

**UNIVERSIDADE FEDERAL DE SANTA CATARINA
BIBLIOTECA UNIVERSITÁRIA**

João Carlos Bez Batti

**UM POUCO DA MATEMÁTICA POR TRÁS DO
ALGORITMO PAGERANK DO GOOGLE**

Florianópolis

2015

João Carlos Bez Batti

**UM POUCO DA MATEMÁTICA POR TRÁS DO
ALGORITMO PAGERANK DO GOOGLE**

Dissertação submetida ao Programa
de Mestrado Profissional em Matemática
da Universidade Federal de Santa Ca-
tarina para a obtenção do Grau de
Mestre em Matemática.

Orientador: Prof. Fernando de La-
cerda Mortari

Florianópolis

2015

Ficha de identificação da obra elaborada pelo autor,
através do Programa de Geração Automática da Biblioteca Universitária da UFSC.

Bez Batti, João Carlos

Um pouco da matemática usada no algoritmo PageRank do Google / João Carlos Bez Batti ; orientador, Fernando de Lacerda Mortari - Florianópolis, SC, 2015.
63 p.

Dissertação (mestrado profissional) - Universidade Federal de Santa Catarina, Centro de Ciências Físicas e Matemáticas. Programa de Pós-Graduação em Matemática.

Inclui referências

1. Matemática. 2. Álgebra Linear. 3. Algoritmo PageRank. I. de Lacerda Mortari, Fernando. II. Universidade Federal de Santa Catarina. Programa de Pós-Graduação em Matemática. III. Título.

João Carlos Bez Batti

**UM POUCO DA MATEMÁTICA POR TRÁS DO
ALGORITMO PAGERANK DO GOOGLE**

Esta Dissertação foi julgada aprovada para a obtenção do Título de “Mestre em Matemática”, e aprovada em sua forma final pelo Programa de Mestrado Profissional em Matemática da Universidade Federal de Santa Catarina.

Florianópolis, 16 de Julho 2015.

Profa. Dra. Maria Inez Cardoso Gonçalves
Coordenador
Universidade Federal de Santa Catarina

Banca Examinadora:

Prof. Fernando de Lacerda Mortari
Orientador

AGRADECIMENTOS

Agradeço a Deus por ter me dado saúde e disposição, mas principalmente, pelas pessoas mais do que especiais que colocou em minha vida.

Agradeço aos meus pais, Silvio Bez Batti Neto e Mariléti Letícia Godinho Bez Batti, que sempre me incentivaram e me apoiaram em cada etapa da minha vida, estimulando e dando força para que seguisse em frente.

A minha esposa Mariana Marcusso Manhães e ao meu filho Bernardo Manhães Bez Batti, agradeço por todo amor, carinho, paciência e compreensão durante todo esse período de estudo, que por várias vezes me ausentei e não pude dar a atenção e carinho que mereciam.

Ao meu amigo e colega de mestrado Jaison Gasperi, pelo incentivo e força nos vários dias que passamos estudando.

Aos meus sócios, Lucas Ramiro Talarico e Ricardo Silveira da Silva, pela compreensão em relação à minha ausência em diversas situações.

Ao amigo e professor, Rafael Descovi Galelli, pelas aulas e pela sua imensurável boa vontade.

A todos os meus amigos que de alguma forma me ajudaram nessa caminhada. Em especial, a Fabiano Carlos Cidral pelo incentivo, lealdade e companheirismo, além de todos os ensinamentos.

Ao meu orientador, Professor Fernando de Lacerda Mortari, por todos os conhecimentos adquiridos durante essa caminhada. Agradeço também pela sua paciência em sanar todas as minhas dúvidas e a sua compreensão em relação ao tempo restrito que tive para dedicar-me a esse trabalho.

A Coordenação de Aperfeiçoamento de Pessoal de Nível Superior pelo apoio financeiro.

*“O bom professor é aquele que se torna
gradativamente desnecessário.”
Autor desconhecido*

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. Por fim, apresentaremos um roteiro para se inserir, de modo intuitivo, a essência do algoritmo PageRank no Ensino Médio, estimulando os estudantes ao aprendizado de temas como Matrizes, Determinantes e Sistemas de Equações Lineares usando como objeto a internet, a qual está cada vez mais presente no dia a dia dos alunos e dessa forma servindo como fonte de inspiração para o aprendizado da matemática.

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

ABSTRACT

In this work we deal, in a simplified way, with the mathematical concepts involved in one of the algorithms used in Google's page ranking system called: PageRank. We discuss informally the idea of assigning an importance score for websites and then do a mathematical formalization, with the help of Linear Algebra. Finally, we present a class script to insert, intuitively, the idea of the PageRank algorithm in high school, encouraging students to learn themes such as matrices, determinants and systems of linear equations using the internet as the subject, which is becoming more present in the day-to-day lives of students and thus serving as a source of inspiration for the learning of Mathematics.

Keywords: Internet. Google. PageRank. Linear Algebra.

LISTA DE FIGURAS

Figura 1	Rastreadores	23
Figura 2	Janela do Google	23
Figura 3	Modelo de grafo dirigido.....	24
Figura 4	Modelo de internet de quatro páginas.....	26
Figura 5	Modelo de sumidouro.....	30
Figura 6	Grafo dirigido.....	31
Figura 7	Modelo de internet desconexa.....	36
Figura 8	Grafo	49

SUMÁRIO

INTRODUÇÃO	19
1 A IDEIA	21
1.1 O PAGERANK	24
2 MODELANDO O PROBLEMA	29
2.1 A MATRIZ DO GRAFO	29
2.2 MATRIZES COLUNA-ESTOCÁSTICAS E OS AUTO- VALORES	33
2.3 DEFICIÊNCIAS DO MÉTODO	35
2.3.1 INTERNET DESCONECTADA	35
2.3.2 CORRIGINDO AS DEFICIÊNCIAS	48
2.4 APLICANDO O MÉTODO	50
3 ROTEIRO PARA ABORDAR O TEMA NO EN- SINO MÉDIO	55
3.1 ROTEIRO	55
CONCLUSÃO	59
REFERÊNCIAS	61

INTRODUÇÃO

“Google”, trocadilho feito com o termo matemático “Googol”: número representado pelo dígito um seguido de cem dígitos zeros, deu origem ao nome de uma das maiores e mais evidentes empresas da atualidade: a Google. A utilização da expressão traduz a vontade de Larry Page e Sergey Brin, os fundadores do Google, de organizar uma volume muito grande de dados na internet.

Tudo começou em meados da década de 90, quando faziam parte do programa de doutorado da Universidade de Stanford e começaram juntos um trabalho de pesquisa, com o intuito de criar um sistema de busca mais eficaz do que os existiam na época. Esse sistema foi criado e foi nomeado BackRub. Diferentemente dos sistemas que já existiam, os quais apenas procuravam registros em um banco de dados pré-cadastrados, o BackRub, através de algoritmos, contava os links que se dirigiam às páginas e os interpretava como votos para essas páginas. Dessa forma, quanto mais links um site recebia, mais relevante esse site se tornava nos resultados de uma pesquisa.

Posteriormente, os algoritmos do BackRub foram testados para organizar e classificar milhões de sites públicos. Há boatos de que esse teste tenha sobrecarregado os servidores de Stanford deixando toda a Universidade sem acesso à internet.

Em 1997, vivenciando o êxito do BackRub, numa reunião com diversos estudantes, Larry Page encontrou um nome que representasse a grande quantidade de dados que seu sistema conseguia processar. Contam as referências que um dos estudantes mencionou oralmente a palavra “Googolplex”, e Larry Page prontamente complementou de forma mais sucinta: “Googol”. Sem intenção, cometeram um erro ortográfico ao digitar o termo enquanto procuravam pela disponibilidade de seu domínio - escreveram “Google” em vez de “Googol- obtendo o domínio disponível. Em seguida o domínio foi registrado, nascendo assim, o “Google Search Engine”, a versão pública do BackRub.

Maiores informações sobre a história do Google, como surgimento, evoluções e desenvolvimentos de produtos, o leitor pode seguir em [3]. Para um conhecimento mais técnico sobre a elaboração do algoritmo PageRank sugerimos a leitura de [6].

No Capítulo 1 deste trabalho, apresentaremos as principais estratégias e técnicas atualmente utilizadas pelo Google, de maneira informal, com o simples intuito de transmitir ao leitor uma ideia geral

sobre como funciona esse motor de busca. No entanto, o objetivo central desse trabalho não é explicar o funcionamento do Google nos dias de hoje, e sim apresentar os conceitos matemáticos envolvidos na ideia inicial de Larry Page e Sergey Brin para o desenvolvimento do algoritmo PageRank, o qual fez com que o Google se tornasse um dos mais eficazes buscadores da internet. Tais conceitos serão apresentados formalmente no Capítulo 2, dando portanto, embasamento teórico aos processos descritos no Capítulo 1. Por fim, no Capítulo 3, daremos uma sugestão de como abordar as ideias principais deste trabalho no Ensino médio. Faremos isso através de um roteiro de aula.

