Universidade Federal do Piauí

Centro de Ciências da Natureza

Pós-graduação em Matemática

Mestrado Profissional em Matemática - PROFMAT

Aplicação Matemática no Mecanismo de Pesquisa do Google

Rubens de Carvalho Almondes

Teresina - 2016

Rubens de Carvalho Almondes

Dissertação de Mestrado:

Aplicação Matemática no Mecanismo de Pesquisa do Google

Dissertação submetida ao Programa de Pós-Graduação - Mestrado Profissional em Matemática em Rede Nacional como requisito parcial para a obtenção do grau de Mestre em Matemática.

Orientador:
Prof. Dr. Paulo Alexandre Araújo Sousa

Teresina - 2016

Copia da folha de rosto assinada pelos membros da banca examinadora.

PROFMAT

UNIVERSIDADE FEDERAL DO PIAUÍ
CENTRO DE CIÊNCIAS DA NATUREZA
CENTRO DE EDUCAÇÃO ABERTA E À DISTÂNCIA
MESTRADO PROFISSIONAL EM MATEMÁTICA EM REDE NACIONAL

SBM

Dissertação de Mestrado submetida à coordenação Acadêmica Institucional, na Universidade Federal do Piauí, do Programa de Mestrado Profissional em Matemática em Rede Nacional para obtenção do grau de **mestre em matemática** intitulada: **Aplicação Matemática no Mecanismo de Pesquisa do Google,** defendida por **Rubens de Carvalho Almondes** em 18/08/2016 e aprovada pela banca constituída pelos professores:

Paulo Alexandre Araújo Sousa
Presidente da Banca Examinadora

Examinador

Antonio Francisco de Oliveira Veloso
Examinador Externo

Ficha catalográfica editada pela biblioteca setorial do CCN.

Almondes, R. C.

xxxx Aplicação Matemática no Mecanismo de Pesquisa do Google.

Teresina, 2016.

Orientador: Prof. Dr. Paulo Alexandre Araújo Sousa

1. Área de Concentração

CDD 516.36

Dedico este trabalho a meus avós; Vó Maria (In memoriam), Vô Manoel (In memoriam), Vó Teresa e Vô José (In memoriam).

Agradecimentos

Agradeço primeiramente a Deus, Pai Criador, pelo dom da vida vivenciado a cada novo amanhecer;

Agradeço a minha família: pai (Rufino), mãe (Bernadete), esposa (Jéssica) e irmãs (Rute, Raquel, Rafaela, Rebeca e Renata), pelo irrestrito incentivo e motivação diária, principalmente nos momentos mais difíceis;

Agradeço ao meu orientador, Prof. Dr. Paulo Alexandre Araújo Sousa, pela paciência e insistência durante todo processo;

Agradeço aos amigos de turma que caminharam, vivenciaram comigo e, por muitas vezes, me motivaram a prosseguir essa jornada. De modo especial e fraterno Bruno, Delano, Gideone, Gilson, Huerllen, Jerson, Pedro, Perivaldo, Queiroz, Raimundo, Renato, Renné, Samuel e Viviam;

Agradeço aos amigos de Picos-PI que entenderam minha ausência mas nunca deixaram de acreditar e se fazerem presentes em meu dia a dia;

Agradeço ao IFPI, instituição a qual faço parte e que permitiu a minha presença semanalmente aos encontros com a flexibilização de meus horários de trabalho;

Agradeço À PROFMAT/UFPI que oportunizou a realização deste sonho;

Agradeço aos professores da UFPI que deram o suporte necessário para a concretização desta jornada;

Agradeço a CAPES pelo apoio financeiro que foi de fundamental importância para a concretização deste momento.

"*Nota-se entre os matemáticos, uma imaginação assombrosa... Repetimos: Havia mais imaginação na cabeça de Arquimedes do que na de Homero*".

François Marie Arouet (Voltaire).

Resumo

Neste trabalho trataremos de forma simplificada os conceitos matemáticos envolvidos em um dos algoritmos utilizados no sistema de classificação das páginas pelo Google: o *PageRank*. Abordaremos de maneira informal, a ideia de se atribuir uma pontuação de importância para as páginas da internet e posteriormente formalizaremos matematicamente, com o auxílio da Álgebra Linear, bem como do ponto de vista probabilístico. O presente trabalho tem por objetivo estimular os estudantes ao aprendizado de temas como Matrizes, Determinantes e Sistemas de Equações Lineares tendo como plano de fundo a internet, a qual possui uma presença quase que constante na vida dos alunos e assim servir como fonte de inspiração para o aprendizado da matemática.

Palavras-chave: Internet. Google. PageRank. Álgebra Linear

Abstract

In this deal work simplified mathematical concepts involved in one of the algorithms used in the classification system of the pages by Google: the *PageRank*. We discuss informally the idea of assigning a score of importance to websites and then mathematically formalize, with the help of linear algebra and probabilistic point of view. This work aims to encourage students to learning topics such as matrices, determinants and linear equations systems having as background the internet, which has a presence almost constant in the lives of students and thus serve as a source of inspiration for learning of mathematics.

Keywords: Internet. Google. PageRank. Linear Algebra

Sumário

Capítulo 1

Introdução

Durante as duas últimas décadas assistimos a um impulso tecnológico. No ano de 1994, haviam menos de 3 mil sites. Cerca de 20 anos depois, em 2014, a marca de 1 bilhão de sites em todo o mundo é alcançada, no entanto, devido ao número de páginas desativadas, houve um decréscimo, conforme figura 1.1. No início de 2016 a marca de 1 bilhão foi alcançada novamente e desde março estabilizou acima dessa quantidade, de acordo com o site *Internet Live Stats*, que fornece estatísticas em tempo real sobre o uso da rede.

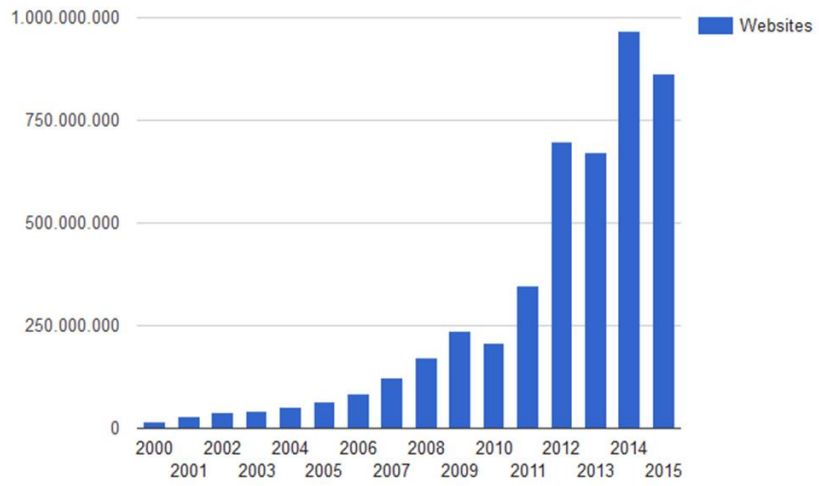


Figura 1.1: 1.50.5 Quantidade de sites na internet
Fonte: http://www.internetlivestats.com/total-number-of-websites/

É notório o crescimento exponencial da Internet. Mesmo navegando de forma ocasional é o suficiente para que sejamos convencidos de que há uma enorme quantidade de informações e *links* disponíveis. No entanto, todas estas informações se tornam inúteis,

a menos que tenhamos uma forma de classificação e pesquisa. Vivemos na era da informação, a internet faz parte de nossas vidas e boa parte desta informação fica a apenas um clique de distância. Basta abrirmos um site de busca, como por exemplo Google, Bing ou Yahoo, digitar a palavra ou termo que se deseja procurar, que serão exibidos uma lista de páginas relevantes para a sua pesquisa. Mas como é que um site de busca realmente funciona? Desde o primeiro mecanismo de busca, no início da década de 1990, para os modernos motores de busca que usamos hoje, o problema de decidir a relevância das informações disponíveis *on-line* tem sido uma questão crucial.

É razoável imaginar que um site de busca mantém um índice de todas as páginas da web, e quando o usuário digita uma pesquisa, o mecanismo de busca contido no site navega através de seu índice e conta as ocorrências das palavras, ou termos, digitados em cada arquivo web. As páginas com maior número de ocorrência das palavras pesquisadas serão exibidas para o usuário.

Isto costumava ser a imagem correta no início dos anos 1990, quando os primeiros motores de busca utilizavam sistemas de classificação baseados em textos para decidir quais páginas que são mais relevantes em uma determinada pesquisa. Existia, no entanto, uma série de problemas com essa abordagem. Uma pesquisa sobre um termo comum, como "internet" era problemático, bem como a possibilidade de encontrar páginas contendo centenas do mesmo termo pesquisado. Além disso, suponha que queiramos encontrar alguma informação sobre UFPI. Ao fazer uma busca sobre "UFPI" esperamos que "www.ufpi.br" seria o resultado mais relevante para a nossa consulta. Contudo, pode haver milhões de páginas na web que usam termo UFPI e "www.ufpi.br" pode não ser onde o termo apareça com tanta frequência. Suponhamos ainda que decidimos construir site que contenha apenas a palavra "UFPI" repetida inúmeras de vezes, então para um motor de busca que conta o número de ocorrências das palavras, o nosso site seria o primeiro a ser exibido, o que não seria relevante para quem busca.

