Introdução às Redes de Interação – MO804 (MC908)
Centralidade de vértices e arestas

Prof. Dr. Ruben Interian

Instituto de Computação, UNICAMP

Resumo

Resumo

Revisão do conteúdo

O que é um vértice (ou uma aresta) **importante em uma rede**?

Revisão do conteúdo

O que é um vértice (ou uma aresta) **importante em uma rede**?

- **Centralidade de grau**: aquele vértice com o maior grau, $f(v_i) = d(v_i) = k_i$.

Revisão do conteúdo

O que é um vértice (ou uma aresta) **importante em uma rede**?

- **Centralidade de grau**: aquele vértice com o maior grau, $f(v_i) = d(v_i) = k_i$.
- **Centralidade de proximidade**, *closeness centrality*: o mais próximo dos outros vértices da rede, $f(v_i) = C(v_i) = \frac{n-1}{\sum_{j \neq i} d(v_i, v_j)}$.

Revisão do conteúdo

O que é um vértice (ou uma aresta) **importante em uma rede**?

- **Centralidade de grau**: aquele vértice com o maior grau, $f(v_i) = d(v_i) = k_i$.
- **Centralidade de proximidade**, *closeness centrality*: o mais próximo dos outros vértices da rede, $f(v_i) = C(v_i) = \frac{n-1}{\sum_{j \neq i} d(v_i, v_j)}$.
- **Centralidade de intermediação**, *betweenness centrality*: se muitos caminhos mais curtos passam pelo vértice/aresta, $f(v_i) = B(v_i) = \sum_{\substack{s \neq v_i \\ t \neq v_i}} \frac{\sigma(s, v_i, t)}{\sigma(s, t)}$.

Revisão do conteúdo

O que é um vértice (ou uma aresta) **importante em uma rede**?

- **Centralidade de grau**: aquele vértice com o maior grau, $f(v_i) = d(v_i) = k_i$.
- **Centralidade de proximidade**, *closeness centrality*: o mais próximo dos outros vértices da rede, $f(v_i) = C(v_i) = \frac{n-1}{\sum_{j \neq i} d(v_i, v_j)}$.
- **Centralidade de intermediação**, *betweenness centrality*: se muitos caminhos mais curtos passam pelo vértice/aresta, $f(v_i) = B(v_i) = \sum_{\substack{s \neq v_i \\ t \neq v_i}} \frac{\sigma(s, v_i, t)}{\sigma(s, t)}$.
- **Outros indicadores de centralidade**: pontes (+locais), pontos de articulação.

Resumo

Resumo

Centralidade de autovetor

Centralidade de autovetor, *eigenvector centrality*:

- **Ideia**: um vértice v_i tem **centralidade alta** quando ele está conectado a **outros vértices com centralidade alta**; v_i é influente quando ele está conectado a outros nós influentes.

Centralidade de autovetor

Centralidade de autovetor, *eigenvector centrality*:

- **Ideia**: um vértice v_i tem **centralidade alta** quando ele está conectado a **outros vértices com centralidade alta**; v_i é influente quando ele está conectado a outros nós influentes.
  - Mas professor, essa definição não faz sentido! Parece uma definição recursiva!

Centralidade de autovetor

Centralidade de autovetor, *eigenvector centrality*:

- **Ideia**: um vértice v_i tem **centralidade alta** quando ele está conectado a **outros vértices com centralidade alta**; v_i é influente quando ele está conectado a outros nós influentes.
  - Mas professor, essa definição não faz sentido! Parece uma definição recursiva!
- **Sim**, a definição é recursiva. Embora pareça esquisita, esta centralidade está **bem definida**.

Centralidade de autovetor

Formalizando.

- **Notação**: o valor da centralidade do vértice v_i será denotado $f(v_i) = x_i$, como uma variável cujo valor ainda não conhecemos.

Centralidade de autovetor

Formalizando.

