Redes Complexas
Aula 4

Aula passada

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

Aula de hoje

- Centralidade de Autovetor, Katz, PageRank
- Passeios aleatórios (RW)

Centralidade

Como medir a *importância* de um vértice?

- Utilizando apenas a estrutura
- Relativo a outros vértices da rede

- Atribuir a cada vértice um número (função da estrutura)
- Ordenar baseado neste número

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

Centralidade

- **Ideia**: importância do vértice depende da importância dos vizinhos
  - Recursão *to the rescue!*

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 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)

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

autovetor associado ao maior autovalor k_1

Redes Direcionadas

Como definir importância?

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

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

a_{ij} != a_{ji}

Problema?

- Vértices com grau de entrada zero?
- Zero em todos os vértices fora da CC

Centralidade de Katz

Como resolver o problema?

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

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

alpha e beta não dependem da estrutura

- Relação entre alpha e beta determina relação entre estrutura e aptidão externa
  - em geral beta = 1
- Centralidade de Katz

Centralidade de Katz

Qual valor para alpha?

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

- Alpha muito pequeno, centralidade dado por beta (1)
- 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
- Melhor intuição: espalhar importância proporcionalmente

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

Grau de saída de j

- Centralidade de PageRank
- 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!

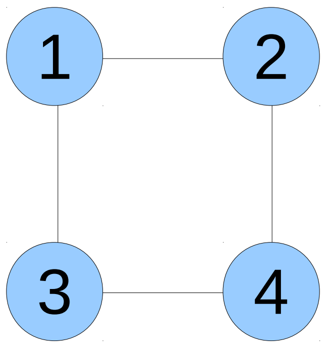
- Surfista dá saltos para qualquer página da web
  - a cada página, decide dar saltos aleatórios

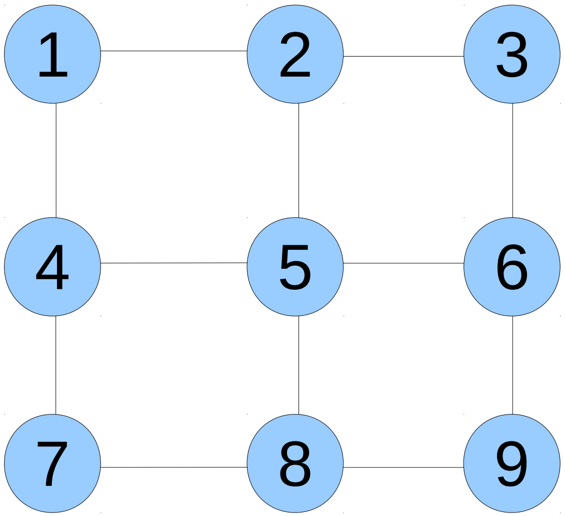
Modelo é uma cadeia de Markov!

- Solução é dada pelas equações do slide anterior!

Random Walks

- Passeios aleatórios: forma aleatória de percorrer um grafo (rede) – BFS é determinística
  - muitas aplicações (busca, centralidade, etc)
- A cada passo, escolher vértice vizinho com igual probabilidade e andar para este vértice





- $s(t)$: vértice no instante t
- $v_i(t)$: número de visitas ao vértice i no intervalo $[0,t]$
- $f_i(t) = v_i(t)/t$: fração de visitas ao vértice i
- Quanto vale $f_i(t)$?

Random Walks

- Probabilidade da walker estar no vértice *i* depois de muito tempo?
  - Prob é igual a fração de visitas!

Intuição?

- Proporcional ao grau de i

$$\lim_{t\to\infty} P[s(t)=i]=\frac{d(i)}{\sum_j d(j)}$$

s(t) é o estado da RW
no instante t

Centralidade de RW

- Considerar todos os caminhos entre dois vértices
  - E não apenas o mais curto (generaliza *betweeness*)

Como usar RW ?

$$n_i(s,t)$$

- Número de visitas ao vértice i realizadas pela uma RW saindo de s e chegando a t
  - variável aleatória (usar média)
- Centralidade é soma para todo s, t
- $n_i(s,t) \neq n_i(t,s)$