1 A IDEIA

As evoluções científicas e tecnológicas advindas do avanço da informática, mais precisamente da internet, são imensuráveis. A facilidade de encontrar uma informação com o simples clicar de um botão tornou muito ágil e prático a construção e a disseminação do conhecimento. Porém, nem sempre foi tão fácil assim. A internet, atualmente, é composta por trilhões de páginas, na qual, ao efetuarmos uma pesquisa procuramos por informações de forma rápida e precisa. As ferramentas que facilitaram tais pesquisas na internet foram os motores de busca, dos quais um deles se destaca: o Google.

Para começarmos a entender os mecanismos de busca do Google, comparemos a internet com uma biblioteca gigantesca, de forma que usar o Google para procurar uma informação na internet é equivalente a procurar um livro que se encontra em alguma das muitas estantes dessa biblioteca. Mais especificamente, encontrar alguma informação contida em alguma parte de uma página de um site compara-se a encontrar um tema discutido em alguma parte de um livro.

Com o auxílio da informática, a busca por um livro numa biblioteca é relativamente fácil. Utilizando um computador, digita-se o título, o autor, ou até mesmo o tema abordado pelo livro e o computador informa um número correspondente a posição do livro dentro da biblioteca. Porém, para que isso aconteça, as informações como tema, título, autor e etc, de todos os livros, devem ser cadastrados manualmente por alguém, o que é uma tarefa muito trabalhosa, visto ainda que palavras específicas contidas no interior do livro não poderiam ser encontradas a não ser que todo o texto do livro fosse cadastrado também de forma manual.

Além disso, ao procurarmos uma informação em uma biblioteca podemos obter como resposta mais de um livro. Tal fato nos levaria, além de tudo, a ter que decidir qual desses seria o livro mais pertinente aos nossos objetivos. Tarefa essa nada fácil de resolver, principalmente se obtivermos uma quantidade muito grande de livros relacionados ao tema pesquisado, haja visto que deveríamos analisar livro por livro para decidir qual o mais relevante.

Imagine ainda que a quantidade de livros nessa biblioteca aumenta diariamente. Sendo assim, se formos hoje a essa biblioteca e procurarmos por livros sobre culinária, por exemplo, e encontramos os livros A, B e C relacionados ao assunto. Após analisarmos os livros determinamos que o livro C é o mais interessante, seguido pelo A e depois

pelo B. Realizando a mesma pesquisa amanhã, pode ser que essa classificação não seja mais a mesma, visto que novos livros D e E podem ter sido acrescentados à biblioteca mudando assim essa classificação. Além do mais, mesmo que novos livros não sejam integrados à biblioteca, novas edições dos próprios livros A, B e C podem ter chegado, gerando assim uma nova classificação para os mesmos livros.

Com isso, algumas perguntas surgem naturalmente:

Ao efetuarmos uma pesquisa na internet usando o Google, algum funcionário do Google pesquisa manualmente o tema que desejamos em todos os sites relacionados, classifica-os e retorna a lista de resultados em centésimos de segundo?

Com certeza há seres humanos incríveis no mundo, com capacidades acima do normal. Entretanto, tais procedimentos são, com certeza, sobre-humanos no intervalo de tempo descrito. A internet é formada por trilhões páginas e a mesma é diariamente enxertada com milhares de novas páginas, o que nos levaria a uma constante atualização dos resultados já obtidos. Além disso, milhares de pesquisas são realizadas a cada segundo, o que demandaria muitas pessoas envolvidas nesse processos.

Portanto, uma nova questão é levantada: Como a pesquisa é realizada sem a interferência humana de forma eficiente?

Quando fazemos uma pesquisa na internet usando o motor de busca Google, não buscamos em toda a rede, e sim, nas páginas “pré-cadastradas” pelo Google que estão armazenadas em seus bancos de dados. De forma prática podemos interpretar a rede como um conjunto de pontos (páginas) e de caminhos ligando os pontos (links entre as páginas).



O Google utiliza programas chamados “aranhas”, rastreadores da Web, para descobrir páginas disponíveis publicamente.

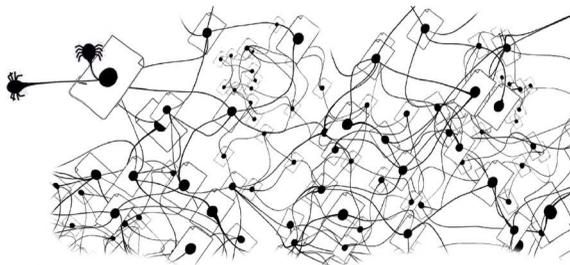


Figura 1 – Rastreadores

Esses rastreadores começam por algumas páginas da web e seguem os links dessas páginas para as quais eles direcionam. Chegando às novas páginas repetem o processo de forma a rastrear uma grande parte da web. A medida que são rastreadas, as páginas são numeradas, bem como a maior parte do seu conteúdo, de modo que quando uma palavra é pesquisada as páginas que estão relacionadas ao tema possam ser localizadas no banco de dados. Por exemplo, imagine que queiramos saber qual o peso de uma baleia azul. Vamos até a janela de busca do Google e digitamos: “PESO BALEIA AZUL” e clicamos ENTER.



Figura 2 – Janela do Google

Um programa pesquisa no banco de dados todas as páginas que incluam esses termos (sem interferência humana alguma). Nesse caso, há centenas de milhares de resultados possíveis, desde os que envolvem as três palavras juntas: “PESO BALEIA AZUL”, aos que envolvem cada uma delas separadas: “PESO”, “BALEIA” e “AZUL”. Dessa

forma, certamente, temos várias páginas que não informam o que realmente queremos saber. Por outro lado, os links que nos interessam, em geral, aparecem nas primeiras colocações da primeira página de resultados. Inclusive são dos primeiros links, da primeira página de resultados, onde encontramos que uma baleia azul tem seu peso variando de 80 a 120 toneladas.

Dessa forma, como o Google decide quais resultados mais nos interessam?

A resposta é: com a análise de mais de 200 variáveis, cada uma com um respectivo peso. Algumas dessas variáveis são: Quantidade de vezes que essa página contém suas palavras-chave, posições em que aparecem as palavras pesquisadas na página: no título da página ou no nome do site?, PageRank dessa página, etc.

A abordagem deste trabalho se restringirá à ponderação do algoritmo PageRank de cada página. Explicaremos a matemática por trás desse algoritmo, como segue.

1.1 O PAGERANK

O PageRank de cada página é dado por uma fórmula matemática, desenvolvida pelos fundadores do Google, Larry Page e Sergey Brin, a qual avalia a importância de uma página em função de páginas externas conectadas a ela através de links, numa espécie de democracia da web, em que esses links são considerados como uma forma de voto.

Abstratamente, representamos a web como um *grafo dirigido*: um conjunto de pontos chamados de *vértices* e um conjunto de flechas ligando esses vértices a outros vértices chamados de *links*.

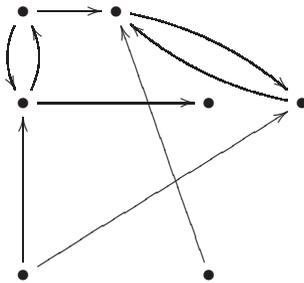


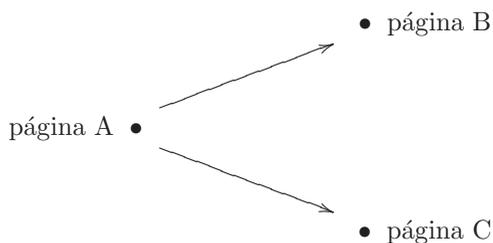
Figura 3 – Modelo de grafo dirigido

Nosso objetivo é atribuir uma pontuação para cada página, isto

é, associar a cada página um número real, maior do que ou igual a zero, de modo que quanto maior o valor do número associado à página, mais importante ela é. Dessa forma, podemos organizar esses números, em ordem crescente, por exemplo, definindo assim, uma ordem de importância para as páginas da internet. Agora, o mais importante: queremos que o computador o faça por conta própria! É impossível para nós avaliarmos a importância de cada página como um todo. Temos, simplesmente, informações demais para processar. Por outro lado, o computador é incapaz de interpretar o conteúdo de cada página e decidir a importância dela embasado apenas no que está escrito nela.

Então, como o fazer?

Uma maneira possível é deixar o conteúdo das páginas de lado e utilizar apenas as informações deixadas por humanos nas páginas que possam ajudar a decidir a importância da página em questão, no contexto da internet como um todo: os links entre as páginas. Ou seja, o Google pontua a importância de uma página em função dos links que se conectam a ela. Um link da página A para a página B indica que a página A “votou” pela importância da página B, isto é, se um site X cita (possui um link para) outro site Y, esse site Y é, mesmo que minimamente, importante para o site X. Além disso, deve-se ponderar a importância que cada link transmite de uma página para outra. Quanto maior a importância da página A, maior deverá ser o peso do seu voto para as páginas que ela possui links. No exemplo que segue, a página A contribui para as importâncias das páginas B e C, isto é, A possui link para B e link para C. Esperamos que essa democracia da web ocorra de forma justa, ou seja, a página A deve contribuir de forma igual para a importância da página B e da página C. Portanto, a página A vota com metade de sua importância para a página B e a outra metade para a página C.