- **Notação**: o valor da centralidade do vértice v_i será denotado $f(v_i) = x_i$, como uma variável cujo valor ainda não conhecemos.
- O valor de x_i deve ser proporcional à media das centralidades dos seus adjacentes:

$$x_i = \frac{1}{\lambda} \sum_{v_j \in N(v_i)} x_j = \frac{1}{\lambda} \sum_{j=1}^{n} a_{ij} x_j,$$

onde $a_{ij} = 1$ se há aresta, e $\frac{1}{\lambda}$ é uma constante.

Centralidade de autovetor

Podemos **reescrever essa equação** assim:

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

$$\lambda x_i = \sum_{j=1}^{n} a_{ij} x_j, \text{ para cada } x_i.$$

Centralidade de autovetor

Podemos **reescrever essa equação** assim:

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

$$\lambda x_i = \sum_{j=1}^{n} a_{ij} x_j, \text{ para cada } x_i.$$

Se definimos um vetor de valores da centralidade $x = (x_1, x_2, \ldots, x_n)$, e consideramos a matriz de adjacência $A_{n \times n} = (a_{ij})$, podemos escrever as equações em **forma matricial**:

$$\lambda x = Ax.$$

Centralidade de autovetor

$$\lambda x = Ax, \text{ equivalente a } (A - \lambda I)x = 0.$$

Precisamos que x seja **não nulo** (de preferência, não negativo).

Centralidade de autovetor

$$\lambda x = Ax, \text{ equivalente a } (A - \lambda I)x = 0.$$

Precisamos que x seja **não nulo** (de preferência, não negativo).

- Da álgebra linear, sabemos que x deve ser um **autovetor** (vetor próprio) da matriz A associado a um **autovalor** (valor próprio) λ.

Centralidade de autovetor

$$\lambda x = Ax, \text{ equivalente a } (A - \lambda I)x = 0.$$

Precisamos que x seja **não nulo** (de preferência, não negativo).

- Da álgebra linear, sabemos que x deve ser um **autovetor** (vetor próprio) da matriz A associado a um **autovalor** (valor próprio) λ.
- Sabemos que a matriz de adjacência pode ter **diversos** (até n) **valores próprios**, com diferentes vetores próprios associados.

Centralidade de autovetor

$$\lambda x = Ax, \text{ equivalente a } (A - \lambda I)x = 0.$$

Precisamos que x seja **não nulo** (de preferência, não negativo).

- Da álgebra linear, sabemos que x deve ser um **autovetor** (vetor próprio) da matriz A associado a um **autovalor** (valor próprio) λ.
- Sabemos que a matriz de adjacência pode ter **diversos** (até n) **valores próprios**, com diferentes vetores próprios associados.
- Felizmente, dado que a matriz de adjacência A é **não negativa**, e considerando que o grafo é **conexo**, o teorema de Perron–Frobenius mostra que existe um autovalor, real e positivo, que sempre tem um autovetor não negativo associado.

Centralidade de autovetor

$$\lambda x = Ax, \text{ equivalente a } (A - \lambda I)x = 0.$$

Precisamos que x seja **não nulo** (de preferência, não negativo).

- Da álgebra linear, sabemos que x deve ser um **autovetor** (vetor próprio) da matriz A associado a um **autovalor** (valor próprio) λ.
- Sabemos que a matriz de adjacência pode ter **diversos** (até n) **valores próprios**, com diferentes vetores próprios associados.
- Felizmente, dado que a matriz de adjacência A é **não negativa**, e considerando que o grafo é **conexo**, o teorema de Perron–Frobenius mostra que existe um autovalor, real e positivo, que sempre tem um autovetor não negativo associado.
- É o **maior autovalor** da matriz A, e ele é o **único** com essas características.