Dentre todos os buscadores, o Google se destacou adquirindo uma supremacia em questões de meses, tudo isto graças ao seu algoritmo de classificação de resultados: o algoritmo *PageRank*.

O presente trabalho tem por objetivo o estudo de matrizes, apresentando a aplicação em um dos mecanismo de pesquisa mais utilizados no mundo. A opção de abordar no contexto da internet, mas especificamente o mecanismo de pesquisa do Google, e utilizá-lo

com plano de fundo, serve de motivação para o aprofundamento do estudo da Álgebra Linear, visto que é algo presente em nosso dia a dia e a curiosidade do "como funciona"faz aguçar nossa imaginação e motivar o aprofundamento do tema.

No capítulo 2 serão abordados os conceitos de Vetores, Matrizes, Autovetores e Autovalores, bem como algumas propriedades que servirão de base para o nosso estudo. O capítulo 3 introduz a definição de Matrizes Coluna Estocástica e sua relação com o processo aleatório denominado Cadeias de Markov aplicando-os para modelarmos o problema de encontrar uma pontuação de importância para cada uma das páginas da internet. No capitulo seguinte é descrito o algorítimo do mecanismo de pesquisa do Google, o *PageRank*, tanto do ponto de vista probabilístico como da Álgebra Linear. Por fim, no último capítulo é destacado a importância da Álgebra Linear na construção de modelos matemáticos para o desenviolamento de diversas áreas.

Capítulo 2

Noções Preliminares

Este capítulo é destinado às noções preliminares onde abordaremos as definições de vetores e matrizes, bem como algumas propriedades e operações, de modo a dar suporte ao desenvolvimento do presente trabalho. Para os leitores que sentirem necessidade de uma explanação mais detalhada dos conceitos citados acima, além daquelas apresentadas no decorrer do texto, indicamos a leitura de [1], [3] e [6]. Para tal, denotemos o conjunto dos números naturais $\{1, 2, 3, \cdots\}$ por $\mathbb{N}$, assim como o conjunto de todos os números reais será representado por $\mathbb{R}$.

2.1 Vetores

Um vetor ν, de comprimento $\mathfrak{n}$, é simplesmente uma lista ordenada contendo $\mathfrak{n}$ elementos, com $\mathfrak{n} \in \mathbb{N}$, sendo estes chamados entradas do vetor. Chamamos de vetor coluna quando a lista de suas entradas é apresentado na vertical. A notação de um vetor coluna ν de entradas $\mathfrak{x}_1, \mathfrak{x}_2, \cdots, \mathfrak{x}_\mathfrak{n}$ é exibida a seguir:

$$\nu = \begin{bmatrix} \mathfrak{x}_1 \\ \mathfrak{x}_2 \\ \vdots \\ \mathfrak{x}_\mathfrak{n} \end{bmatrix}.$$

A princípio, as entradas de um vetor podem ser quaisquer objetos, desde palavras,

cores, números, etc. São exemplos de vetores coluna:

$$\begin{bmatrix} 2 \\ 5 \\ -1 \\ 0 \end{bmatrix}, \begin{bmatrix} \blacklozenge \\ \spadesuit \\ \clubsuit \\ \bigstar \end{bmatrix}, \begin{bmatrix} \text{exemplo} \\ \text{de} \\ \text{um} \\ \text{vetor} \end{bmatrix}.$$

Neste trabalho vamos apenas olhar para vetores cujas entradas são números reais, ou seja, as entradas $x_1, x_2, \cdots, x_n \in \mathbb{R}$.

Definição 2.1. *Dois vetores* u *e* v*, de mesmo tamanho* n*, entradas* $x_1, x_2, \cdots, x_n$ *e* $y_1, y_2, \cdots, y_n$*, respectivamente, são ditos iguais, se e somente se, todas as suas entradas de mesma posição forem iguais, ou seja,* $x_i = y_i$ *para todo* $i \in \mathbb{N}$.

Os vetores colunas de entradas reais são munidos das operações de soma e multiplicação por um número real, sendo este chamado de escalar. Para tal, consideremos os vetores u, v e w, todos de mesmo comprimento n e entradas reais $\{x_1, x_2, \cdots, x_n\}, \{y_1, y_2, \cdots, y_n\}$ e $\{z_1, z_2, \cdots, z_n\}$, respectivamente, onde:

I. O vetor coluna w é dito vetor soma de u e v, representado por $w = u + v$, se, e somente se, suas entradas forem $z_1 = x_1 + y_1, z_2 = x_2 + y_2, \cdots, z_n = x_n + y_n$, ilustrado a seguir:

$$w = u + v \Leftrightarrow \begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_n \end{bmatrix} = \begin{bmatrix} x_1 + y_1 \\ x_2 + y_2 \\ \vdots \\ x_n + y_n \end{bmatrix}.$$

A operação soma de vetores descrita a cima, possui as seguintes propriedades: comutativa, associativa, elemento neutro e elemento simétrico. Para tal, consideremos o vetores u, v, e w, todos de mesmo tamanho.

(a) Comutativa: $u + v = v + u$

(b) Associativa: $u + (v + w) = (u + v) + w$

(c) Elemento neutro: $u + v = u \Leftrightarrow x_i = 0 \in v$, para todo i. Daí $v = o$

(d) Elemento simétrico: $u + v = o \Rightarrow v = -u$

II. O vetor coluna w é o produto de um vetor u por um escalar $a \in \mathbb{R}$, representado por $w = a \cdot u$ se e somente se, suas entradas forem $z_1 = a \cdot x_1, z_2 = a \cdot x_2, \cdots, z_n = a \cdot x_n$, conforme ilustramos abaixo:

$$w = a \cdot u \Leftrightarrow \begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_n \end{bmatrix} = \begin{bmatrix} a \cdot x_1 \\ a \cdot x_2 \\ \vdots \\ a \cdot x_n \end{bmatrix}.$$

A operação de produto por um escalar real descrita a cima, com $a, b \in \mathbb{R}$ e os vetores u e v de mesmo tamanho, possui as seguintes propriedades:

(a) $a(u + v) = au + av$

(b) $(a + b)v = av + bv$

(c) $(ab)v = a(bv)$

(d) $1v = v$

Sob as notações acima, podemos facilmente verificar que:

$$\begin{bmatrix} 3 \\ -2 \\ 1 \\ 5 \end{bmatrix} + \begin{bmatrix} 1 \\ 2 \\ 0 \\ -2 \end{bmatrix} = \begin{bmatrix} 4 \\ 0 \\ 1 \\ 3 \end{bmatrix};$$

$$3 \cdot \begin{bmatrix} 2 \\ 1 \\ -2 \\ 4 \end{bmatrix} = \begin{bmatrix} 6 \\ 3 \\ -6 \\ 12 \end{bmatrix}.$$

2.2 Matrizes

Definição 2.2. *Dados dois números naturais m e n, não nulos, chama-se matriz real de ordem m por n (indica-se $m \times n$) toda tabela M formada por números reais distribuídos em m linhas, que são as filas horizontais, e n colunas, que são as filas verticais.*

Também podemos ver uma matriz como sendo o emparelhamento de n vetores colunas de comprimento m, um ao lado do outro. Para tal, denotaremos uma matriz $A = [a_{ij}]_{m\times n}$ de entradas reais a_{ij}, com $i \in \{1, 2, \cdots, m\}$ e $j \in \{1, 2, \cdots, n\}$ por:

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}.$$

Isso significa que na interseção da linha i com a coluna j encontramos a entrada $a_{ij} \in \mathbb{R}$. Por exemplo, na primeira linha, segunda coluna encontra-se o elemento a_{12}.

$$A = \begin{bmatrix} a_{11} & \textcircled{a_{12}} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}.$$

Observe que um vetor de comprimento m pode ser visto como uma matriz de ordem $m \times 1$, assim como um número $a \in \mathbb{R}$ também pode ser representado por uma matriz de ordem 1×1:

$$\begin{bmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{m1} \end{bmatrix} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{bmatrix} \text{ e } [a_{11}] = a.$$

Definição 2.3. *Dizemos que as matrizes* A *e* B *são iguais, se e somente se, tiverem a mesma ordem e seus elementos de mesma posição forem iguais, ou seja,* $A_{m\times n} = B_{m\times n} \Leftrightarrow a_{ij} = b_{ij}$*, para todo* i *e* j*,*

$$A = B \Leftrightarrow \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} = \begin{bmatrix} b_{11} & b_{12} & \cdots & b_{1n} \\ b_{21} & b_{22} & \cdots & b_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ b_{m1} & b_{m2} & \cdots & b_{mn} \end{bmatrix}.$$

Denominamos A de matriz quadrada de ordem n, sendo denotada por A_n, toda matriz tal que o número de linhas for igual ao número de coluna, ou seja, $m = n$. Em uma matriz quadrada A_n, é chamada diagonal da matriz A a fila que contém os elementos

nos quais sua posiçcão referente a linha e a coluna são as mesmas, ou seja, os elementos $a_{11}, a_{22}, \cdots, a_{nn}$.

$$A = \begin{bmatrix} \textcircled{a_{11}} & a_{12} & \cdots & a_{1n} \\ a_{21} & \textcircled{a_{22}} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & \textcircled{a_{nn}} \end{bmatrix}.$$