Assim, para avaliar a importância de uma página da internet, usamos a importância das páginas que possuem links para ela. No modelo de internet de 4 páginas, representado na figura 4, a página

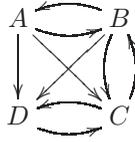


Figura 4 – Modelo de internet de quatro páginas

C recebe links das páginas A, B e D. Dizemos que a importância da página C é gerada pela *soma* das contribuições dadas à página C pelas páginas A, B e D. Mais especificamente, se x_A indica a importância da página A e além disso, a página A vota em três páginas distintas (B, C e D), então a parcela de importância atribuída à página C através da página A é $\frac{x_A}{3}$. Analogamente, se x_B e x_D são, respectivamente, as importâncias das páginas B e D, e além disso, a página B vota em três páginas distintas (A, C e D) e a página D vota em uma única página (C), então a importância atribuída à página C pelas páginas B e D são, respectivamente, $\frac{x_B}{3}$ e $\frac{x_D}{1}$. Logo, a importância da página C, indicada por x_C é dada por:

$$x_C = \frac{x_A}{3} + \frac{x_B}{3} + \frac{x_D}{1},$$

(soma das contribuições). De forma análoga, encontra-se que

$$x_A = \frac{x_B}{3}, \quad x_B = \frac{x_A}{3} + \frac{x_C}{2}, \quad \text{e} \quad x_D = \frac{x_A}{3} + \frac{x_B}{3} + \frac{x_C}{2}.$$

Dessa forma, estabelecemos uma maneira para atribuir a pontuação de importância às páginas. No entanto, percebemos algo curioso: para sabermos a importância de A, precisamos conhecer a importância de B, para sabermos a importância de B, precisamos da importância de A e C, para sabermos a importância de C precisamos das importâncias de A, B e D e por fim, para sabermos a importância de D precisamos das importâncias de A, B e C. Isso gera a impressão que para descobrirmos x_C precisamos saber quanto vale x_C ! Porém, essa não é a questão. Na prática o que acontece é que não precisamos saber de antemão os valores que indicam as importâncias de cada página. Apenas indicamos relações, as quais gostaríamos que os valores referentes às pontuações das páginas obedecessem. Uma vez que essas relações são estabelecidas precisamos saber se existem números que as satisfaçam. Em outras palavras, x_A , x_B , x_C e x_D são *incógnitas* e se relacionam por meio de *equações*. Portanto, se a internet fosse como a representada no exemplo anterior, teríamos uma equação por página,

$$\left\{ \begin{array}{l} x_A = \frac{x_B}{3} \\ x_B = \frac{x_A}{3} + \frac{x_C}{2} \\ x_C = \frac{x_A}{3} + \frac{x_B}{3} + \frac{x_D}{1} \\ x_D = \frac{x_A}{3} + \frac{x_B}{3} + \frac{x_C}{2} \end{array} \right. \quad (1.1)$$

de modo que, determinar a importância de cada página significa encontrar números x_A , x_B , x_C e x_D que satisfaçam, simultaneamente, todas as equações do *Sistema de Equações Lineares* 1.1, que nesse caso possui 4 equações e 4 incógnitas, mais precisamente, temos que determinar, caso exista, uma solução para esse sistema.

Note que o sistema 1.1 possui infinitas soluções, e uma solução possível é $x_A = 3$, $x_B = 9$, $x_C = 16$ e $x_D = 12$ (leitor, verifique!). Colocando as soluções em ordem crescente,

$$x_A = 3 < x_B = 9 < x_D = 12 < x_C = 16.$$

Com isso, ao medirmos a importância das páginas da internet de acordo com os critérios que estabelecemos, obtivemos que a página *C* é a de maior importância. Na sequência temos que a página *D* é a segunda mais importante, seguida pela *B* em terceira colocação e por último a página *A* (página menos importante).

Não obstante, o exemplo anterior de uma internet composta apenas por quatro páginas temos um caso muito simplificado da internet real, composta por trilhões de páginas. O método apresentado acima envolve montar e resolver um sistema de equações com uma equação por página, isto é, no caso real, teríamos trilhões de equações e trilhões de incógnitas. Além disso, a montagem e resolução desse sistema deve ser realizada muito rapidamente - de fato, ao pesquisarmos no Google queremos respostas imediatas. Para resolver esse problema de agilidade e eficiência, necessitamos de conceitos matemáticos bastante avançados, mais especificamente de uma das áreas da Matemática chamada de Álgebra Linear, a qual surgiu a partir do estudo de sistemas de equações lineares.

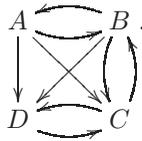
2 MODELANDO O PROBLEMA

Neste capítulo, abordaremos de maneira formal os conceitos matemáticos necessários para modelarmos o problema de encontrar uma pontuação de importância para cada uma das páginas da internet. Mais especificamente, precisaremos dos seguintes conceitos: matriz colunares estocástica, autovalores e autovetores associados a uma matriz quadrada, bem como as definições de base, dimensão e espaço vetorial.

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], [4] e [5].

2.1 A MATRIZ DO GRAFO

No Capítulo 1 determinamos as pontuações de importância para as páginas de uma internet representada pelo grafo



Cada página dessa internet é representada por um vértice do grafo e uma flecha entre dois desses vértices representa um link entre duas páginas.

Considere agora, um modelo de internet composto por n páginas: $P_1, P_2, P_3, \dots, P_n$. Além disso, por uma questão de simplicidade, vamos supor que toda página possui pelo menos um link que sai dela para outra página. Portanto, ao analisarmos o grafo dessa internet, percebemos que todo vértice possui pelo menos uma flecha saindo dele. Quando há um vértice sem flechas saindo dele, o chamaremos de *poço* ou *sumidouro*. No grafo da figura 5, B é um sumidouro.

Agora, admita um grafo com n vértices, que não possui sumidouros, associado a uma internet de n páginas. Representemos os vértices por $1, 2, 3, \dots, n$. Já vimos como associar uma matriz à internet e o mesmo raciocínio se aplica aqui. Para cada $j \in \{1, 2, \dots, n\}$, n_j denota o número de páginas que recebem links da página P_j . Agora, para cada $k \in \{1, 2, \dots, n\}$ definimos L_k como sendo o conjunto de todas as páginas que possuem links para a página k , isto é, L_k denota o con-

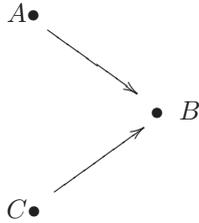
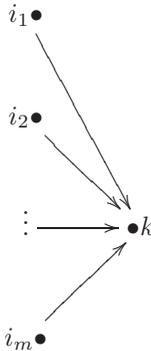


Figura 5 – Modelo de sumidouro

junto de todos os vértices do grafo, os quais possuem flechas saindo deles para o vértice k .



Na ilustração acima, $L_k = \{i_1, i_2, \dots, i_m\}$ em que $i_m \in \{1, 2, \dots, n\}$.

Analogamente ao que fizemos para o modelo da internet da figura 4, denotamos a nota da página P_k por x_k , em que

$$x_k = \sum_{j \in L_k} \frac{1}{n_j} \cdot x_j. \quad (2.1)$$

Esse somatório indica que cada página P_j com um link para P_k , contribui para a nota x_k de P_k com $\frac{1}{n_j}$ de sua própria nota. Tal raciocínio dá origem a um sistema de equações lineares, cuja representação matricial é da forma

$$A \cdot x = x$$

em que A é uma matriz quadrada $n \times n$ e

$$x = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}_{n \times 1}.$$

Denote por a_{ij} , com $i, j \in \{1, 2, \dots, n\}$ as entradas da matriz A . Assim, $a_{ij} = \frac{1}{n_j}$, se P_j possui links para P_i e $a_{ij} = 0$ caso contrário.

Usando a ideia descrita acima para a internet de 4 páginas ilustrada pelo grafo dirigido da figura 6 notamos que a página 1 recebe link

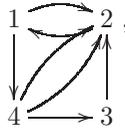


Figura 6 – Grafo dirigido

apenas da página 2. Por outro lado, a página 2 recebe links das páginas 1, 3 e 4, a página 3 recebe link apenas de 4 e finalmente, a página 4 recebe link apenas de 1. Com isso, denotamos $L_1 = \{2\}$, $L_2 = \{1, 3, 4\}$, $L_3 = \{4\}$ e $L_4 = \{2, 3\}$. Além disso, a página 1 possui links para duas páginas, a saber as páginas 2 e 4, com isso, $n_1 = 2$. De forma análoga, contando as páginas, às quais 2, 3 e 4 possuem links, obtemos, respectivamente, que $n_2 = 1$, $n_3 = 1$ e $n_4 = 2$. (Observação: Lembre-se que n_j denota o número de páginas que recebem links da página j e não o número de links que saem da página j , dessa forma, n_4 por exemplo, vale 2 e não 3 como pode parecer se contarmos o número de links que saem da página 4). Com isso, as expressões para a pontuação das páginas 1, 2, 3 e 4 são, respectivamente:

$$x_1 = \sum_{j \in L_1} \frac{1}{n_j} \cdot x_j = \frac{1}{n_2} \cdot x_2 = \frac{x_2}{1},$$

$$x_2 = \sum_{j \in L_2} \frac{1}{n_j} \cdot x_j = \frac{1}{n_1} \cdot x_1 + \frac{1}{n_3} \cdot x_3 + \frac{1}{n_4} \cdot x_4 = \frac{x_1}{2} + \frac{x_3}{1} + \frac{x_4}{2},$$

$$x_3 = \sum_{j \in L_3} \frac{1}{n_j} \cdot x_j = \frac{1}{n_4} \cdot x_4 = \frac{x_4}{2}$$

e

$$x_4 = \sum_{j \in L_4} \frac{1}{n_j} \cdot x_j = \frac{1}{n_1} \cdot x_1 = \frac{x_1}{2}.$$