Centralidade de autovetor

- **Em resumo**, o **maior autovalor** da matriz de adjacência A de um grafo conexo tem associado um **autovetor** $x = (x_1, x_2, \ldots, x_n)$, **não negativo**, que cumpre:

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

Centralidade de autovetor

- **Em resumo**, o **maior autovalor** da matriz de adjacência A de um grafo conexo tem associado um **autovetor** $x = (x_1, x_2, \ldots, x_n)$, **não negativo**, que cumpre:

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

- Cada valor x_i representa o **índice de centralidade relativa** do vértice v_i na rede: v_i é importante quando ele está conectado a outros vértices importantes.

Centralidade de autovetor

- **Em resumo**, o **maior autovalor** da matriz de adjacência A de um grafo conexo tem associado um **autovetor** $x = (x_1, x_2, \ldots, x_n)$, **não negativo**, que cumpre:

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

- Cada valor x_i representa o **índice de centralidade relativa** do vértice v_i na rede: v_i é importante quando ele está conectado a outros vértices importantes.
- Como todos os vetores **proporcionais** αx são autovetores, precisamos escolher um deles. Geralmente escolhemos aquele **normalizado**, no qual a soma dos valores de cada x_i é igual à unidade.

Centralidade de autovetor

- **Em resumo**, o **maior autovalor** da matriz de adjacência A de um grafo conexo tem associado um **autovetor** $x = (x_1, x_2, \ldots, x_n)$, **não negativo**, que cumpre:

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

- Cada valor x_i representa o **índice de centralidade relativa** do vértice v_i na rede: v_i é importante quando ele está conectado a outros vértices importantes.
- Como todos os vetores **proporcionais** αx são autovetores, precisamos escolher um deles. Geralmente escolhemos aquele **normalizado**, no qual a soma dos valores de cada x_i é igual à unidade.
- Este valor x_i é chamado **centralidade de autovetor** do vértice v_i na rede.

Centralidade de autovetor

A **centralidade do autovetor** foi descoberta e redescoberta diversas vezes, em contextos diferentes:

Centralidade de autovetor

A **centralidade do autovetor** foi descoberta e redescoberta diversas vezes, em contextos diferentes:

- E. Landau: "On the relative value of tournament results" (1895).

Centralidade de autovetor

A **centralidade do autovetor** foi descoberta e redescoberta diversas vezes, em contextos diferentes:

- E. Landau: "On the relative value of tournament results" (1895).
- T. Wei: "The algebraic foundations of ranking theory" (Cambridge, 1952).

Centralidade de autovetor

A **centralidade do autovetor** foi descoberta e redescoberta diversas vezes, em contextos diferentes:

- E. Landau: "On the relative value of tournament results" (1895).
- T. Wei: "The algebraic foundations of ranking theory" (Cambridge, 1952).
- M. Kendall: "Further contributions to the theory of paired comparisons" (1955).

Centralidade de autovetor

Que **algoritmos** podem ser usados para calcular x, o autovetor de centralidades?

- Algoritmo mais usado: **método das potências** ("Iteração de **Von Mises**").
  É um algoritmo iterativo. Inicia-se com um vetor aleatório b_0, e em cada iteração o atualiza multiplicando por A e normalizando: $b_{k+1} = \frac{A \cdot b_k}{\|A \cdot b_k\|}$.

Centralidade de autovetor

Que **algoritmos** podem ser usados para calcular x, o autovetor de centralidades?

- Algoritmo mais usado: **método das potências** ("Iteração de **Von Mises**").
  É um algoritmo iterativo. Inicia-se com um vetor aleatório b_0, e em cada iteração o atualiza multiplicando por A e normalizando: $b_{k+1} = \frac{A \cdot b_k}{\|A \cdot b_k\|}$.
- É **muito eficiente** para grafos esparsos grandes (com matrizes de adjacência esparsas), se a implementação é apropriada.

Observação: Existem formas de lidar com matrizes grandes e esparsas sem usar $O(n^2)$ de espaço!

Centralidade de autovetor: algoritmos

Algoritmo para calcular o autovetor de centralidades: **método das potências**.