Uma matriz quadrada cujos os elementos fora da diagonal principal são todos iguais a zero é chamada de *matriz diagonal*, caso os elementos da diagonal desta matriz sejam iguais, denomina-se esta de *matriz escalar*. Se o escalar na diagonal for igual a 1, a matriz é chamada de *matriz identidade* e denotaremos por I_n, com $n \in \mathbb{N}$:

$$\underset{\text{Matriz Diagonal}}{\begin{bmatrix} a_{11} & 0 & 0 & \cdots & 0 \\ 0 & a_{22} & 0 & \cdots & 0 \\ 0 & 0 & a_{33} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & a_{nn} \end{bmatrix}}; \quad \underset{\text{Matriz Escalar}}{\begin{bmatrix} x & 0 & 0 & \cdots & 0 \\ 0 & x & 0 & \cdots & 0 \\ 0 & 0 & x & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & x \end{bmatrix}}; \quad \underset{\text{Matriz Identidade}}{\begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{bmatrix}}.$$

As operações de adição de matrizes e multiplicação de matriz por escalar, são feitas de forma semelhante às operações definidas com vetores. Adicionando as matrizes $A = [a_{ij}]_{m\times n}$ e $B = [b_{ij}]_{m\times n}$ retorna uma matriz $C = [c_{ij}]_{m\times n}$, de mesma ordem $m \times n$ e cuja entrada na linha i e coluna j igual à soma das entradas correspondentes de A e B, ou seja $c_{ij} = a_{ij} + b_{ij}$:

$$A+B = C \Leftrightarrow \begin{bmatrix} a_{11}+b_{11} & a_{12}+b_{12} & \cdots & a_{1n}+b_{1n} \\ a_{21}+b_{21} & a_{22}+b_{22} & \cdots & a_{2n}+b_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1}+b_{m1} & a_{m2}+b_{m2} & \cdots & a_{mn}+b_{mn} \end{bmatrix} = \begin{bmatrix} c_{11} & c_{12} & \cdots & c_{1n} \\ c_{21} & c_{22} & \cdots & c_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ c_{m1} & c_{m2} & \cdots & c_{mn} \end{bmatrix}.$$

Assim como a operação soma de vetores, a soma de matrizes possui as seguintes propriedades: comutativa, associativa, elemento neutro e elemento simétrico. Para tal, consideremos as matrizes A, B, e C, todas de mesmo tamanho $m \times n$.

I. Comutativa: $A + B = B + A$

II. Associativa: $A + (B + C) = (A + B) + C$

III. Elemento neutro: $A + B = A \Leftrightarrow b_{ij} = 0 \in \nu$, para todo i. Daí $B = O$

IV. Elemento simétrico: $A + B = O \Rightarrow B = -A$

Multiplicando matriz $A = [a_{ij}]_{m\times n}$ com um número real k é o mesmo que multiplicar cada elemento de A por o escalar k, obtendo uma matriz $B = [b_{ij}]_{m\times n}$ em que cada uma de suas entradas é dada por $b_{ij} = k \cdot a_{ij}$, para todo i e j:

$$k \cdot A = B \Leftrightarrow \begin{bmatrix} k\cdot a_{11} & k\cdot a_{12} & \cdots & k\cdot a_{1n} \\ k\cdot a_{21} & k\cdot a_{22} & \cdots & k\cdot a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ k\cdot a_{m1} & k\cdot a_{m2} & \cdots & k\cdot a_{mn} \end{bmatrix} = \begin{bmatrix} b_{11} & b_{12} & \cdots & b_{1n} \\ b_{21} & b_{22} & \cdots & b_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ b_{m1} & b_{m2} & \cdots & b_{mn} \end{bmatrix}.$$

A operação de produto por um escalar real descrita a cima, com $k, l \in \mathbb{R}$ e as matrizes A e B de mesmo tamanho $m \times n$, possui as seguintes propriedades:

I. $k(A + B) = kA + kB$

II. $(k + l)A = kA + lA$

III. $(kl)A = k(lA)$

IV. $1A = A$

Definição 2.4. *Dada duas matrizes $A = [a_{ij}]_{m\times n}$ e $B = [b_{jk}]_{n\times p}$, ambas de entradas reais, chama-se a matriz produto $A \cdot B$, a matriz $C_{m\times p}$ de entradas c_{ik} tal que, para todo $i \in \{1, 2, \cdots, m\}$ e todo $k \in \{1, 2, \cdots, p\}$, teremos:*

$$c_{ik} = a_{i1} \cdot b_{1k} + a_{i2} \cdot b_{2k} + \cdots + a_{i1} \cdot b_{nk} = \sum_{j=1}^{n} a_{ij} b_{jk}$$

, i.e.,

$$A \cdot B = C \Leftrightarrow \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \cdot \begin{bmatrix} b_{11} & b_{12} & \cdots & b_{1p} \\ b_{21} & b_{22} & \cdots & b_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ b_{n1} & b_{n2} & \cdots & b_{np} \end{bmatrix} =$$

$$\begin{bmatrix} a_{11}\cdot b_{11}+a_{12}\cdot b_{21}+\cdots+a_{1n}\cdot b_{n1} & \cdots & a_{11}\cdot b_{1p}+a_{12}\cdot b_{2p}+\cdots+a_{1n}\cdot b_{np} \\ a_{21}\cdot b_{11}+a_{22}\cdot b_{21}+\cdots+a_{2n}\cdot b_{n1} & \cdots & a_{21}\cdot b_{1p}+a_{22}\cdot b_{2p}+\cdots+a_{2n}\cdot b_{np} \\ \vdots & \ddots & \vdots \\ a_{m1}\cdot b_{11}+a_{m2}\cdot b_{21}+\cdots+a_{mn}\cdot b_{n1} & \cdots & a_{m1}\cdot b_{1p}+a_{m2}\cdot b_{2p}+\cdots+a_{mn}\cdot b_{np} \end{bmatrix} =$$

$$\begin{bmatrix} c_{11} & c_{12} & \cdots & c_{1p} \\ c_{21} & c_{22} & \cdots & c_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ c_{m1} & c_{m2} & \cdots & c_{mp} \end{bmatrix}.$$

É fácil ver que para existir o produto de duas matrizes é condição necessária e suficiente que o número de colunas da matriz que está à esquerda da operação seja igual ao de linhas da matriz que encontra-se a direita do operador, onde a matriz produto herda a quantidade de linhas da matriz à esquerda e o número de colunas da matriz à direita.

Podemos observar que, desde que as operações sejam possíveis, o produto de matrizes possui as propriedades operatórias:

I. Distributiva à esquerda: $A(B+C) = AB + AC$

II. Distributiva à direita: $(A+B)C = AC + BC$

III. Associativa: $A(BC) = (AB)C$

IV. Elemento Identidade: $AI = IA = A$

Também vale salientar que o produto de duas matrizes é uma operação que não goza da propriedade comutativa, ou seja, nem sempre é verdadeiro a igualdade $AB = BA$.

Definição 2.5. *Dada a matriz quadrada* A_n *e* $k \in \mathbb{N}$ *com* $k \geqslant 2$*, definimos potência natural de matrizes:*

$$A^0 = I_n,$$

$$A^1 = A,$$

$$\vdots$$

$$A^k = \underbrace{AA\cdots A}_{k \text{ } vezes}.$$

O produto de uma matriz de ordem $m \times n$ por um vetor com n entadas reais é feito tomando o vetor como sendo uma matriz coluna, ou seja, de ordem $n \times 1$, resultando um novo vetor coluna de tamanho m ou uma matriz de coluna $m \times 1$, que pode ser observado com clareza no exemplo a seguir:

$$\begin{bmatrix} 1 & -2 & 3 \\ 2 & 1 & 0 \end{bmatrix} \cdot \begin{bmatrix} -2 \\ 1 \\ -1 \end{bmatrix} = \begin{bmatrix} 1 \cdot (-2) + (-2) \cdot 1 + 3 \cdot (-1) \\ 2 \cdot (-2) + 1 \cdot 1 + 0 \cdot (-1) \end{bmatrix} = \begin{bmatrix} -7 \\ -3 \end{bmatrix}.$$

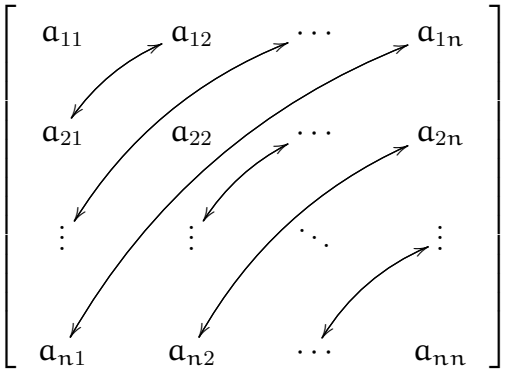
Definição 2.6. *Dada uma matriz* $A = [a_{ij}]_{m \times n}$*, chama-se transposta de* A *a matriz* $A^t = [a'_{ji}]_{n \times m}$ *tal que* $a'_{ji} = a_{ij}$*, para todo* i *e* j*.*

$$A^t = \begin{bmatrix} a'_{11} & a'_{12} & \cdots & a'_{1m} \\ a'_{21} & a'_{22} & \cdots & a'_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ a'_{n1} & a'_{n2} & \cdots & a'_{nm} \end{bmatrix} = \begin{bmatrix} a_{11} & a_{21} & \cdots & a_{m1} \\ a_{12} & a_{22} & \cdots & a_{m2} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1n} & a_{2n} & \cdots & a_{nm} \end{bmatrix}.$$

Na prática, como o próprio nome sugere, a matriz A^t é a transposição das colunas para linhas da matriz A e vice e versa.