Com essas equações montamos o sistema linear

$$\left\{ \begin{array}{l} x_1 = \frac{x_2}{1} \\ x_2 = \frac{x_1}{2} + \frac{x_3}{1} + \frac{x_4}{2} \\ x_3 = \frac{x_4}{2} \\ x_4 = \frac{x_1}{2} \end{array} \right. ,$$

o qual pode ser escrito na forma $A \cdot x = x$:

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

em que

$$x = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix}$$

é uma matriz coluna 4×1 cujas entradas x_1 , x_2 , x_3 e x_4 representam as variáveis do problema, isto é, as pontuações de importância, respectivamente, das páginas 1, 2, 3 e 4. Além disso, a matriz

$$A = \begin{pmatrix} 0 & 1 & 0 & 0 \\ \frac{1}{2} & 0 & 1 & \frac{1}{2} \\ 0 & 0 & 0 & \frac{1}{2} \\ \frac{1}{2} & 0 & 0 & 0 \end{pmatrix}$$

é denominada *matriz do grafo*.

Definição 1. Define-se um vetor como uma matriz coluna $n \times 1$ com entradas reais. Denotaremos por

$$v = \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix},$$

em que $v_i \in \mathbb{R}$, com $1 \leq i \leq n$, ou ainda, $v \in \mathbb{R}^n$.

Observação 2. $0_{n \times 1}$ denota o vetor nulo de ordem $n \times 1$.

2.2 MATRIZES COLUNA-ESTOCÁSTICAS E OS AUTOVALORES

No exemplo anterior, a matriz do grafo possui a soma das entradas de cada coluna igual a 1.

Definição 3. Uma matriz quadrada com entradas não negativas, cuja somas das entradas de cada coluna é igual a 1 é chamada *matriz coluna-estocástica*.

As matrizes dos grafos associadas a modelos de internet sem sumidouros são sempre matrizes coluna-estocástica. Isso ocorre visto que as entradas da coluna j correspondem às contribuições da página P_j para as notas das páginas que recebem links de P_j . Ou seja, cada entrada não nula nessa coluna é igual a $\frac{1}{n_j}$ e, como há n_j entradas não nulas nessa coluna - existem n_j páginas que recebem links de P_j - a soma das entradas da coluna j é dada por

$$\underbrace{\frac{1}{n_j} + \frac{1}{n_j} + \cdots + \frac{1}{n_j}}_{n_j \text{ parcelas}} = 1.$$

Sejam A uma matriz quadrada de ordem n e x um vetor em \mathbb{R}^n . No nosso caso, encontrar uma solução para um sistema da forma $A \cdot x = x$, ou ainda $(A - I) \cdot x = 0$, é equivalente a determinar uma

lista de notas, uma para cada página. Obviamente,

$$x = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 0 \end{pmatrix}_{n \times 1}$$

é solução desse sistema. No entanto, esta solução nos diz que todas as páginas são igualmente importantes, visto que cada uma delas recebeu a mesma pontuação igual a zero. No entanto, estamos interessados em determinar possíveis soluções não *triviais* para esse sistema. Para isso usaremos alguns conceitos de Álgebra Linear.

Definição 4. Sejam A uma matriz quadrada de ordem n e $v \in \mathbb{R}^n$ um vetor não nulo. Se existe um escalar α tal que $A \cdot v = \alpha \cdot v$, então esse escalar α é denominado um *autovalor* de A e v um *autovetor* de A associado a esse escalar α .

Agora, note que se um vetor $b \in \mathbb{R}^n$, não nulo, é solução do sistema $A \cdot x = x$, então $\lambda = 1$ é um autovalor da matriz A . Por outro lado, se $\lambda = 1$ é um autovalor associado à matriz A , então todo autovetor associado ao autovalor 1 será uma solução desse sistema. No entanto, perceba que nem toda matriz quadrada possui autovalor igual a 1. Contudo, isso sempre ocorre para matrizes coluna-estocásticas, o que nos leva ao seguinte teorema.

Teorema 5. Toda matriz B coluna-estocástica possui 1 como autovalor.

Prova:

Sejam B uma matriz coluna-estocástica $n \times n$, B^T sua transposta e v um vetor coluna $n \times 1$ com todas as entradas iguais a 1. Como B é coluna-estocástica, a soma das entradas de cada coluna é igual a 1, e portanto, em B^T , a soma das entradas de cada linha é igual a 1, donde

$$\begin{aligned}
B^T \cdot v &= \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix} \cdot \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} \\
&= \begin{pmatrix} a_{11} + a_{12} + \cdots + a_{1n} \\ a_{21} + a_{22} + \cdots + a_{2n} \\ \vdots \\ a_{n1} + a_{n2} + \cdots + a_{nn} \end{pmatrix} \\
&= \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} \\
&= v
\end{aligned}$$

comprovando que 1 é autovalor de B^T associado ao autovetor v . Como B e B^T possuem os mesmos autovalores, temos que 1 é autovalor de B . Isso ocorre pois, da Álgebra Linear, um sistema da forma $B \cdot x = \lambda \cdot x$ pode ser reescrito como o sistema homogêneo $(B - \lambda I) \cdot x = 0$. Tal sistema possui soluções não triviais quando $\det(B - \lambda I) = 0$, visto que $x \neq 0$. Essas soluções são os autovalores de B . No entanto, perceba que $\det(B - \lambda I) = \det(B - \lambda I)^T$, donde $\det(B - \lambda I) = \det(B^T - \lambda I)$, o que indica que B e B^T possuem os mesmos autovalores.

Dessa forma, para uma matriz da internet sem sumidouros, existem soluções não nulas do sistema $A \cdot x = x$, as quais podemos usar para atribuir notas às páginas da internet do modo como foi exposto: utilizando um autovetor de A associado ao seu autovalor 1.

2.3 DEFICIÊNCIAS DO MÉTODO

2.3.1 INTERNET DESCONECTADA

Até agora nos preocupamos em encontrar uma forma de atribuir uma pontuação de importância para as páginas da internet. Tivemos

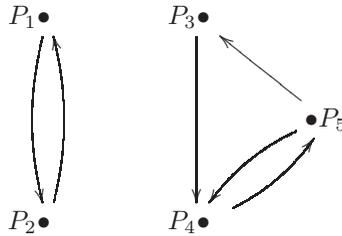


Figura 7 – Modelo de internet desconexa

êxito, porém, alguns problemas acabam surgindo e temos que solucioná-los. Um deles vem do fato de que o autovalor $\lambda = 1$ possui infinitos autovetores associados, ou seja, podemos atribuir notas de infinitas maneiras diferentes, todas satisfazendo as condições exigidas. Dessa forma, a ordem de importância das páginas pode mudar dependendo de qual dos autovetores escolhemos. Por exemplo, considere a internet como a representada pelo grafo

A matriz do grafo é

$$A = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{2} \\ 0 & 0 & 1 & 0 & \frac{1}{2} \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix}$$

donde duas possíveis soluções para o sistema $A \cdot x = x$, isto é, dois autovetores associados ao autovalor $\lambda = 1$, são

$$v_1 = \begin{pmatrix} \frac{1}{2} \\ \frac{1}{2} \\ 0 \\ 0 \\ 0 \end{pmatrix} \quad \text{e} \quad v_2 = \begin{pmatrix} 0 \\ 0 \\ \frac{1}{2} \\ 1 \\ 1 \end{pmatrix}.$$

Veja que na solução v_1 , as páginas P_1 e P_2 são as mais importantes, mas na solução v_2 , as páginas P_4 e P_5 são as mais importantes! Com isso, algumas perguntas devem ser respondidas: Qual dos autovetores devemos usar para a classificação das páginas? Em quais condições obteremos mais de um autovetor para a classificação das páginas?

Para tal, considere $V_1(A)$ o *auto-espaço* de A associado ao autovalor $\lambda = 1$, isto é,

$$V_1(A) = \{x \in \mathbb{R}^n \mid A \cdot x = x\}.$$

Já sabemos que $V_1(A)$ é *subespaço vetorial* do espaço $M_{n \times 1}(\mathbb{R})$ das matrizes coluna $n \times 1$. No exemplo acima, $\{v_1, v_2\}$ forma uma base de $V_1(A)$, logo $\dim(V_1(A)) = 2$.