- A convergência tipicamente ocorre com poucas (algumas dezenas de) iterações.

Centralidade de autovetor: algoritmos

Algoritmo para calcular o autovetor de centralidades: **método das potências**.

- A convergência tipicamente ocorre com poucas (algumas dezenas de) iterações.
- A convergência é mais lenta apenas em casos quando o segundo maior autovalor λ_2 é muito próximo do maior autovalor λ_1 da matriz A, ou seja, se $\lambda_1/\lambda_2 \approx 1$.

Centralidade de autovetor: algoritmos

Algoritmo para calcular o autovetor de centralidades: **método das potências**.

- A convergência tipicamente ocorre com poucas (algumas dezenas de) iterações.
- A convergência é mais lenta apenas em casos quando o segundo maior autovalor λ_2 é muito próximo do maior autovalor λ_1 da matriz A, ou seja, se $\lambda_1/\lambda_2 \approx 1$.
- É um método numérico. Ele pode ser **generalizado** para que a matriz A tenha valores reais que representam a força de cada vínculo em grafos ponderados.

Centralidade de autovetor: algoritmos

Algoritmo para calcular o autovetor de centralidades: **método das potências**.

- A convergência tipicamente ocorre com poucas (algumas dezenas de) iterações.
- A convergência é mais lenta apenas em casos quando o segundo maior autovalor λ_2 é muito próximo do maior autovalor λ_1 da matriz A, ou seja, se $\lambda_1/\lambda_2 \approx 1$.
- É um método numérico. Ele pode ser **generalizado** para que a matriz A tenha valores reais que representam a força de cada vínculo em grafos ponderados.

Observação: Na prática, o algoritmo é muito rápido. A velocidade de convergência é exponencial, e varia aproximadamente como $(\lambda_1/\lambda_2)^k$.

Centralidade de autovetor: variantes

Existem diversas **variantes** da centralidade de autovetor:

- Uma das mais conhecidas é a **centralidade de Katz**, *Katz centrality*, na qual a influência de um vizinho de v na centralidade de v é proporcional a uma fração α, a de um vizinho do vizinho é proporcional a α^2...

Centralidade de autovetor: variantes

Existem diversas **variantes** da centralidade de autovetor:

- Uma das mais conhecidas é a **centralidade de Katz**, *Katz centrality*, na qual a influência de um vizinho de v na centralidade de v é proporcional a uma fração α, a de um vizinho do vizinho é proporcional a α^2...
  - **Formula**: $x_i = \sum_{k=1}^{\infty} \sum_{j=1}^{N} \alpha^k (A^k)_{ji}$. O valor de α precisa ser menor ou igual do que o recíproco do maior autovalor de A.

Centralidade de autovetor: variantes

Existem diversas **variantes** da centralidade de autovetor:

- Uma das mais conhecidas é a **centralidade de Katz**, *Katz centrality*, na qual a influência de um vizinho de v na centralidade de v é proporcional a uma fração α, a de um vizinho do vizinho é proporcional a α^2...
  - **Formula**: $x_i = \sum_{k=1}^{\infty} \sum_{j=1}^{N} \alpha^k (A^k)_{ji}$. O valor de α precisa ser menor ou igual do que o recíproco do maior autovalor de A.
- **Centralidade alpha**: quase idêntica à **centralidade de Katz**, difere em um fator constante que não muda a ordem de centralidade dos vértices.

Centralidade de autovetor: variantes

Existem diversas **variantes** da centralidade de autovetor:

- Uma das mais conhecidas é a **centralidade de Katz**, *Katz centrality*, na qual a influência de um vizinho de v na centralidade de v é proporcional a uma fração α, a de um vizinho do vizinho é proporcional a α^2...
  - **Formula**: $x_i = \sum_{k=1}^{\infty} \sum_{j=1}^{N} \alpha^k (A^k)_{ji}$. O valor de α precisa ser menor ou igual do que o recíproco do maior autovalor de A.
- **Centralidade alpha**: quase idêntica à **centralidade de Katz**, difere em um fator constante que não muda a ordem de centralidade dos vértices.
- Talvez a variante mais conhecida e usada é o **PageRank**.