Definição 2.7. *A matriz* $A = [a_{ij}]_{n \times n}$ *é uma matriz simétrica, se, e somente se,* $a_{ij} = a_{ji}$*.*

Ou seja, os elementos simetricamente dispostos em relação à digonal são iguais. Com isso, é fácil ver que toda matriz quadrada A_n, tal que $A = A^t$, é uma *matriz simétrica*.



A transposição da matriz A^t é a própria matriz A, em outras palavras, podemos escrever $(A^t)^t = A$. Também é possível verificar que, para quaisquer matrizes A, B, as propriedades de distributividade com relação a soma e com relação ao produto são válidas para transposição de matrizes:

I. $(A + B)^t = A^t + B^t$

II. $(A \cdot B)^t = B^t \cdot A^t$

2.3 Autovalores e Autovetores

Vamos agora introduzir duas noções importantes na teoria sobre matrizes: autovetor e autovalor.

> O adjetivo germânico *eigen* significa "próprio" ou "característico de". Valores próprios e vetores próprios, ou autovalores e autovetores, são característicos de uma matriz no sentido de conterem informações importantes sobre a natureza da matriz. A letra λ (lambda), letra grega equivalente ao L em português, é utilizada para designar autovalores porque anteriormente esses números também eram chamados de *valores latentes*. A pronúncia fonética do prefixo alemão *eingen* é "áiguen".(POOLE, 2004, p. 232)

Definição 2.8. *Seja A uma matriz quadrada de ordem n. Um escalar λ é chamado de autovalor da matriz A se existir um vetor não nulo v tal que $A \cdot v = \lambda \cdot v$. Tal vetor é chamado de um autovetor da matriz A correspondente ao autovalor λ.*

A expressão da definição acima pode ser reescrita da seguinte forma:

$$A \cdot v = \lambda \cdot v \Leftrightarrow A \cdot v = \lambda \cdot I \cdot v \Leftrightarrow A \cdot v - \lambda \cdot I \cdot v = 0 \Leftrightarrow (A - \lambda I) \cdot v = 0 \Leftrightarrow$$

$$\left(\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix} - \lambda \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix} \right) \cdot \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \Leftrightarrow$$

$$\begin{bmatrix} a_{11} - \lambda & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} - \lambda & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} - \lambda \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \Leftrightarrow$$

$$\begin{cases} (a_{11} - \lambda) \cdot x_1 + a_{12} \cdot x_2 + \cdots + a_{1n} \cdot x_n = 0 \\ a_{21} \cdot x_1 + (a_{22} - \lambda) \cdot x_2 + \cdots + a_{2n} \cdot x_n = 0 \\ \vdots \\ a_{n1} \cdot x_1 + a_{n2} \cdot x_2 + \cdots + a_{nn} \cdot (x_n - \lambda) = 0 \end{cases}$$

Para um vetor v não nulo, o sistema homogêneo admitirá solução além da trivial se $\det(A - \lambda I) = 0$, sendo assim possível determinar os valores de λ. O conjunto de todos os autovetores correspondentes a um autovalor λ de uma matriz quadrada A, acrescido do vetor nulo, é chamado de auto-subespaço de λ, e denotaremos por E_λ.

Teorema 1 (Teorema Espectral para matrizes simétricas)**.** *Seja A uma matriz real e simétrica. Então todos os autovalores de A são reais.*

Demonstração. Para demonstração deste teorema 1, vamos considerar o vetor coluna u de entradas complexas ($\mathbb{C}$) $z_j = a_j + b_{ji}$, para todo $j \in 1, 2, \cdots, n$, $i = \sqrt{-1}$ sendo o nosso elemento imaginário e $a, b \in \mathbb{R}$. Chamamos de conjugado de z, denotado por $\bar{z}$, o complexo $\bar{z} = a - bi$. Com isso, denotaremos a transposta Hermitiana de u o vetor $u^H = \bar{u}^t = [\bar{z_1}, \bar{z_2}, \cdots, \bar{z_n}]$. No espaço $\mathbb{C}^n$ o produto interno canônico é definido como sendo $\langle (z_1, z_2, \cdots, z_n) \rangle \cdot \langle (w_1, w_2, \cdots, w_n) \rangle = z_1\bar{w_1} + z_2\bar{w_2} + \cdots + z_n\bar{w_n}$, com $z_j \in \mathbb{C}$ e $w_j \in \mathbb{C}$ para todo j. Seja a norma de um vetor u, denotado por $\| u \|$, o valor $\| u \| = \sqrt{\langle u, u \rangle}$. Daí, teremos que $\| u \|^2 = (\sqrt{z_1\bar{z_1} + z_2\bar{z_2} + \cdots + z_n\bar{z_n}})^2 = u^H u$.

Seja A uma matriz simétrica, com u e λ, respectivamente, seus autovetor e autovalor, com u de entradas complexas e $\lambda \in \mathbb{C}$. Temos que $Au = \lambda u$, multiplicando a equação à esquerda por u^H, obteremos $u^H Au = u^H \lambda u = \lambda u^H u = \lambda \| u \|^2$. Por outro lado, $u^H Au = (uA)^H u = (\lambda u)^H u = \bar{\lambda} u^H u = \bar{\lambda} \| u \|^2$. Como $u \neq 0$, chegamos que $\lambda = \bar{\lambda}$, Logo $\lambda \in \mathbb{R}$. ■

Logo, se assumirmos que A_n é uma matriz simétrica, então sabemos que ela possui n autovalores, sendo estes reais, e todos os autovetores também são de entradas reais. Na verdade, uma matriz quadrada de ordem n pode ter no máximo n autovalores. Portanto, se A é uma matriz quadrada de ordem n, o sistema homogêneo tem solução não-trivial e pode ter no máximo n autovalores.

Uma consequência importante para o nosso estudo é que qualquer matriz A tem os mesmos autovalores que sua transposta A^t, visto que $\det(A) = \det(A^t)$. Porém, em geral, uma matriz A e sua transposta A^t não possuem os mesmos autovetores que correspondem a autovalores comuns.

Definição 2.9. *Dada uma matriz não negativa* A*, ela é chamada de coluna estocástica se a soma das entradas em cada coluna forem iguais a* 1*.*

Portanto, para uma matriz de coluna estocástica A_n qualquer, temos que $\sum_{i=1}^{n} a_{ij} = 1$, para todo j.

Com isso, podemos afirmar que toda matriz coluna estocástica tem $\lambda = 1$ como autovalor. Para isto, primeiro mostramos que a matriz transposta A^t possui como autovalor $\lambda = 1$, com seu correspondente autovetor. Então, seja A_n uma matriz coluna estocástica, tome o vetor coluna v com todos seus elementos iguais a 1, logo $x_j = 1$, para todo $j \leqslant n$. Note que:

$$\begin{bmatrix} a'_{11} & a'_{12} & \cdots & a'_{1n} \\ a'_{21} & a'_{22} & \cdots & a'_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a'_{n1} & a'_{n2} & \cdots & a'_{nn} \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} = \begin{bmatrix} a'_{11} + a'_{12} + \cdots + a'_{1n} \\ a'_{21} + a'_{22} + \cdots + a'_{2n} \\ \vdots \\ a'_{n1} + a'_{n2} + \cdots + a'_{nn} \end{bmatrix}$$

Sendo A^t a matriz transposta de A, então temos que $a'_{ji} = a_{ij}$, e como A é coluna estocástica, a soma dos elementos de mesma coluna é sempre igual a 1, daí, na matriz A^t a soma dos elementos de mesma linha também será sempre 1, logo $\sum_{i=1}^{n} a'_{ji} = 1$. Assim,

$$\begin{bmatrix} a_{11} + a_{21} + \cdots + a_{n1} \\ a_{12} + a_{22} + \cdots + a_{n2} \\ \vdots \\ a_{1n} + a_{2n} + \cdots + a_{nn} \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} = 1 \cdot \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}.$$

Portanto, temos que $A^t \cdot v = 1 \cdot v$, daí $\lambda = 1$ é auto valor da matriz A^t associado ao vetor coluna $v = [1, 1, \cdots, 1]^t$. Como toda matriz tem o mesmo determinante de sua transposta, temos que 1 também é um autovalor para a matriz A. É importante salientar que o vetor v de entradas todas iguais a 1 não é o autovetor correspondente ao autovalor $\lambda = 1$. Deseja-se encontrar um autovetor que corresponde a este autovalor para A, seja u esse autovetor. A fim de encontrar $u = [x_1, x_2, \cdots, x_n]^t$ explicitamente escrevemos a equação $(A - I) \cdot u = 0$ onde

$$\begin{bmatrix} a_{11} - 1 & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} - 1 & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} - 1 \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \Leftrightarrow$$

$$\begin{cases} (a_{11}-1)\cdot x_1 + a_{12}\cdot x_2 + \cdots + a_{1n}\cdot x_n = 0 \\ a_{21}\cdot x_1 + (a_{22}-1)\cdot x_2 + \cdots + a_{2n}\cdot x_n = 0 \\ \vdots \\ a_{n1}\cdot x_1 + a_{n2}\cdot x_2 + \cdots + a_{nn}\cdot (x_n - 1) = 0 \end{cases} \Leftrightarrow$$

$$\begin{cases} x_1 = a_{11}\cdot x_1 + a_{12}\cdot x_2 + \cdots + a_{1n}\cdot x_n \\ x_2 = a_{21}\cdot x_1 + a_{22}\cdot x_2 + \cdots + a_{2n}\cdot x_n \\ \vdots \\ x_n = a_{n1}\cdot x_1 + a_{n2}\cdot x_2 + \cdots + a_{nn}\cdot x_n \end{cases}.$$

Portanto, o autovetor u que corresponde ao autovalor $\lambda = 1$, terá suas entradas $x_i = \sum_{j=1}^{n} a_{ij}x_j$, para todos $i \leqslant n$ e $j \leqslant n$.