Com isso, sempre que $\dim(V_1(A)) > 1$, teremos este problema: existirão ao menos 2 vetores linearmente independentes em $V_1(A)$ e cada um deles nos dá uma forma diferente de atribuir notas para as páginas da internet. Em outras palavras, haverão ao menos dois vetores, dos quais um deles não é múltiplo do outro, que servirão para classificarmos a importância das páginas da internet. Além disso, note que o grafo do exemplo anterior é composto de dois “subgrafos”: as duas partes desconectadas dessa internet. Tal fato, gera na matriz do grafo, dois blocos,

$$A = \left(\begin{array}{cc|ccc} B_{2 \times 2} & & 0_{2 \times 3} & & & \\ & & & & & \\ \hline & & C_{3 \times 3} & & & \end{array} \right) = \left(\begin{array}{cc|ccc} 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & \frac{1}{2} \\ 0 & 0 & 1 & 0 & \frac{1}{2} \\ 0 & 0 & 0 & 1 & 0 \end{array} \right),$$

os quais podem ser interpretados como as matrizes dos grafos de cada uma desses subgrafos desconectados, o que força $\dim(V_1(A))$ a ser maior do que 1. Mais precisamente, temos o seguinte resultado:

Teorema 6. Se o grafo de uma internet sem sumidouros é constituído de r subgrafos, em que $r > 1$, então $\dim(V_1(A)) \geq r$.

Prova: Inicialmente, observemos que se temos uma internet com n páginas, P_1, P_2, \dots, P_n , então na matriz da internet há uma coluna para cada página, bem como uma linha para cada página.

A página P_j está associada à coluna j da matriz, ou ainda, as entradas desta coluna revelam se P_j possui links para as demais páginas, mais precisamente, a i -ésima entrada na coluna j indica se P_j possui links para P_i . Dessa forma, interpretamos que P_j possui links para P_i se, e somente se, $a_{ij} \neq 0$. Similarmente, a página P_i está associada à linha i da matriz. Temos que P_i recebe links de P_j se, e somente se, $a_{ij} \neq 0$.

Considere portanto, uma internet com n páginas, representada por um grafo com n vértices, constituído por r subgrafos desconectados:

W_1, \dots, W_r , com $r \geq 2$. Numeremos os vértices do grafo de modo que os vértices de W_1 sejam indicados por $1, 2, 3, \dots, N_1$, os vértices de W_2 por $N_1 + 1, \dots, N_2$, e assim sucessivamente, até os vértices de W_r , indicados por $N_{r-1} + 1$ até $N_r = n$. Se estipularmos $N_0 = 1$, podemos observar que o número de vértices em cada subgrafo é dado por

$$n_k = N_k - N_{k-1}.$$

Assim, podemos escrever, $n = n_1 + n_2 + \dots + n_r$. Seja A a matriz do grafo. Esta matriz é quadrada de ordem n e cada linha/coluna de A está associada a um vértice do grafo, pelo que observamos acima. Podemos agrupar as linhas de A em r blocos, um para cada subgrafo, como ilustrado a seguir:

$$\begin{array}{l}
 W_1 \left\{ \begin{array}{l} 1 \\ \vdots \\ N_1 \end{array} \right. \\
 \\
 W_2 \left\{ \begin{array}{l} N_1 + 1 \\ \vdots \\ N_2 \end{array} \right. \\
 \vdots \\
 \\
 W_r \left\{ \begin{array}{l} N_{r-1} + 1 \\ \vdots \\ N_r \end{array} \right.
 \end{array}
 \left(\begin{array}{c} \text{Bloco 1} \\ \hline \text{Bloco 2} \\ \hline \vdots \\ \hline \text{Bloco r} \end{array} \right)$$

O mesmo pode ser feito para as colunas:

$$\left(\begin{array}{c|c|c|c} \overbrace{1, \dots, N_1}^{W_1} & \overbrace{N_1 + 1, \dots, N_2}^{W_2} & \dots & \overbrace{N_{r-1} + 1, \dots, N_r}^{W_r} \\ \hline \text{Bloco 1} & \text{Bloco 2} & \dots & \text{Bloco r} \end{array} \right)$$

Combinando essas duas decomposições podemos subdividir a matriz A em r^2 sublocos:

$$\begin{array}{c} W_1 \quad \dots \quad W_l \quad \dots \quad W_r \\ \begin{array}{c} W_1 \\ \vdots \\ W_k \\ \vdots \\ W_r \end{array} \left(\begin{array}{c|c|c|c|c} & & & & \\ \hline & & & & \\ \hline & & B_{kl} & & \\ \hline & & & & \\ \hline & & & & \\ \hline & & & & \end{array} \right) \end{array}$$

Para cada $k, l \in \{1, \dots, r\}$, $B_{kl} = \text{Bloco } kl = [a_{ij}]$ é a matriz em que i varia de $N_{k-1} + 1$ até N_k e j varia de $N_{l-1} + 1$ até N_l . Note que B_{kl} é uma matriz $n_k \times n_l$. Seja $j \in \{1, \dots, n\}$ um vértice qualquer do grafo. Digamos que este vértice pertença ao subgrafo W_l , para algum $l \in \{1, \dots, r\}$. Sabemos que a página P_j da internet possui apenas links para páginas que correspondem a vértices do mesmo subgrafo W_l de j . Isto quer dizer que se i é um vértice que não pertence a W_l , então a entrada a_{ij} da matriz do subgrafo W_l é igual a zero. Em outras palavras, se $i \in \{1, \dots, n\}$ e $i \notin \{N_{l-1} + 1, \dots, N_l\}$, então $a_{ij} = 0$. Como consequência, se $k \in \{1, \dots, r\}$ e $k \neq l$, então o bloco B_{kl} da matriz A

possui apenas entradas nulas. Assim, a matriz A pode ser escrita como,

$$A = \left(\begin{array}{c|c|c|c|c} B_{11} & 0 & 0 & \dots & 0 \\ \hline 0 & B_{22} & 0 & \dots & 0 \\ \hline \vdots & \vdots & \vdots & \vdots & \vdots \\ \hline 0 & 0 & 0 & B_{r-1r-1} & 0 \\ \hline 0 & 0 & 0 & \dots & B_{rr} \end{array} \right).$$

Não obstante, as matrizes em blocos, da forma B_{ll} , com $l \in \{1, \dots, r\}$, podem ser interpretadas como a matriz associada ao subgrafo W_l , com $l \in \{1, \dots, r\}$ da internet de n páginas com r subgrafos desconexos.

Perceba ainda, que cada matriz quadrada B_{ii} de ordem n_i é coluna-estocástica, portanto, pelo Teorema 5 possui 1 como autovalor, e dessa forma, existe algum autovetor $v_i \in \mathbb{R}^{n_i}$ associado ao autovalor 1. Agora, para cada $i \in \{1, \dots, r\}$ seja $w_i \in \mathbb{R}^n$ um vetor subdivido em r blocos, da seguinte forma: para cada $j \in \{1, \dots, r\}$, o bloco j de w_i é igual a: o vetor nulo $n_j \times 1$ se $j \neq i$, ou o vetor v_i no caso $j = i$. Com isso, montamos o vetor w_i , como segue:

$$w_i = \left(\begin{array}{c} 0 \\ \vdots \\ 0 \\ 0 \\ \vdots \\ 0 \\ \vdots \\ v_i \\ \vdots \\ 0 \\ \vdots \\ 0 \end{array} \right) \left. \begin{array}{l} \vphantom{\left(\begin{array}{c} 0 \\ \vdots \\ 0 \\ 0 \\ \vdots \\ 0 \\ \vdots \\ v_i \\ \vdots \\ 0 \\ \vdots \\ 0 \end{array} \right)} \\ \vphantom{\left(\begin{array}{c} 0 \\ \vdots \\ 0 \\ 0 \\ \vdots \\ 0 \\ \vdots \\ v_i \\ \vdots \\ 0 \\ \vdots \\ 0 \end{array} \right)} \\ \vphantom{\left(\begin{array}{c} 0 \\ \vdots \\ 0 \\ 0 \\ \vdots \\ 0 \\ \vdots \\ v_i \\ \vdots \\ 0 \\ \vdots \\ 0 \end{array} \right)} \\ \vphantom{\left(\begin{array}{c} 0 \\ \vdots \\ 0 \\ 0 \\ \vdots \\ 0 \\ \vdots \\ v_i \\ \vdots \\ 0 \\ \vdots \\ 0 \end{array} \right)} \\ \vphantom{\left(\begin{array}{c} 0 \\ \vdots \\ 0 \\ 0 \\ \vdots \\ 0 \\ \vdots \\ v_i \\ \vdots \\ 0 \\ \vdots \\ 0 \end{array} \right)} \\ \vphantom{\left(\begin{array}{c} 0 \\ \vdots \\ 0 \\ 0 \\ \vdots \\ 0 \\ \vdots \\ v_i \\ \vdots \\ 0 \\ \vdots \\ 0 \end{array} \right)} \\ \vphantom{\left(\begin{array}{c} 0 \\ \vdots \\ 0 \\ 0 \\ \vdots \\ 0 \\ \vdots \\ v_i \\ \vdots \\ 0 \\ \vdots \\ 0 \end{array} \right)} \\ \vphantom{\left(\begin{array}{c} 0 \\ \vdots \\ 0 \\ 0 \\ \vdots \\ 0 \\ \vdots \\ v_i \\ \vdots \\ 0 \\ \vdots \\ 0 \end{array} \right)} \\ \vphantom{\left(\begin{array}{c} 0 \\ \vdots \\ 0 \\ 0 \\ \vdots \\ 0 \\ \vdots \\ v_i \\ \vdots \\ 0 \\ \vdots \\ 0 \end{array} \right)} \\ \vphantom{\left(\begin{array}{c} 0 \\ \vdots \\ 0 \\ 0 \\ \vdots \\ 0 \\ \vdots \\ v_i \\ \vdots \\ 0 \\ \vdots \\ 0 \end{array} \right)} \\ \vphantom{\left(\begin{array}{c} 0 \\ \vdots \\ 0 \\ 0 \\ \vdots \\ 0 \\ \vdots \\ v_i \\ \vdots \\ 0 \\ \vdots \\ 0 \end{array} \right)} \end{array} \right\} \begin{array}{l} v_1 \\ \\ v_2 \\ \\ \vdots \\ \\ v_{n_i} \end{array}.$$