Resumo

PageRank

Centralidade **PageRank**:

- Pode ser vista como uma **variante** ou evolução da centralidade de autovetor.

PageRank

Centralidade **PageRank**:

- Pode ser vista como uma **variante** ou evolução da centralidade de autovetor.
- Veremos a sua versão mais conhecida e usada, para grafos **direcionados**.

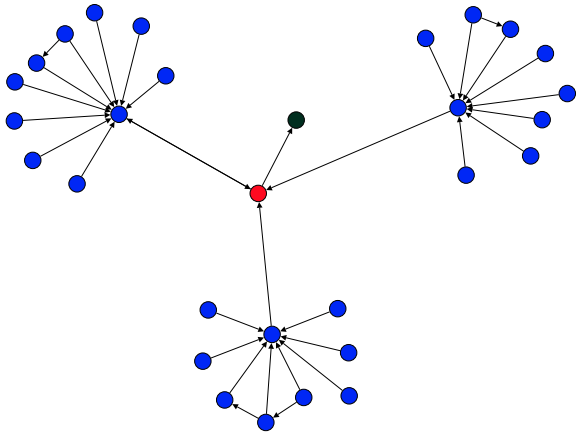
PageRank

Centralidade **PageRank**:

- Pode ser vista como uma **variante** ou evolução da centralidade de autovetor.
- Veremos a sua versão mais conhecida e usada, para grafos **direcionados**.
- **Ideia**: um vértice v_i tem **centralidade alta** quando ele recebe **muitos arcos** ("votos") de **outros vértices com centralidade alta**, considerando que quanto **maior** o número de arcos que saem do vértice, **menor** é a importância ou valor de cada arco ("voto").

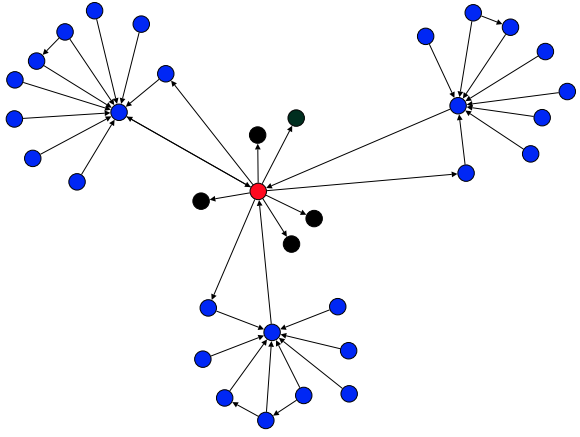
PageRank

O vértice preto recebe um "voto" do vértice vermelho, que é importante.



PageRank

Os vértices pretos recebem "votos" do vértice vermelho, que é importante.



PageRank

Centralidade $x_i = PR(v_i)$ do vértice v_i no PageRank:

$$x_i = PR(v_i) = \frac{1-\alpha}{N} + \alpha \sum_j a_{ji} \frac{x_j}{L_j}.$$

PageRank

Centralidade $x_i = PR(v_i)$ do vértice v_i no PageRank:

$$x_i = PR(v_i) = \frac{1-\alpha}{N} + \alpha \sum_j a_{ji} \frac{x_j}{L_j}.$$

Diferenças fundamentais com a centralidade do autovetor:

- Divisão por $L_j = \sum_i a_{ji}$, que é o **grau de saída** do vértice v_j.

PageRank

Centralidade $x_i = PR(v_i)$ do vértice v_i no PageRank:

$$x_i = PR(v_i) = \frac{1-\alpha}{N} + \alpha \sum_j a_{ji} \frac{x_j}{L_j}.$$

Diferenças fundamentais com a centralidade do autovetor:

- Divisão por $L_j = \sum_i a_{ji}$, que é o **grau de saída** do vértice v_j.
- Presença da componente $\frac{1-\alpha}{N}$. O que é ela representa?