Capítulo 3

Cadeias de Markov

Neste capítulo, abordaremos de maneira formal os conceitos matemáticos de Matriz Coluna Estocástica e Cadeias de Markov, que serão necessários para modelarmos o problema de encontrar uma pontuação de importância para cada uma das páginas da internet. Para os leitores que sentirem necessidade de uma explanação mais detalhada dos conceitos citados acima, além daquelas apresentadas no decorrer do texto, indicamos a leitura de [2] e [3].

> A palavra estocástico é derivada do adjetivo grego *stokhastikos*, que significa "capaz de aproximar" (ou adivinhar). É aplicada em qualquer coisa governada pelas leis da probabilidade, no sentido de que probabilidade faz previsões sobre a chance das coisas acontecerem. Na teoria das probabilidades, os "processos estocásticos" são uma generalização das **cadeias de Markov**. (POOLE, 2004, p. 204)

Chamamos vetor de probabilidade todo vetor com entradas reais não negativas cujo somatório de todas as suas entradas seja igual a 1. Portanto, sendo u um vetor de probabilidade com n entradas reais não negativas, ou seja, $x_i \geqslant 0$, para todo $i \in \mathbb{N}$, então:

$$x_1 + x_2 + \cdots + x_n = \sum_{i=1}^{n} x_i = 1.$$

Uma forma de transformar um vetor v de comprimento n e entradas reais positivos $y_1, y_2, \cdots, y_n$ em um vetor de probabilidade u, é dividir cada uma de suas entradas pelo somatório de suas entradas:

$$x_1 = \frac{y_1}{\sum_{i=1}^{n} y_i}, x_2 = \frac{y_2}{\sum_{i=1}^{n} y_i}, \cdots, x_n = \frac{y_n}{\sum_{i=1}^{n} y_i}.$$

Assim como em vetores, dizemos que uma matriz é "não negativa" se todas suas entradas forem maiores ou iguais a zero. E é dita positiva se todas as entradas forem maiores que zero. Como podemos observar, pela definição 2.9, uma matriz coluna estocástica de tamanho $m \times n$ pode ser vista como o emparelhamento de n vetores probabilidade com m entradas em cada vetor.

Um resultado interessante é que o produto de duas matrizes coluna estocástica, A_n e B_n, também é uma matriz coluna estocástica. Como $a_{ij}, b_{ij} \in [0, 1]$, as entradas da matriz $A \cdot B$ serão não negativas e

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix} \cdot \begin{bmatrix} b_{11} & b_{12} & \cdots & b_{1n} \\ b_{21} & b_{22} & \cdots & b_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ b_{n1} & b_{n2} & \cdots & b_{nn} \end{bmatrix} =$$

$$\begin{bmatrix} \sum_{j=1}^{n} a_{1j}b_{j1} & \sum_{j=1}^{n} a_{1j}b_{j2} & \cdots & \sum_{j=1}^{n} a_{1j}b_{jn} \\ \sum_{j=1}^{n} a_{2j}b_{j1} & \sum_{j=1}^{n} a_{2j}b_{j2} & \cdots & \sum_{j=1}^{n} a_{2j}b_{jn} \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{j=1}^{n} a_{nj}b_{j1} & \sum_{j=1}^{n} a_{nj}b_{j2} & \cdots & \sum_{j=1}^{n} a_{nj}b_{jn} \end{bmatrix}.$$

Observe que a soma de cada uma das colunas é igual a 1:

$$a_{11} \cdot b_{11} + a_{12} \cdot b_{21} + \cdots + a_{1n} \cdot b_{n1} +$$
$$a_{21} \cdot b_{11} + a_{22} \cdot b_{21} + \cdots + a_{2n} \cdot b_{n1} +$$
$$\vdots$$
$$a_{n1} \cdot b_{11} + a_{n2} \cdot b_{21} + \cdots + a_{nn} \cdot b_{n1} =$$

$$b_{11}\underbrace{(a_{11} + a_{21} + \cdots + a_{n1})}_{1} + b_{21}\underbrace{(a_{12} + a_{22} + \cdots + a_{n2})}_{1} + \cdots + b_{n1}\underbrace{(a_{1n} + a_{2n} + \cdots + a_{nn})}_{1} =$$
$$b_{11} + b_{21} + \cdots + b_{n1} = 1.$$

Como consequência, teremos que A^k, como $k \in \mathbb{N}$, também é uma matriz coluna estocástica. Outrossim, como podemos operar vetores coluna como matrizes de ordem $n \times 1$ (no caso do vetor linha $1 \times n$), vale ressaltar que o produto de uma matriz coluna estocástica por um vetor de probabilidade obtemos também um vetor de probabilidade.

Outra observação importante é que se A é uma matriz coluna estocástica, então A^t não é. No entanto, a soma dos elementos de cada linha de A^t é igual a 1, logo $\sum_{j=1}^{n} a'_{ij} = 1$.

> Andrei A. Markov (1856-1922) foi um matemático russo que estudou e posteriormente lecionou na Universidade de São Petersburgo. Era interessado na teoria de números, em análise, e na teoria de frações contínuas, uma área recém-surgida que Markov aplicava na teoria de probabilidade. Markov também tinha interesse em poesia, e um dos usos que ele deu às cadeias de Markov foi a análise de padrões em poemas e outros textos literários. (POOLE, 2004, p. 202)

Um processo aleatório ou estocástico $X_n, n = 0, 1, 2, 3, \cdots$ é uma família de variáveis aleatórias parametrizadas pelo inteiro n. Assumimos que cada uma destas variáveis aleatórias X_n toma seus valores num conjunto finito T. Em outras palavras, é um espaço de amostras em que cada elemento é associado a uma função do tempo. Podemos ver como uma probabilidade condicional $P(A|B)$ que é a probabilidade de que o evento A ocorra, dado que o evento B tenha ocorrido.

Definição 3.1. *Seja* $\{X_n, n = 0, 1, 2, 3, \cdots\}$ *um processo aleatório, tomando seus valores num conjunto* $T = \{A, B, C, \cdots\}$. *Dizemos que* $\{X_n\}$ *é uma cadeia de Markov se a probabilidade* $P(X_n = i), i \in T$, *depender somente do valor do processo no passo anterior,* X_{n-1}, *e não em qualquer dos passos anteriores* $X_{n-2}, X_{n-3}, \cdots$. *Definimos* $n \in \mathbb{N}$ *como o número de elementos em* T.

Dados os vetores v_k e v_{k+1} de mesmo tamanho $n \in \mathbb{N}$ e uma matriz $P = [p_{ij}]_{n \times n}$, tal que $v_{k+1} = Pv_k$, para $k \in \mathbb{N}$, então os vetores v_k e v_{k+1} são chamados de vetores de estado e a matriz P de matriz de transição.

Cadeias de Markov possui um comportamento caracterizado por seu estado inicial e por uma matriz de transição dada por $P(X_n = i|X_{n-1} = j) = p_{ij}$. Ou seja, não tem memória dos estados passados e o estado futuro é determinado completamente pelo estado atual e por uma matriz de transição.

Definição 3.2. *Seja* P *uma matriz quadrada, é dita matriz de transição de uma Cadeia de Markov, se e somente se, estiver definida por* $P(X_n = i|X_{n-1} = j) = p_{ij}$, *com* $p_{ij} \in [0, 1]$ *e* $\sum_{i \in T} p_{ij} = 1$ *para todo* $i, j \in T$.

Observando o comportamento das mudanças sucessivas de estado, temos:

$$\begin{aligned}
\nu_1 &= P\nu_0 \\
\nu_2 &= P\nu_1 \\
&\vdots \\
\nu_k &= P\nu_{k-1}.
\end{aligned}$$

Então, por recorrência temos:

$$\begin{aligned}
\nu_1 &= P\nu_0 \\
\nu_2 &= P(P\nu_0) = P^2\nu_0 \\
\nu_3 &= P(P^2\nu_0) = P^3\nu_0 \\
&\vdots \\
\nu_k &= P^{k-1}\nu = P^k\nu_0.
\end{aligned}$$

Logo, $(p^k)_{ij}$ é a probabilidade de se passar do estado j ao estado i em k transições. Assim a sequência de vetores de probabilidade $\nu_1, \nu_2, \cdots, \nu_n$ juntamente com a matriz de transição P forma uma Cadeia de Markov.