Veja que os vetores w_i , com $1 \leq i \leq r$, da forma

$$w_1 = \begin{pmatrix} v_1 \\ 0 \\ 0 \\ \vdots \\ 0 \\ 0 \end{pmatrix}, w_2 = \begin{pmatrix} 0 \\ v_2 \\ 0 \\ \vdots \\ 0 \\ 0 \end{pmatrix}, \dots, w_i = \begin{pmatrix} 0 \\ \vdots \\ v_i \\ \vdots \\ 0 \\ 0 \end{pmatrix}, \dots,$$

são autovetores de A associados ao autovalor $\lambda = 1$, pois,

$$A \cdot w_i = \begin{pmatrix} A_1 & 0 & 0 & \dots & 0 \\ 0 & A_2 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & A_{r-1} & 0 \\ 0 & 0 & 0 & \dots & A_r \end{pmatrix} \cdot \begin{pmatrix} 0 \\ \vdots \\ v_i \\ \vdots \\ 0 \\ 0 \end{pmatrix} = w_i,$$

em que cada $A_i = B_{ii}$, com $1 \leq i \leq r$.

Além disso, é fácil ver que os vetores w_i , com $1 \leq i \leq r$, são linearmente independentes, visto que a equação

$$\alpha_1 \cdot w_1 + \alpha_2 \cdot w_2 + \dots + \alpha_r \cdot w_r = 0_{n \times 1}$$

é satisfeita somente quando

$$\alpha_1 = \alpha_2 = \dots = \alpha_r = 0.$$

De fato,

$$\alpha_1 \cdot w_1 + \alpha_2 \cdot w_2 + \dots + \alpha_r \cdot w_r = 0_{n \times 1} \quad (2.2)$$

$$\alpha_1 \cdot \begin{pmatrix} v_1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix} + \alpha_2 \cdot \begin{pmatrix} 0 \\ v_2 \\ 0 \\ \vdots \\ 0 \end{pmatrix} + \dots + \alpha_r \cdot \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ v_r \end{pmatrix} = 0_{n \times 1}$$

$$\begin{pmatrix} \alpha_1 \cdot v_1 \\ \alpha_2 \cdot v_2 \\ \alpha_3 \cdot v_3 \\ \vdots \\ \alpha_r \cdot v_r \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}.$$

Como cada v_i , com $1 \leq i \leq r$, é não nulo, pois são autovetores de A associados ao autovalor $\lambda = 1$, a única solução para a equação 2.2 é $\alpha_1 = \alpha_2 = \dots = \alpha_r = 0$.

Portanto, $\dim(V_1(A)) \geq r$.

Com base no que provamos no Teorema 6, precisamos determinar uma forma de que $\dim(V_1(A)) = 1$, visto que com essa informação conseguimos uma única classificação de importância para as páginas da internet.

Definição 7. Uma matriz $M_{m \times n}$ é *positiva* se $a_{ij} > 0$ para todo $1 \leq i \leq m$ e $1 \leq j \leq n$.

Proposição 8. Se $M_{n \times n}$ é uma matriz positiva e coluna-estocástica, então qualquer autovetor $v \in V_1(M)$ possui todas as entradas positivas ou todas as entradas negativas.

Prova: Sejam

$$v = \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix}$$

e

$$M = \begin{pmatrix} m_{11} & \cdots & m_{1i} & \cdots & m_{1n} \\ m_{21} & \cdots & m_{2i} & \cdots & m_{2n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ m_{n1} & \cdots & m_{ni} & \cdots & m_{nn} \end{pmatrix},$$

tal que $v \in V_1(M)$. Suponha, por absurdo, que v não possui, todas as entradas positivas ou todas as entradas negativas, isto é, possui simultaneamente entradas positivas e entradas negativas.

Como $v \in V_1(M)$,

$$M \cdot v = v,$$

$$\begin{pmatrix} m_{11} & \cdots & m_{1i} & \cdots & m_{1n} \\ m_{21} & \cdots & m_{2i} & \cdots & m_{2n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ m_{n1} & \cdots & m_{ni} & \cdots & m_{nn} \end{pmatrix} \cdot \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix} = \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix},$$

$$\begin{pmatrix} m_{11}v_1 + \cdots + m_{1i}v_i + \cdots + m_{1n}v_n \\ m_{21}v_1 + \cdots + m_{2i}v_i + \cdots + m_{2n}v_n \\ \vdots \\ m_{n1}v_1 + \cdots + m_{ni}v_i + \cdots + m_{nn}v_n \end{pmatrix} = \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix}.$$

$$\begin{cases} m_{11}v_1 + \cdots + m_{1i}v_i + \cdots + m_{1n}v_n = v_1 \\ m_{21}v_1 + \cdots + m_{2i}v_i + \cdots + m_{2n}v_n = v_2 \\ \vdots \\ m_{n1}v_1 + \cdots + m_{ni}v_i + \cdots + m_{nn}v_n = v_n \end{cases}.$$

Aplicando módulo nos dois membros de cada equação vem que

$$\begin{cases} |m_{11}v_1 + \cdots + m_{1i}v_i + \cdots + m_{1n}v_n| = |v_1| \\ |m_{21}v_1 + \cdots + m_{2i}v_i + \cdots + m_{2n}v_n| = |v_2| \\ \vdots \\ |m_{n1}v_1 + \cdots + m_{ni}v_i + \cdots + m_{nn}v_n| = |v_n| \end{cases}$$

Usando a desigualdade triangular juntamente com o fato de que v tem entradas mistas e cada $m_{ij} > 0$, visto que M é uma matriz positiva, temos

$$\begin{cases} |m_{11}v_1| + \cdots + |m_{1i}v_i| + \cdots + |m_{1n}v_n| \geq |v_1| \\ |m_{21}v_1| + \cdots + |m_{2i}v_i| + \cdots + |m_{2n}v_n| \geq |v_2| \\ \vdots \\ |m_{n1}v_1| + \cdots + |m_{ni}v_i| + \cdots + |m_{nn}v_n| \geq |v_n| \end{cases}$$

donde,

$$\begin{cases} m_{11}|v_1| + \cdots + m_{1i}|v_i| + \cdots + m_{1n}|v_n| > |v_1| \\ m_{21}|v_1| + \cdots + m_{2i}|v_i| + \cdots + m_{2n}|v_n| > |v_2| \\ \vdots \\ m_{n1}|v_1| + \cdots + m_{ni}|v_i| + \cdots + m_{nn}|v_n| > |v_n| \end{cases}$$

Somando as equações do sistema membro a membro obtemos

$$(m_{11} + \cdots + m_{n1})|v_1| + \cdots + (m_{1n} + \cdots + m_{nn})|v_n| > |v_1| + \cdots + |v_n|.$$

Porém, $m_{1i} + m_{2i} + \cdots + m_{ni}$, $1 \leq i \leq n$, representa a soma dos elementos da coluna i de M que é coluna-estocástica, e portanto, $m_{1i} + m_{2i} + \cdots + m_{ni} = 1$ para todo $1 \leq i \leq n$. Dessa forma,

$$|v_1| + \cdots + |v_n| > |v_1| + \cdots + |v_n|,$$

o que é um absurdo! Portanto, v possui todas as entradas positivas ou todas as entradas negativas.

Proposição 9. Sejam $m \in \mathbb{N}$, $m \geq 2$ e $v, w \in \mathbb{R}^m$, linearmente independentes. Existem $\alpha, \beta \in \mathbb{R}$ não simultaneamente nulos, tais que o vetor $u = \alpha \cdot v + \beta \cdot w$ possui entradas positivas e negativas simultaneamente, isto é, entradas com sinais opostos.

Prova: Sejam $v, w \in \mathbb{R}^m$, linearmente independentes, tais que

$$v = \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_m \end{pmatrix}$$

e

$$w = \begin{pmatrix} w_1 \\ w_2 \\ \vdots \\ w_m \end{pmatrix}.$$

Para demonstrar a proposição dividiremos em casos:

- **Caso 1:**

Considere que ao menos um dos vetores v ou w já possuam entradas positivas e negativas simultaneamente. Logo, basta tomar

$\alpha = 1$ e $\beta = 0$ ou vice-versa.