PageRank

$$x_i = PR(v_i) = \frac{1-\alpha}{N} + \alpha \sum_j a_{ji} \frac{x_j}{L_j}.$$

Estamos supondo que todos os vértices, **mesmo com grau de entrada** igual a **zero**, que não receberam "votos", possuem um valor **"básico"** de centralidade $\frac{1-\alpha}{N}$:

PageRank

$$x_i = PR(v_i) = \frac{1-\alpha}{N} + \alpha \sum_j a_{ji} \frac{x_j}{L_j}.$$

Estamos supondo que todos os vértices, **mesmo com grau de entrada** igual a **zero**, que não receberam "votos", possuem um valor **"básico"** de centralidade $\frac{1-\alpha}{N}$:

- Ele é inversamente proporcional à quantidade de nós na rede N;

PageRank

$$x_i = PR(v_i) = \frac{1-\alpha}{N} + \alpha \sum_j a_{ji} \frac{x_j}{L_j}.$$

Estamos supondo que todos os vértices, **mesmo com grau de entrada** igual a **zero**, que não receberam "votos", possuem um valor **"básico"** de centralidade $\frac{1-\alpha}{N}$:

- Ele é inversamente proporcional à quantidade de nós na rede N;
- Ele é diretamente proporcional à $1 - \alpha$.

  O valor α é chamado de **fator de amortecimento**.

PageRank

$$x_i = PR(v_i) = \frac{1-\alpha}{N} + \alpha \sum_j a_{ji} \frac{x_j}{L_j}.$$

Interpretação do PageRank:

O valor do PageRank do vértice v_i é a probabilidade de chegar a v_i durante um passeio aleatório especial que começa em um vértice aleatório do grafo, e no qual há uma determinada probabilidade que o passeio possa ser reiniciado a qualquer momento. Se chagamos a um sorvedouro, reiniciamos o passeio.

PageRank

$$x_i = PR(v_i) = \frac{1-\alpha}{N} + \alpha \sum_j a_{ji} \frac{x_j}{L_j}.$$

Interpretação do PageRank:

> *O valor do PageRank do vértice v_i é a probabilidade de chegar a v_i durante um passeio aleatório especial que começa em um vértice aleatório do grafo, e no qual há uma determinada probabilidade que o passeio possa ser reiniciado a qualquer momento. Se chagamos a um sorvedouro, reiniciamos o passeio.*

Com probabilidade α, nós **continuamos** o passeio, escolhendo o próximo vértice do passeio dentre os adjacentes, com igual chance.

PageRank

$$x_i = PR(v_i) = \frac{1-\alpha}{N} + \alpha \sum_j a_{ji} \frac{x_j}{L_j}.$$

Interpretação do PageRank:

> *O valor do PageRank do vértice v_i é a probabilidade de chegar a v_i durante um passeio aleatório especial que começa em um vértice aleatório do grafo, e no qual há uma determinada probabilidade que o passeio possa ser reiniciado a qualquer momento. Se chagamos a um sorvedouro, reiniciamos o passeio.*

Com probabilidade α, nós **continuamos** o passeio, escolhendo o próximo vértice do passeio dentre os adjacentes, com igual chance.

Com probabilidade $1 - \alpha$, nós **reiniciamos** o passeio em um vértice aleatório, com chance $\frac{1}{N}$ do vértice v_i ser escolhido neste caso.

PageRank

PageRank e a equivalência com as cadeias de Markov:

- O grafo pode ser visto como uma **cadeia de Markov** na qual os estados são os vértices, e as transições são os arcos.

PageRank

PageRank e a equivalência com as cadeias de Markov:

- O grafo pode ser visto como uma **cadeia de Markov** na qual os estados são os vértices, e as transições são os arcos.
- É possível provar que o valor do **PageRank** de um vértice é a probabilidade de chegar a esse vértice após um número grande (infinito) de passos em um processo de Markov.