Capítulo 4

Algoritmo *PageRank*

Modernos motores de busca empregam métodos de classificação de resultados para que seja fornecido primeiramente os "melhores", ao invés de simplesmente um *ranking* das páginas que apresentam o texto pesquisado. Um dos algoritmos mais conhecidos e influentes para determinar a relevância das páginas na internet é o algoritmo *PageRank*, usado pelo motor de busca Google. Foi criado e desenvolvido por Larry Page e Sergey Brin, enquanto eles eram estudantes de pós-graduação em Stanford, e tornou-se uma marca Google em 1998. A ideia inicial para o desenvolvimento do *PageRank* foi a de que, a importância de qualquer página da internet pode ser julgada olhando para as páginas que apontam para ela. Quer falemos de popularidade ou autoridade, podemos, de maneira repetida, atribuir uma classificação a cada página, com base nas fileiras das outras páginas que a apontam. É bem semelhante ao conceito popular de moda, se muitas pessoas usam um mesmo estilo de roupa ou mesmo uma determinada celebridade o usa, aquilo vira moda a ser seguida, logo julga-se importante uma loja ter aquela roupa ou acessório em sua vitrine, já que será procurado por outros compradores.

Outra justificativa intuitiva é que uma página de internet pode ter um alto *PageRank* se existem muitas páginas que apontam para ela, ou se algumas páginas que a apontam têm um alto *PageRank*. Intuitivamente, as páginas que são bem citados a partir de muitos lugares ao redor da internet valem a pena olhar. Além disso, as páginas que têm, talvez, apenas uma citação de algo como a *homepage* do site da CAPES! geralmente vale a pena olhar. Se uma página não for de alta qualidade, ou era um *link* quebrado, é bastante provável que a página inicial da CAPES não iria apontar para ele. O algoritmo *PageRank* lida com esses dois casos e tudo mais por recursividade através da estrutura de *links* da

própria internet.

A utilidade de um motor de busca depende da relevância do conjunto de resultados que é retornado. Não pode ser, obviamente, milhões de páginas de internet que incluem uma determinada palavra ou frase, no entanto algumas dessas páginas serão mais relevantes, ou popular, ou com maior autoridade do que outras. Um usuário não tem a habilidade, ou paciência, para fazer a varredura através de todas as páginas que contêm as palavras consultadas. Uma pessoa ao fazer uma pesquisa espera que as páginas relevantes sejam exibidas dentre as primeiras devolvido pelo motor de busca.

4.1 A Matemática do Google

O motor de busca Google tem características importantes que o ajudam a produzir resultados de alta precisão. Ele faz uso da estrutura de *links* já existente na internet para calcular um *ranking* de qualidade para cada página da web. Esta classificação é chamado *PageRank*, que é o objeto de estudo deste trabalho e o descrevemos na sessão seguinte.

4.1.1 Descrição do Cálculo PageRank

O modelo de citação da literatura acadêmica tem sido aplicado à web, em grande parte, pela contagem de citações, ou *backlinks*, para uma determinada página. Isto dá alguma aproximação de importância ou qualidade de uma página. O algoritmo do *PageRank* estende esta ideia, por não contar os *links* em todas as páginas igualmente, e por normalizar pelo número de *links* em uma página.

Assumimos que uma página A tenha $T_1, T_2, \cdots, T_n$ páginas com *links* que apontam para ela, ou seja, são citações, *backlinks*. Tomaremos o parâmetro $d \in \mathbb{R}$, como sendo um fator de amortecimento que pode ser ajustado no intervalo entre 0 e 1. Há mais detalhes sobre d na próxima seção. Definiremos $C(A)$ como o número de *links* que a página A possui apontando para outras páginas. Seja $PR(A)$, o *PageRank* de uma página A, que é dado por:

$$PR(A) = (1-d) + d\left(\frac{PR(T_1)}{C(T_1)} + \frac{PR(T_2)}{C(T_2)} + \cdots + \frac{PR(T_n)}{C(T_n)}\right).$$

Note-se que o *PageRank* formam uma distribuição de probabilidade sobre as páginas da web, de modo que a soma de todos os *PageRanks* das páginas da web também será

um deles.

PageRank ou PR(A) pode ser calculada usando um algoritmo iterativo simples, e corresponde ao principal autovetor da matriz de ligação normalizada da web. Além disso, um PageRank de 26 milhões de páginas da web pode ser calculado em algumas horas com uma estação de trabalho de tamanho médio.

O *PageRank* pode ser pensado como um modelo de comportamento para os usuários da internet. Assumindo que um usuário resolve sair clicando de forma aleatória nos *links* das páginas na web e continua a clicar nos *links*, mas eventualmente fica entediado e começa em outra página aleatória. A probabilidade de que este usuário visite uma determinada página é o que chamamos de *PageRank*. E o fator de amortecimento d é a probabilidade do usuário sair da página atual e acessar outra sem usar link.

Para este objetivo, começamos retratando a rede Web como um grafo direcionado, com nós representados por páginas web e bordas representadas pelas ligações entre eles. Suponha, por exemplo, que temos uma pequena internet que consiste em apenas 4 sites, www.pagina1.com, www.pagina2.com, www.pagina3.com e www.pagina4.com, referenciando um ao outro da maneira sugerida pela figura 4.1:

Figura 4.1: uma pequena internet

Nós podemos "traduzir"a imagem em um grafo direcionado com 4 nós, um para cada página. Quando um site i referencia um site j, nós adicionamos uma aresta dirigida entre o nó i e o nó j no gráfico. Para efeitos de computação seu *PageRank*, ignoramos

quaisquer *links* de navegação, tais como os botões próximo, avançar e outros, já que só nos preocupamos com as conexões entre diferentes sites da web. Por exemplo, como exitem ligações da pagina1 para todas as outras páginas, então o nó 1 no gráfico terá aresta de saída para todos os outros nós. Como a pagina3 tem apenas um *link*, para a pagina1, portanto o nó 3 terá uma aresta de saída para o nó 1. Depois de analisar cada página da web, termos o gráfico a seguir:

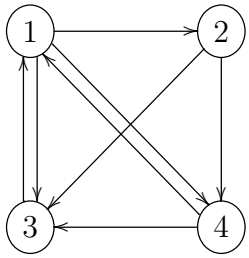
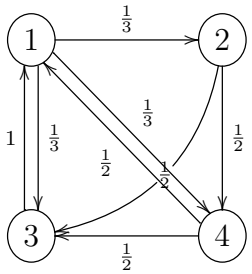


Figura 4.2: Grafo representando nossa pequena internet

Em nosso modelo, cada página deve transferir uniformemente a sua importância para as páginas relacionadas a ela. O nó 1 possui 3 *links* que apontam para as páginas $2, 3$ e 4, de modo que vai passar $\dfrac{1}{3}$ da sua importância para cada um dos outros 3 nós. O nó 3 tem apenas uma extremidade de saída, por isso vai passar toda a sua importância para o nó 1. Em geral, se um nó tem k arestas de saída, ele vai passar $\frac{1}{\mathsf{k}}$ de sua importância para cada um de nós que as vincula. Vamos visualizar melhor o processo atribuindo pesos a cada aresta.



Onde podemos denotar o gráfico pela matriz de transição $\mathsf{A} = [\mathsf{a}_{\mathsf{ij}}]_{4\times4}$, onde cada entrada a_{ij} da matriz representa a probabilidade de estando em i ir para j.

$$\mathsf{A} = \begin{bmatrix} 0 & 0 & 1 & \frac{1}{2} \\ \frac{1}{3} & 0 & 0 & 0 \\ \frac{1}{3} & \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{3} & \frac{1}{2} & 0 & 0 \end{bmatrix}$$

Suponha-se que inicialmente a importância é distribuída uniformemente entre os 4 nós, obtendo $\frac{1}{4}$ para cada. Denote por ν o vetor do *ranking* inicial, tendo todas as entradas iguais a $\frac{1}{4}$. Cada *link* de entrada aumenta a importância de uma página web, então no passo 1, nós atualizamos a classificação de cada página adicionando ao valor atual a importância das ligações recebidas. Isto é o mesmo que multiplicarmos a matriz A com ν. Após o passo 1, o novo vetor de importância será $\nu_1 = A\nu$. Podemos repetir o processo, assim, no passo 2, o vetor de importância é atualizado $\nu_2 = A(A\nu) = A^2\nu$. Assim sucessivamente até um passo k :

$$\nu = \begin{bmatrix} 0,25 \\ 0,25 \\ 0,25 \\ 0,25 \end{bmatrix}, A\nu = \begin{bmatrix} 0,375 \\ 0,08\bar{3} \\ 0,\bar{3} \\ 0,208\bar{3} \end{bmatrix}, A^2\nu = \begin{bmatrix} 0,4375 \\ 0,125 \\ 0,2708\bar{3} \\ 0,1\bar{6} \end{bmatrix}, A^3\nu = \begin{bmatrix} 0,3541\bar{6} \\ 0,1458\bar{3} \\ 0,291\bar{6} \\ 0,208\bar{3} \end{bmatrix},$$

$$A^4\nu = \begin{bmatrix} 0,3958\bar{3} \\ 0,1180\bar{5} \\ 0,29513\bar{8} \\ 0,19097\bar{2} \end{bmatrix}, A^5\nu = \begin{bmatrix} 0,390625 \\ 0,1319\bar{4} \\ 0,286458\bar{3} \\ 0,19097\bar{2} \end{bmatrix}, A^6\nu = \begin{bmatrix} 0,3819\bar{4} \\ 0,130208\bar{3} \\ 0,291\bar{6} \\ 0,196180\bar{5} \end{bmatrix},$$

$$A^7\nu = \begin{bmatrix} 0,3897569\bar{4} \\ 0,12731\bar{4}8 \\ 0,290509\bar{2}5 \\ 0,192418981\bar{4} \end{bmatrix}, A^8\nu = \begin{bmatrix} 0,38671875 \\ 0,1299189\bar{1}4 \\ 0,28978587\bar{9}62 \\ 0,1935763\bar{8} \end{bmatrix}, A^9\nu = \begin{bmatrix} 0,3865\bar{7}40 \\ 0,12890625 \\ 0,290653935\bar{1}8 \\ 0,193865\bar{7}40 \end{bmatrix},$$

$$A^{10}\nu = \begin{bmatrix} 0,387586805 \\ 0,12885802\bar{4}691 \\ 0,290244020061728 \\ 0,193311149691358 \end{bmatrix}, \cdots$$

Notamos que a sequência de interações $\nu, A\nu, \cdots, A^k\nu$ tende ao valor de equilíbrio em que, aproximadamente:

$$\nu^{n-1} = \nu^n = \begin{bmatrix} 0,3870 \\ 0,1290 \\ 0,2903 \\ 0,1935 \end{bmatrix}.$$

Este vetor de equilíbrio ν^n é o vetor *PageRank* do nosso gráfico web.