• **Caso 2:**

Suponha agora que w possua exatamente uma entrada não nula, ou seja, w_i não é zero para um único i com $1 \leq i \leq m$. Como v e w são linearmente independentes, v possui ao menos uma entrada $v_j \neq 0$, em que $1 \leq j \leq m$ e $i \neq j$. Dessa forma, o vetor

$$u = \alpha \cdot v + \beta \cdot w$$

é dado por

$$u = \begin{pmatrix} \alpha \cdot v_1 + \beta \cdot 0 \\ \alpha \cdot v_2 + \beta \cdot 0 \\ \vdots \\ \alpha \cdot v_j + \beta \cdot 0 \\ \vdots \\ \alpha \cdot v_i + \beta \cdot w_i \\ \vdots \\ \alpha \cdot v_m + \beta \cdot 0 \end{pmatrix} = \begin{pmatrix} \alpha \cdot v_1 \\ \alpha \cdot v_2 \\ \vdots \\ \alpha \cdot v_j \\ \vdots \\ \alpha \cdot v_i + \beta \cdot w_i \\ \vdots \\ \alpha \cdot v_m \end{pmatrix}$$

– Subcaso: $v_i = 0$.

I) Se v_j e w_i tiverem sinais opostos, tome $\alpha = 1$ e $\beta = 1$. Dessa forma, o vetor u possui as entradas $u_i = w_i$ e $u_j = v_j$ com sinais opostos.

II) Se v_j e w_i tiverem sinais iguais, tome $\alpha = 1$ e $\beta = -1$. Com isso, o vetor u possui as entradas $u_i = -w_i$ e $u_j = v_j$ com sinais opostos.

– Subcaso: $v_i \neq 0$ e $v_j > 0$.

I) Se $v_i > 0$ e $w_i > 0$, então tome $\alpha = 1$ e $\beta = -d$, em que $d > \frac{v_i}{w_i}$. Dessa forma, u possui as entradas u_i e u_j com sinais opostos, pois, $d > \frac{v_i}{w_i} \Rightarrow u_i = v_i - d \cdot w_i < 0$ e $u_j = v_j > 0$.

II) Se $v_i > 0$ e $w_i < 0$, então tome $\alpha = 1$ e $\beta = -d$, em que $d < \frac{v_i}{w_i}$. Dessa forma, u possui as entradas u_i e u_j com sinais opostos, pois, $d < \frac{v_i}{w_i} \Rightarrow u_i = v_i - d \cdot w_i < 0$ e $u_j = v_j > 0$.

III) Se $v_i < 0$, temos o **Caso 1**.

Os casos em que $v_i \neq 0$ e $v_j < 0$ são análogos.

Suporemos a partir de agora que w possui ao menos duas entradas não nulas.

• **Caso 3:**

Considere $0 \leq w_i \leq v_i$, para todo $1 \leq i \leq m$. Para as entradas $w_i \neq 0$ analise os quocientes $\frac{v_i}{w_i}$. Perceba que há ao menos dois quocientes diferentes da forma $\frac{v_i}{w_i}$, caso contrário, $\frac{v_i}{w_i} = \frac{v_j}{w_j} = k$, e dessa forma, $v_i = k \cdot w_i$, o que caracterizaria que os vetores v e w seriam linearmente dependentes. Com isso, sejam, sem perda de generalidade,

$$v = \begin{pmatrix} v_1 \\ \vdots \\ v_i \\ \vdots \\ v_j \\ \vdots \\ v_m \end{pmatrix} \quad e \quad w = \begin{pmatrix} w_1 \\ \vdots \\ w_i \\ \vdots \\ w_j \\ \vdots \\ w_m \end{pmatrix},$$

em que, $\frac{v_i}{w_i} < \frac{v_j}{w_j}$. Com isso, tome $\alpha = 1$ e $\beta = -d$, de forma que $\frac{v_i}{w_i} < d < \frac{v_j}{w_j}$.

Dessa forma,

$$u = \alpha \cdot v + \beta \cdot w$$

é dado por

$$u = \begin{pmatrix} v_1 - d \cdot w_1 \\ \vdots \\ v_i - d \cdot w_i \\ \vdots \\ v_j - d \cdot w_j \\ \vdots \\ v_m - d \cdot w_m \end{pmatrix}.$$

No entanto, $d < \frac{v_j}{w_j} \Rightarrow u_j = v_j - d \cdot w_j > 0$ e $\frac{v_i}{w_i} < d \Rightarrow u_i = v_i - d \cdot w_i < 0$, o que indica que u possui, ao menos, as entradas u_i e u_j com sinais opostos.

O caso em que $v_i \leq w_i \leq 0$ é análogo.

• **Caso 4:**

Considere que existam $i, j \in \{1, \dots, m\}$ tais que $v_i > w_i$ e $v_j < w_j$. Nesse caso basta tomar $\alpha = 1$ e $\beta = -1$.

De fato, sejam, sem perda de generalidade,

$$v = \begin{pmatrix} v_1 \\ \vdots \\ v_i \\ \vdots \\ v_j \\ \vdots \\ v_m \end{pmatrix} \quad e \quad w = \begin{pmatrix} w_1 \\ \vdots \\ w_i \\ \vdots \\ w_j \\ \vdots \\ w_m \end{pmatrix}.$$

Dessa forma,

$$u = \alpha \cdot v + \beta \cdot w$$

é dado por

$$u = \begin{pmatrix} v_1 - w_1 \\ \vdots \\ v_i - w_i \\ \vdots \\ v_j - w_j \\ \vdots \\ v_m - w_m \end{pmatrix}.$$

Como, $v_i > w_i \Rightarrow u_i = v_i - w_i > 0$ e $v_j < w_j \Rightarrow u_j = v_j - w_j < 0$, temos que u possui, ao menos, as entradas u_i e u_j com sinais opostos.

Não obstante, em todos os casos conseguimos encontrar números reais, não simultaneamente nulos, α e β , tais que $u = \alpha \cdot v + \beta \cdot w$ possui entradas com sinais opostos.

Teorema 10. Se $M_{n \times n}$ é positiva e coluna-estocástica, então a dimensão de $V_1(M)$ é igual a 1.

Prova: Provaremos por contradição. Considere que existam dois autovetores linearmente independentes, v_1 e $v_2 \in V_1(M)$. Para quaisquer $\alpha, \beta \in \mathbb{R}$, não simultaneamente nulos, temos que x , definido como $\alpha \cdot v_1 + \beta \cdot v_2$, está no subespaço $V_1(M)$ e é não nulo, pois, v_1 e v_2 são linearmente independentes, logo é autovetor de M associado ao autovalor 1. Dessa forma, pela Proposição 8, x possui todas as entradas positivas ou todas as entradas negativas. Por outro lado, pela Proposição 9, existem $\alpha, \beta \in \mathbb{R}$, não simultaneamente nulos, tais que $x = \alpha \cdot v_1 + \beta \cdot v_2$ possui entradas positivas e negativas simultaneamente. Contradição! Com isso, não existem dois vetores linearmente independentes pertencentes a $V_1(M)$ e, portanto, $\dim(V_1(M)) = 1$.

Agora, com o resultado do Teorema 10 conseguimos gerar, de forma única, a pontuação de importância para uma internet constituída de r subgrafos desconectados, como segue.

2.3.2 CORRIGINDO AS DEFICIÊNCIAS

Para modelos de internet com n páginas sem sumidouros, podemos gerar pontuações de importâncias únicas até mesmo nos casos desconectados. Para tal, considere a matriz $A_{n \times n}$ de um grafo dirigido com n vértices sem sumidouros e a matriz $S_{n \times n}$ cujas entradas são todas iguais a $\frac{1}{n}$. Note que a matriz

$$S = \begin{pmatrix} \frac{1}{n} & \frac{1}{n} & \cdots & \frac{1}{n} & \frac{1}{n} \\ \frac{1}{n} & \frac{1}{n} & \cdots & \frac{1}{n} & \frac{1}{n} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \frac{1}{n} & \frac{1}{n} & \cdots & \frac{1}{n} & \frac{1}{n} \\ \frac{1}{n} & \frac{1}{n} & \cdots & \frac{1}{n} & \frac{1}{n} \end{pmatrix}$$

pode ser interpretada como a matriz do grafo representado na figura 8, em que cada um dos n vértices desse grafo possui um link para cada um dos outros vértices. Claramente, a matriz S é coluna-estocástica e todas as suas entradas são positivas, logo, pelo Teorema 10, $\dim(V_1(S)) = 1$. Agora, para um valor $m \in [0, 1]$, considere a matriz

$$M = (1 - m) \cdot A + m \cdot S.$$

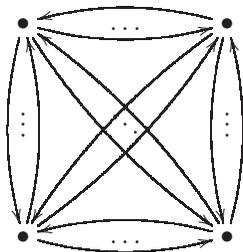


Figura 8 – Grafo

Veja que M definida dessa forma é igual a A para $m = 0$, porém igual a S para $m = 1$. A matriz M é uma “perturbação” da matriz A .