PageRank

$$x_i = PR(v_i) = \frac{1-\alpha}{N} + \alpha \sum_j a_{ji} \frac{x_j}{L_j}.$$

Algoritmos: cálculo **iterativo** do PageRank:

- No inicio, $x_i(t=0) = \frac{1}{N}$ para todo i.
- Em cada passo, precisamos ter $x_i(t+1) = \frac{1-\alpha}{N} + \alpha \sum_j a_{ji} \frac{x_j(t)}{L_j}$.

PageRank

$$x_i = PR(v_i) = \frac{1-\alpha}{N} + \alpha \sum_j a_{ji} \frac{x_j}{L_j}.$$

Algoritmos: cálculo **iterativo** do PageRank:

- A partir da matriz de adjacência A, é possível construir uma matriz $\widehat{\mathcal{M}}$ de forma que $x(t+1) = \widehat{\mathcal{M}} x(t)$ (J_N denota a matriz $N \times N$ de 1's).

$$\mathcal{M}_{ij} = \begin{cases} 1/L_j, & \text{se existe arco } (v_j, v_i) \\ 0, & \text{caso contrário} \end{cases}; \qquad \widehat{\mathcal{M}} = \alpha \mathcal{M} + \frac{1-\alpha}{N} J_N.$$

PageRank

$$x_i = PR(v_i) = \frac{1-\alpha}{N} + \alpha \sum_j a_{ji} \frac{x_j}{L_j}.$$

Algoritmos: cálculo **iterativo** do PageRank:

- Usando a matriz $\widehat{\mathcal{M}}$, executamos uma série de iterações $x(t+1) = \widehat{\mathcal{M}} x(t)$.

PageRank

$$x_i = PR(v_i) = \frac{1-\alpha}{N} + \alpha \sum_j a_{ji} \frac{x_j}{L_j}.$$

Algoritmos: cálculo **iterativo** do PageRank:

- Usando a matriz $\widehat{\mathcal{M}}$, executamos uma série de iterações $x(t+1) = \widehat{\mathcal{M}}x(t)$.
- O vetor solução x é o autovetor associado a ao **maior autovalor** da matriz $\widehat{\mathcal{M}}$.

PageRank

$$x_i = PR(v_i) = \frac{1-\alpha}{N} + \alpha \sum_j a_{ji} \frac{x_j}{L_j}.$$

Algoritmos: cálculo **iterativo** do PageRank:

- Usando a matriz $\widehat{\mathcal{M}}$, executamos uma série de iterações $x(t+1) = \widehat{\mathcal{M}}x(t)$.
- O vetor solução x é o autovetor associado a ao **maior autovalor** da matriz $\widehat{\mathcal{M}}$.
- O valor geralmente escolhido para α é 0,85.

PageRank

$$x_i = PR(v_i) = \frac{1-\alpha}{N} + \alpha \sum_j a_{ji} \frac{x_j}{L_j}.$$

Algoritmos: cálculo **iterativo** do PageRank:

- Usando a matriz $\widehat{\mathcal{M}}$, executamos uma série de iterações $x(t+1) = \widehat{\mathcal{M}}x(t)$.
- O vetor solução x é o autovetor associado a ao **maior autovalor** da matriz $\widehat{\mathcal{M}}$.
- O valor geralmente escolhido para α é 0,85.

Observação: devido à forma como foi construída a matriz $\widehat{\mathcal{M}}$, os valores de PageRank podem ser obtidos com grande precisão após poucas iterações, pois a diferença entre os dois maiores autovalores **não é muito pequena**.

PageRank

Um pouco de **história**:

- L. Page e S. Brin desenvolveram PageRank em **Stanford**, em 1996, como parte de um **projeto de pesquisa**.