4.1.2 Do ponto de vista da Álgebra Linear

Observando as interações descritas na sessão 4.1.1 onde tendem a um valor de equilíbrio, tal que

$$\nu^{n-1} = \nu^n \Rightarrow A \cdot \nu^n = \nu^n.$$

Denotemos por $\nu = [x_1, x_2, x_3, x_4]^t$ o vetor de equilíbrio que representa a importância das quatro páginas. Analisando a situação em cada nó temos:

$$\begin{bmatrix} 0 & 0 & 1 & \frac{1}{2} \\ \frac{1}{3} & 0 & 0 & 0 \\ \frac{1}{3} & \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{3} & \frac{1}{2} & 0 & 0 \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} \Leftrightarrow$$

$$\begin{cases} x_1 = 1 \cdot x_3 + \frac{1}{2} \cdot x_4 \\ x_2 = \frac{1}{3} \cdot x_1 \\ x_3 = \frac{1}{3} \cdot x_1 + \frac{1}{2} \cdot x_2 + \frac{1}{2} \cdot x_4 \\ x_4 = \frac{1}{3} \cdot x_1 + \frac{1}{2} \cdot x_2 \end{cases}$$

Fazendo $x_1 = t$, com $t \in \mathbb{R}$, encontramos o sistema equivalente:

$$\begin{cases} x_1 = t \\ x_2 = \frac{1}{3} \cdot t \\ x_3 = \frac{3}{4} \cdot t \\ x_4 = \frac{1}{2} \cdot t \end{cases}$$

Como as entradas do vetor formam uma distribuição de probabilidade, ou seja, a soma de todas suas entradas é igual a 1, logo $x_1 + x_2 + x_3 + x_4 = 1 \Rightarrow t + \frac{1}{3} \cdot t + \frac{3}{4} \cdot t + \frac{1}{2} \cdot t = 1 \Rightarrow t = \frac{12}{31}$. Daí temos, com aproximação por omissão das demais casas decimais seguintes: $x_1 = 0,3870; x_2 = 0,1290; x_3 = 0,2903; x_4 = 0,1935$ os valores do *PageRank* de cada uma das páginas.

Por outro lado, podemos observar que o vetor ν é o autovetor associado ao autovalor 1 da matriz A

$$\begin{bmatrix} 0 & 0 & 1 & \frac{1}{2} \\ \frac{1}{3} & 0 & 0 & 0 \\ \frac{1}{3} & \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{3} & \frac{1}{2} & 0 & 0 \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = 1 \cdot \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix}.$$

4.1.3 Ponto de vista probabilístico

Uma vez que a importância de uma página web é medida por sua popularidade, ou seja, quantas ligações a apontam, podemos ver a importância de uma página i como a probabilidade de que um usuário de forma aleatória a visite, sendo que o mesmo, ao navegar pela internet, abra qualquer página e comece a clicar pelos *hiperlinks* até chegar na página i. Interpretaremos os pesos atribuídos aos nós e às arestas direcionadas do gráfico de uma forma probabilística: Um usuário que está atualmente a visualizar a pagina2, tem probabilidade $\frac{1}{2}$ de ir para pagina3, e probabilidade $\frac{1}{2}$ de ir para a pagina4. É possível modelar o processo como um passeio aleatório em gráfico. Para a nossa "pequena internet", cada página tem igual probabilidade de $\frac{1}{4}$ de ser escolhido como ponto de partida. Assim, a distribuição de probabilidade inicial é dada pelo vetor coluna $v = \left[\frac{1}{4}\frac{1}{4}\frac{1}{4}\frac{1}{4}\right]^t$. A probabilidade de que a página i seja visitada dado o primeiro clique, que chamaremos de passo, é igual a Av, após o segundo clique, ou passo, é de $A(Av) = A^2v$, e assim por diante. Daí a probabilidade de que a página i seja visitada depois de $k \in \mathbb{N}$ passos é igual a A^kv. A sequência $Av, A^2v, A^3v, \cdots, A^kv, \cdots$ converge, neste caso, a um vetor estacionário v^*, que é único. Neste contexto v^* será o nosso vetor *PageRank*. Além disso, a i-ésima entrada no vetor v^* é simplesmente a probabilidade de que a cada momento i uma página receba visita do nosso usuário aleatório. Os cálculos são idênticos aos que fizemos na sessão 4.1.1, apenas o significado que atribuímos a cada passo é ligeiramente diferente.

4.1.4 Grafos desconectados

É possível calcular por métodos diferentes o vetor *PageRank* v^*, no qual indica que a pagina1 é a mais relevante. Isso pode parecer surpreendente, pois a pagina1 possui 2 *backlinks*, enquanto a pagina3 tem 3 *backlinks*. Se dermos uma olhada no gráfico da figura 4.2, veremos que o nó 3 tem apenas uma extremidade de saída para o nó 1, por isso transfere toda a sua importância para o nó 1. De forma equivalente, uma vez que um internauta que segue clicando aleatoriamente nos *links* presentes em cada página, ao visitar pagina3, ele só poderá ir para a pagina1. Observe também, como a classificação de cada página não é somente feita com a soma ponderada das arestas que entram no nó. Intuitivamente, no instante 1, um nó recebe voto de importância de seus vizinhos diretos, no passo 2, recebe dos vizinhos de seus vizinhos, e assim sucessivamente.

A web é muito heterogêneo, pela sua natureza, e, certamente, enorme, por isso não esperamos que o seu gráfico seja todo conectado. Da mesma forma, haverá páginas que são simples descritivo e não contêm *links* de saída. O que deve ser feito nesse caso? Precisamos de um sentido não ambíguo da classificação de uma página, para qualquer gráfico web dirigido com n nós.

Afim de superar esse problema, é fixada uma constante real positiva d entre 0 e 1, que foi chamada de fator de amortecimento (um valor típico para d é de $0,15$). Daí, definimos a matriz de *PageRank* (também conhecida como a matriz Google) do gráfico por $M = (1-d)\cdot A + d\cdot B$ onde:

$$B = \frac{1}{n}\begin{bmatrix} 1 & 1 & \cdots & 1 \\ 1 & 1 & \cdots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \cdots & 1 \end{bmatrix}.$$

É importante notar que M continua sendo uma matriz coluna estocástica com entradas positivas. Veja que:

$$M = (1-d)\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix} + d\begin{bmatrix} \frac{1}{n} & \frac{1}{n} & \cdots & \frac{1}{n} \\ \frac{1}{n} & \frac{1}{n} & \cdots & \frac{1}{n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{1}{n} & \frac{1}{n} & \cdots & \frac{1}{n} \end{bmatrix} \Rightarrow$$

$$M = \begin{bmatrix} (1-d)a_{11} & (1-d)a_{12} & \cdots & (1-d)a_{1n} \\ (1-d)a_{21} & (1-d)a_{22} & \cdots & (1-d)a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ (1-d)a_{n1} & (1-d)a_{n2} & \cdots & (1-d)a_{nn} \end{bmatrix} + \begin{bmatrix} \frac{d}{n} & \frac{d}{n} & \cdots & \frac{d}{n} \\ \frac{d}{n} & \frac{d}{n} & \cdots & \frac{d}{n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{d}{n} & \frac{d}{n} & \cdots & \frac{d}{n} \end{bmatrix} \Rightarrow$$

$$M = \begin{bmatrix} (1-d)a_{11} + \frac{d}{n} & (1-d)a_{12} + \frac{d}{n} & \cdots & (1-d)a_{1n} + \frac{d}{n} \\ (1-d)a_{21} + \frac{d}{n} & (1-d)a_{22} + \frac{d}{n} & \cdots & (1-d)a_{2n} + \frac{d}{n} \\ \vdots & \vdots & \ddots & \vdots \\ (1-d)a_{n1} + \frac{d}{n} & (1-d)a_{n2} + \frac{d}{n} & \cdots & (1-d)a_{nn} + \frac{d}{n} \end{bmatrix}.$$

Observe que a soma dos elementos de qualquer uma das colunas i da matriz M é igual a 1 e suas entradas são sempre positivo:

$$(1-d)a_{1i} + \frac{d}{n} + (1-d)a_{2i} + \frac{d}{n} + \cdots + (1-d)a_{ni} + \frac{d}{n} =$$
$$(1-d)(a_{1i} + a_{2i} + \cdots + a_{ni}) + n\frac{d}{n} = (1-d) + d = 1.$$

