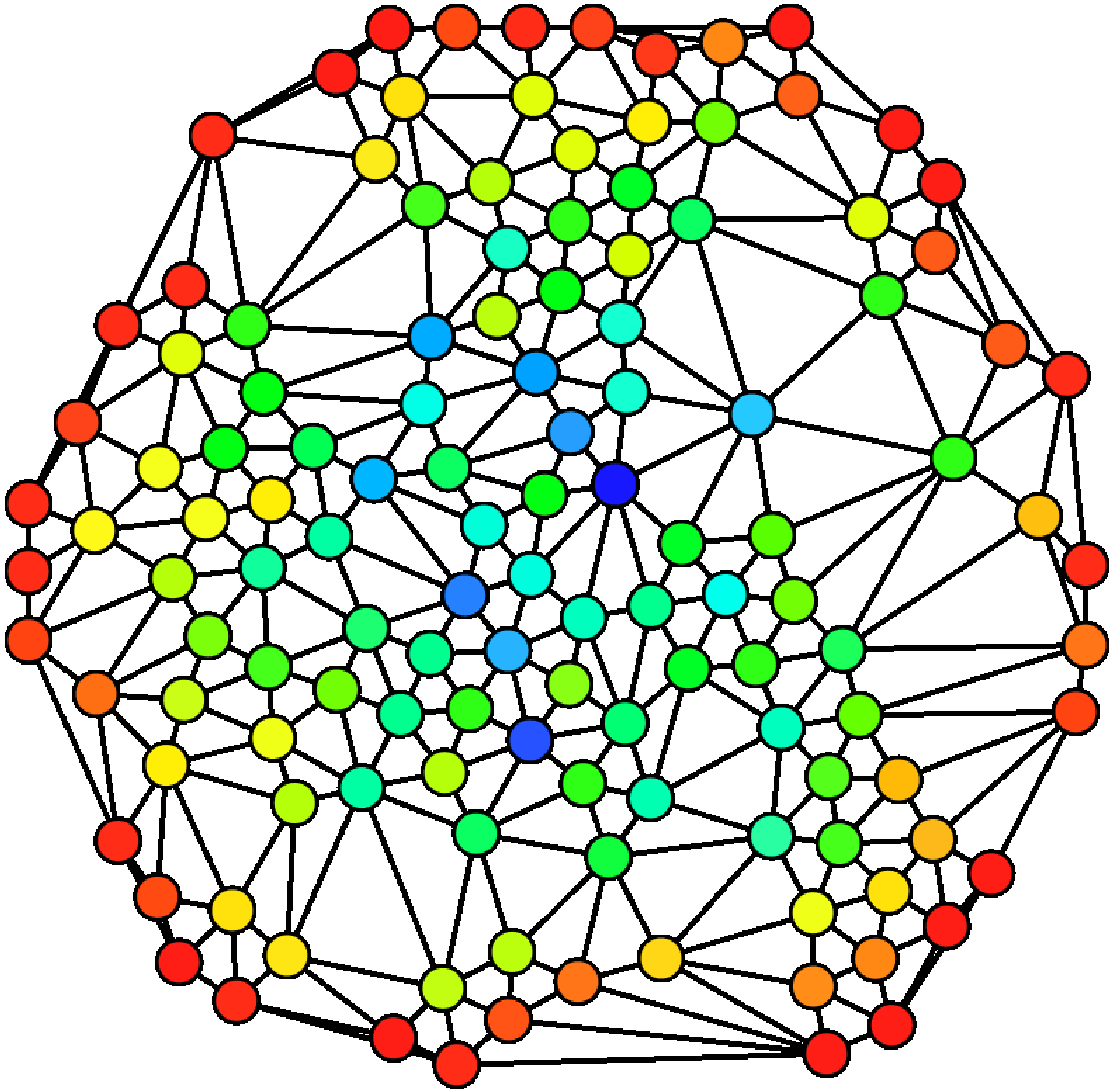
Número de caminhos mínimos entre s e t

- ou

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

- C_v 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 quão central é o vértice

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

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

Centralidade de Excentricidade

- Maior distância para algum vértice do grafo
  - excentricidade (*eccentricity*) do vértice

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

Centralidade Recursiva

- **Ideia**: importância do vértice depende da importância dos vizinhos
  - recursão para definir centralidade

Como formalizar ideia?

- Seja x_i a importância do vértice *i*

$$x_i = \sum_{j=1}^{n} a_{ij} x_j$$

Somatório da importância dos vizinhos de *i*

a_{ij}: matriz de adjacência

Como calcular x_i?

Centralidade de Autovetor

- Processo iterativo
- Iniciar com vetor x(0) = (x_1(0), ..., x_n(0))
- Forma matricial: $x(1) = A\,x(0)$
- Depois de *t* iterações

$$x(t) = A^t\,x(0)$$

Processo converge!
(normalizar *x* depois de cada passo)

$$A\,x = \kappa_1\,x$$

x é o autovetor associado ao autovalor k_1 (maior autovalor de A)

Redes Direcionadas

Como definir importância?

- **Ideia**: importância do vértice depende da importancia dos vértices que apontam para ele

$$x_i = \sum_{j=1}^{n} a_{ji} x_j \longleftarrow a_{ij} \mathrel{!}= a_{ji}$$

Problema?

- Vértices com grau de entrada zero?
- Mais de uma componente conexa?

Centralidade de Katz

Como resolver o problema?

- **Idéia**: Todo vértice tem pequena importância intrínsica

$$x_i = \alpha \sum_{j=1}^{n} a_{ij} x_j + \beta$$

α e β não dependem da estrutura da rede

- Relação entre α e β determina relação entre estrutura e aptidão externa
  - em geral, $\beta = 1$
- Centralidade de Katz, definida em 1953 por Leo Katz para medir influência em redes sociais (reais)

Centralidade de Katz

Qual valor para alpha?

$$x_i = \alpha \sum_{j=1}^{n} a_{ij} x_j + 1$$

- alpha muito pequeno, centralidade será determinada por por beta
- alpha muito grande, processo iterativo diverge
- Restrição para convergência:

$$\alpha < 1/\kappa_1$$ maior autovalor de A

Centralidade de PageRank

Problema com Katz?

- Vértice importantes espalham importância igualmente, independente do grau de saída
- Outra ideia: espalhar importância proporcionalmente ao grau

$$x_i = \alpha \sum_{j=1}^{n} a_{ji} \frac{x_j}{d_j^s} + \beta$$

Grau de saída de j

- Centralidade de PageRank, com $\beta=(1-\alpha)/N$
- Proposto e utilizada pelo Google

PageRank

- Interpretação original: surfista passeando pela Web de forma aleatória (*random surfer model*)
  - a cada página, escolhe hiperlink de maneira uniforme
- Importância: fração de visitas a cada página

Problema: Web não é conexa!

- Surfista dá saltos para página qualquer da web
  - a cada página, decide dar salto aleatório

Modelo é uma cadeia de Markov!

- distribuição estacionária é dada pelas equações do slide anterior!

Personalized PageRank

- PageRank: ranqueamento absoluto dos vértices da rede
- Como obter ranqueamento relativo a um vértice?
  - importância do ponto de vista do vértice

Personalized PageRank (PPR)

- **Ideia:** salto retorna para um vértice específico
  - difusão a partir deste vértice (i*)

$$x_i = \alpha \sum_{j=1}^{n} a_{ji} \frac{x_j}{d_j^s} + \beta$$

$\beta=(1-\alpha)$, se i = i*
$\beta=0$, se i != i*

Ranqueamento relativo a i*