PageRank

Um pouco de **história**:

- L. Page e S. Brin desenvolveram PageRank em **Stanford**, em 1996, como parte de um **projeto de pesquisa**.
- Foi o **primeiro algoritmo** usado quando eles fundaram **Google** em 1998.

PageRank

Um pouco de **história**:

- L. Page e S. Brin desenvolveram PageRank em **Stanford**, em 1996, como parte de um **projeto de pesquisa**.
- Foi o **primeiro algoritmo** usado quando eles fundaram **Google** em 1998.
- Hoje, PageRank **não é** mais o **único algoritmo** usado para ordenar os resultados de uma pesquisa no Google, mas ele **ainda é usado** com algumas modificações.

PageRank

Um pouco de **história**:

- L. Page e S. Brin desenvolveram PageRank em **Stanford**, em 1996, como parte de um **projeto de pesquisa**.
- Foi o **primeiro algoritmo** usado quando eles fundaram **Google** em 1998.
- Hoje, PageRank **não é** mais o **único algoritmo** usado para ordenar os resultados de uma pesquisa no Google, mas ele **ainda é usado** com algumas modificações.
- No dia 24 de setembro de 2019, todas as **patentes** associadas ao PageRank **expiraram**.

PageRank

PageRank: **eficiência** do algoritmo:

- No artigo original, Brin e Page afirmaram que o algoritmo para calcular PageRank para uma rede de 322 milhões de arcos converge (com precisão suficiente) com apenas 52 iterações.

PageRank

PageRank: **eficiência** do algoritmo:

- No artigo original, Brin e Page afirmaram que o algoritmo para calcular PageRank para uma rede de 322 milhões de arcos converge (com precisão suficiente) com apenas 52 iterações.
- Se implementado adequadamente, usando estruturas de dados eficientes, a complexidade é aproximadamente $O(E \cdot k)$, onde k é o número de iterações até a convergência.

PageRank

PageRank: **eficiência** do algoritmo:

- No artigo original, Brin e Page afirmaram que o algoritmo para calcular PageRank para uma rede de 322 milhões de arcos converge (com precisão suficiente) com apenas 52 iterações.
- Se implementado adequadamente, usando estruturas de dados eficientes, a complexidade é aproximadamente $O(E \cdot k)$, onde k é o número de iterações até a convergência.
- Em redes muito grandes, o número de iterações k necessário para a convergência cresce aproximadamente como $O(\log n)$, onde n é o número de vértices.

PageRank

Casos de uso do PageRank:

- Com algumas modificações, PageRank **ainda é usado** para ordenar os resultados de uma pesquisa no **Google**. Outros fatores contribuem para a classificação dos resultados: localização geográfica, preferências do usuário, entre outros.

PageRank

Casos de uso do PageRank:

- Com algumas modificações, PageRank **ainda é usado** para ordenar os resultados de uma pesquisa no **Google**. Outros fatores contribuem para a classificação dos resultados: localização geográfica, preferências do usuário, entre outros.
- Existem abordagens semelhantes ao PageRank para quantificar de forma mais justa o **impacto científico de pesquisadores**, criando uma classificação para publicações individuais e para seus autores.

PageRank

Casos de uso do PageRank:

- Com algumas modificações, PageRank **ainda é usado** para ordenar os resultados de uma pesquisa no **Google**. Outros fatores contribuem para a classificação dos resultados: localização geográfica, preferências do usuário, entre outros.
- Existem abordagens semelhantes ao PageRank para quantificar de forma mais justa o **impacto científico de pesquisadores**, criando uma classificação para publicações individuais e para seus autores.
- **Twitter** usa uma variante do PageRank para apresentar aos usuários **sugestões de contas** a serem seguidas.

Material bibliográfico

F. A. Rodrigues: “Network centrality: an introduction” (2019).

M. Newman: ”The mathematics of networks”, (2008).

S.Brin & L. Page: “The Anatomy of a Large-Scale Hypertextual Web Search Engine”, (1998).

Dúvidas

Dúvidas?