A matriz M modela o nosso usuário aleatório da seguinte forma: na maioria das vezes, ele vai seguir os *links* de uma página, a partir de uma página i, e passar para um dos vizinhos de i. Uma porcentagem menor, mas positiva do tempo, o usuário vai sair da página atual e escolher arbitrariamente uma outra diferente a partir da *web*, digitando seu endereço e "teletransporta-se" lá. O fator de amortecimento d reflete a probabilidade de o usuário fechar a página atual e ir para uma nova sem utilizar *links*. Desde que seja possível efetuar esse "teletransporte" para qualquer página da internet, cada uma tem probabilidade de ser escolhida. Isto justifica a estrutura da matriz $\frac{1}{n}B\cdot$

Intuitivamente, a matriz M faz a conexão do gráfico e se livra do problema com nós desconectados. Um nó que não possui saída tem probabilidade $\frac{p}{n}$ de se mover para qualquer outro nó. Rigorosamente, para a matriz M, aplicam-se as seguintes teoremas:

Teorema 2 (Perron-Frobenius)**.** *Se M é uma matriz coluna estocástica positiva, então:*

I. 1 *é um Autovalor de multiplicidade única.*

II. todos os outros autovalores têm um valor absoluto menor do que 1*, logo o maior Autovalor será* 1*.*

III. os autovetores correspondentes ao autovalor 1 *tem apenas entradas positivas ou apenas entradas negativas. Em particular, para o autovalor* 1 *existe um único autovetor com a soma das suas entradas iguais a* 1*.*

Teorema 3. *Seja M_n uma matriz coluna estocástica positiva. Denote v^* sendo seu autovetor correspondente ao autovalor* 1*. Seja v o vetor coluna com todas entradas iguais a $\frac{1}{n}$. Então, a sequência $v, Mv, \cdots, M^k v$ converge para o vetor v^*.*

Com isso, podemos concluir que, o vetor *PageRank* para um gráfico da internet com matriz de transição A, e fator de amortecimento d, é o Autovetor probabilístico original da matriz M, correspondente ao Autovalor 1.

Do ponto de vista matemático, uma vez que temos M, encontrar os Autovetores correspondentes ao Autovalor 1 é, pelo menos em teoria, uma tarefa simples. Consiste

apenas em resolver o sistema $Ax = x$. Porém quando a matriz M tem tamanho 30 bilhões, como no caso da matriz para o gráfico real da Web, mesmo um *software* matemático, como *Matlab* ou *Mathematica* são claramente sobrecarregado. Uma forma alternativa de calcular o Autovetor probabilístico correspondente ao Autovalor 1 é pelo método da potência. O teorema 3 garante que funciona para matrizes coluna estocástica positivas. Em termos computacionais, é muito mais fácil, a partir do vetor com todas as entradas 1, para multiplicar $x, Mx, ..., M^n x$ até a convergência, do que é para calcular os Autovetores de M. Na verdade, neste caso, a pessoa precisa apenas calcular os primeiros pares de iterações a fim de obter uma boa aproximação do vetor *PageRank*. Para uma matriz aleatória, o método de interação é, em geral, conhecido por convergir lentamente. O que torna o trabalho rápido neste caso é o fato de que o gráfico web é escasso, ou seja, que em um nó será pequeno número de ligações de saída, na melhor das hipóteses, duas centenas, que é extremamente pequeno quando comparado aos 30 milhões de nós que poderia, teoricamente, apontar. Daí a matriz de transição A tem um monte de entradas nulas, o que deixa o cálculo bem mais rápido.

Capítulo 5

Considerações Finais

O desenvolvimento da *World Wide Web* aponta para todo um novo campo da matemática chamado Matemática da Internet. É uma mistura de probabilidade, álgebra linear, teoria dos grafos e sistemas dinâmicos, concebidos para responder a problemas rigorosos sobre como a informação se propaga através da rede.

Vimos como o Google usa o *PageRank* para determinar a relevância das páginas da web, tornando uma busca na Internet simples e eficiente. Da perspectiva de um *web designer*, é importante não só para criar um site agradável, com efeitos gráficos interessantes, mas é importante que outras páginas apontem para o seu web site. Um bom *PageRank* pode transformar seu negócio algo muito rentável. Do ponto de vista do Google, é importante manter os cálculos *PageRank* exatos, e evitar tentativas de fraude. *Link* de spam é definido como a tentativa intencional para melhorar o ranking de uma página no motor de busca, aumentando o número de páginas que apontam a ela. As fazendas de ligações envolvem a criação de comunidades de páginas da web referenciando um a outra, conhecidas como "sociedades de admiração mútuas". Embora algumas fazendas de ligações possam ser criadas manualmente, a maioria delas são implementadas através de programas e serviços automatizados. É importante, portanto, reconhecer subgrafos densos do gráfico web projetado para fazer exatamente isso, e para eliminá-los a partir de cálculos *PageRank*. Embora em seu núcleo, o *PageRank* seja como descrito neste trabalho, no mundo real, o Google usa algoritmos melhorados de busca para lidar com estes problemas, podendo atribuir pesos diferentes para os *links* de saída, ou decidir que determinadas ligações não devem transferir seu *PageRank* em tudo, isto para citar apenas alguns dos recursos extras. Atualmente, os algoritmos do Google utilizam mais de *200*

sinais ou "pistas" diferentes para adivinhar o que você realmente procura. Esses sinais incluem coisas como os termos em *websites*, a atualização do conteúdo, a região do usuário e o *PageRank*.

A *World Wide Web* está em constante mudança, as páginas são adicionados e excluídos a cada momento, a compreensão da estrutura do gráfico Internet é um problema-chave de pesquisa. Desenvolvimento de modelos para um gráfico de Internet que tem bilhões de nós, é, no entanto, longe de ser trivial. A web é muitas vezes descrito como tendo uma estrutura de gravata borboleta. O nó da gravata borboleta é representado por um componente fortemente ligado do gráfico, chamado de núcleo. Um lado da gravata borboleta consiste em páginas que apontam para o núcleo, enquanto o outro lado da gravata contém páginas que com pelo menos uma ligação a partir do núcleo. O restante das páginas da web são, ou em componentes desconectados, ou em gavinhas que só apontam para páginas de fora do núcleo. Nesta teia de aranha, redes sociais, comunidades na web são definidos como subgráficos com *links* mais internos, em seguida, os externos. Métodos espectrais são usados para analisar a distribuição destas comunidades.

Compreender o gráfico da Internet pode ajudar a responder perguntas sobre como a informação se propaga através da rede. Propagação do vírus de computador sobre a rede é de particular interesse. Compreender a taxa em que os vírus de computador podem ajudar a propagar os fabricantes de software antivírus para antecipar e áreas de quarentena infectado.

Nesse contexto, a Álgebra Linear é importante aliada na construção de modelos matemáticos lineares para auxiliar no desenvolvimento da informática, porém, sua aplicação se estende em outras áreas como aviação, economia, circuitos eletrônicos ou exploração petrolífera. Alguns conceitos, como matrizes, determinantes e sistemas lineares são trabalhados durante o ensino médio e que por muitas vezes somos questionados por aplicações na qual usaremos esse conhecimento.

Dessa forma apresentamos aqui uma breve introdução no estudo da Álgebra Linear, utilizando o algorítimo do motor de busca do Google como motivação, com o intuito de instigar a curiosidade sobre o tema, visando não um estudo completo e definitivo sobre o mesmo, mas sim como forma de incentivar uma busca mais aprofundada de suas características e propriedades.

Referências Bibliográficas

[1] **HEFEZ, A., FERNANDEZ, C. S.** - *Introdução à Álgebra Linear*. Coleção PROFMAT - Rio de Janeiro: SBM. 2012.

[2] **ROUSSEU, C., SAINT-AUBIN, Y.** - *Matemática e Atualidades, Volume 1*. Coleção PROFMAT - 1 ed. Rio de Janeiro: SBM. 2015.

[3] **POOLE, D**. - *Álgebra Linear*. Tradução de Mertha Salerno Monteiro, et al. São Paulo: Pioneira Thomson Learning, 2004.

[4] **LIMA, E. L.** - *Análise Real, Volume 2, Funções de n Variáveis*. 6 ed. Rio de Janeiro: IMPA. 2013.

[5] **HANING, C.S.** - *Aplicação da Topologia à Análise*. Projeto Euclides - IMPA, 1976.

[6] **IEZZE, G.** - *Fundamentos da Matemática Elementar: Sequências, Matrizes, Determinantes e Sistemas, Volume 4*. 2 ed. Atual, 1977.

[7] **BRIAN, S., Page, L.** - *The Anatomy of a Large-Scale Hypertextual Web Search Engine*. Computer Science Department, Stanford University, Stanford, CA 94305. Disponível em:<`http://infolab.stanford.edu/~backrub/google.html`> Acesso em: 07 de junho de 2016.

[8] **INTERNET LIVE STATS**. Disponível em <`http://www.internetlivestats.com/`>. Acesso em: 07 de junho de 2016.

[9] **PAGE, L., BRIN, S., MOTWANI, R., WINOGRAD, T.** - *The PageRank Citation Ranking: Bringing Order to the Web. Technical Report*. Stanford InfoLab, 1999. Disponível em: <`http://ilpubs.stanford.edu:8090/422/1/1999-66.pdf`>. Acesso em: 20 de junho de 2016.

[10] **BARROS, C. D. V.** - *O Teorema do Ponto Fixo de Banach e Algumas Aplicações*