Na prática, ao substituir a matriz A por M , estamos atribuindo, de maneira igual, uma pequena pontuação para cada uma das páginas, ou ainda, como se cada uma das páginas tivesse, por menor que seja, importância para cada uma das demais. Note ainda, que para qualquer valor de $m \in [0, 1]$,

$$M = (1 - m) \cdot A + m \cdot S$$

$$M = (1 - m) \cdot \begin{pmatrix} a_{11} & \cdots & a_{1i} & \cdots & a_{1n} \\ a_{21} & \cdots & a_{2i} & \cdots & a_{2n} \\ a_{31} & \cdots & a_{3i} & \cdots & a_{3n} \\ \vdots & \vdots & \cdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{ni} & \cdots & a_{nn} \end{pmatrix} + m \cdot \begin{pmatrix} \frac{1}{n} & \frac{1}{n} & \cdots & \frac{1}{n} & \frac{1}{n} \\ \frac{1}{n} & \frac{1}{n} & \cdots & \frac{1}{n} & \frac{1}{n} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \frac{1}{n} & \frac{1}{n} & \cdots & \frac{1}{n} & \frac{1}{n} \\ \frac{1}{n} & \frac{1}{n} & \cdots & \frac{1}{n} & \frac{1}{n} \end{pmatrix}$$

$$= \begin{pmatrix} (1 - m)a_{11} + \frac{m}{n} & \cdots & (1 - m)a_{1i} + \frac{m}{n} & \cdots & (1 - m)a_{1n} + \frac{m}{n} \\ (1 - m)a_{21} + \frac{m}{n} & \cdots & (1 - m)a_{2i} + \frac{m}{n} & \cdots & (1 - m)a_{2n} + \frac{m}{n} \\ (1 - m)a_{31} + \frac{m}{n} & \cdots & (1 - m)a_{3i} + \frac{m}{n} & \cdots & (1 - m)a_{3n} + \frac{m}{n} \\ \vdots & \vdots & \cdots & \ddots & \vdots \\ (1 - m)a_{n1} + \frac{m}{n} & \cdots & (1 - m)a_{ni} + \frac{m}{n} & \cdots & (1 - m)a_{nn} + \frac{m}{n} \end{pmatrix}$$

a soma das entradas da coluna j , com $1 \leq j \leq n$, da matriz M é dada

por

$$\sum_{i=1}^n m_{ij} = \sum_{i=1}^n \left[(1-m)a_{ij} + \frac{m}{n} \right] = (1-m) \sum_{i=1}^n a_{ij} + \frac{m}{n} \sum_{i=1}^n 1.$$

Como A é coluna-estocástica,

$$\sum_{i=1}^n a_{ij} = 1$$

e

$$\sum_{i=1}^n 1 = n,$$

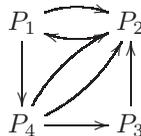
então,

$$\sum_{i=1}^n m_{ij} = (1-m) \times 1 + \frac{m}{n} \times n = 1,$$

o que indica que M também é coluna-estocástica. Além disso, para todo $m \in]0, 1[$, a matriz M possui todas as entradas positivas, tendo em vista que a matriz A possui todas as entradas a_{ij} não negativas ($a_{ij} \geq 0$) para todo i e j . Como a matriz S possui todas as entradas iguais a $s_{ij} = \frac{1}{n} > 0$, todas as entradas de M são da forma $m_{ij} = (1-m) \cdot a_{ij} + m \cdot s_{ij}$ e portanto, $m_{ij} > 0$ para todo i e j . Dessa forma, pelo Teorema 10, $\dim(V_1(M)) = 1$, gerando assim, apenas uma forma de classificar a importância das páginas da internet considerada.

2.4 APLICANDO O MÉTODO

Considere o seguinte grafo da internet de 4 páginas como segue.



Desse grafo, extraímos as equações que descrevem as relações entre as pontuações de importância das páginas. Usando a Equação 2.1 montamos o sistema

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

o qual pode ser escrito na forma $A \cdot x = x$:

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

e cuja solução geral é dada por

$$S = \left\{ \left(2\alpha, 2\alpha, \frac{\alpha}{2}, \alpha \right), \alpha \in \mathbb{R} \right\}.$$

Fazendo $x_1 + x_2 + x_3 + x_4 = 1$, obtemos a seguinte solução para o sistema: $x_1 \approx 0,364$, $x_2 \approx 0,364$, $x_3 \approx 0,09$ e $x_4 \approx 0,182$. Dessa forma, x_1 e x_2 são as páginas mais importantes e x_3 a página menos importante.

Agora podemos utilizar a matriz $M = (1 - m) \cdot A + m \cdot S$ no lugar de A para determinar a ordem de importância das páginas do modelo de internet anterior e verificar se há mudanças significativas nessa ordem. Para tal, considere a matriz

$$S = \begin{pmatrix} \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \end{pmatrix}.$$

Usaremos $m = 0,15$ (valor divulgado pelo Google segundo [2]). Dessa forma,

$$M = (1 - 0,15) \cdot \begin{pmatrix} 0 & 1 & 0 & 0 \\ \frac{1}{2} & 0 & 1 & \frac{1}{2} \\ 0 & 0 & 0 & \frac{1}{2} \\ \frac{1}{2} & 0 & 0 & 0 \end{pmatrix} + 0,15 \cdot \begin{pmatrix} \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \end{pmatrix}$$

$$M = \begin{pmatrix} 0,0375 & 0,8875 & 0,0375 & 0,0375 \\ 0,4625 & 0,0375 & 0,8875 & 0,4625 \\ 0,0375 & 0,0375 & 0,0375 & 0,4625 \\ 0,4625 & 0,0375 & 0,0375 & 0,0375 \end{pmatrix}$$

Resolvendo o sistema

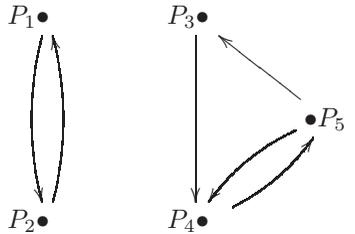
$$M \cdot x = x$$

encontramos a solução geral

$$x_1 = \frac{51853\alpha}{27713}, x_2 = \frac{108653\alpha}{55426}, x_3 = \frac{34907\alpha}{55426} \text{ e } x_4 = \alpha$$

para todo $\alpha \in \mathbb{R}$. Fazendo $x_1 + x_2 + x_3 + x_4 = 1$, obtemos a seguinte solução para o sistema: $x_1 \approx 0,343$, $x_2 \approx 0,359$, $x_3 \approx 0,115$ e $x_4 \approx 0,183$. Isso mostra que, mesmo utilizando a matriz M , as páginas 1 e 2 ainda são as mais importantes, porém, com uma leve vantagem para a página 2 e também que a página 3 ainda é a menos importante.

No caso da internet desconexa, do início da Seção 2.3, dada pelo grafo



em que

$$A = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{2} \\ 0 & 0 & 1 & 0 & \frac{1}{2} \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix}$$

é a matriz do grafo, tínhamos mais de uma maneira de classificar suas páginas, pois, $V_1(A) > 1$. No entanto, ao utilizarmos a matriz

$$M = (1 - m).A + m.S,$$

com

$$m = 0,15$$

e

$$S = \begin{pmatrix} \frac{1}{5} & \frac{1}{5} & \frac{1}{5} & \frac{1}{5} & \frac{1}{5} \\ \frac{1}{5} & \frac{1}{5} & \frac{1}{5} & \frac{1}{5} & \frac{1}{5} \\ \frac{1}{5} & \frac{1}{5} & \frac{1}{5} & \frac{1}{5} & \frac{1}{5} \\ \frac{1}{5} & \frac{1}{5} & \frac{1}{5} & \frac{1}{5} & \frac{1}{5} \\ \frac{1}{5} & \frac{1}{5} & \frac{1}{5} & \frac{1}{5} & \frac{1}{5} \end{pmatrix},$$

encontramos infinitos autovetores associados ao autovalor $\lambda = 1$ para a matriz M , porém todos são múltiplos uns dos outros. Além disso, temos um único autovetor cuja soma das entradas vale 1, e portanto, podemos utilizá-lo para classificar a importância das páginas como segue:

$$M = (1 - 0,15) \cdot \begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{2} \\ 0 & 0 & 1 & 0 & \frac{1}{2} \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix} + 0,15 \cdot \begin{pmatrix} \frac{1}{5} & \frac{1}{5} & \frac{1}{5} & \frac{1}{5} & \frac{1}{5} \\ \frac{1}{5} & \frac{1}{5} & \frac{1}{5} & \frac{1}{5} & \frac{1}{5} \\ \frac{1}{5} & \frac{1}{5} & \frac{1}{5} & \frac{1}{5} & \frac{1}{5} \\ \frac{1}{5} & \frac{1}{5} & \frac{1}{5} & \frac{1}{5} & \frac{1}{5} \\ \frac{1}{5} & \frac{1}{5} & \frac{1}{5} & \frac{1}{5} & \frac{1}{5} \end{pmatrix}$$

$$M = \begin{pmatrix} 0,03 & 0,88 & 0,03 & 0,03 & 0,03 \\ 0,88 & 0,03 & 0,03 & 0,03 & 0,03 \\ 0,03 & 0,03 & 0,03 & 0,03 & 0,455 \\ 0,03 & 0,03 & 0,88 & 0,03 & 0,455 \\ 0,03 & 0,03 & 0,03 & 0,88 & 0,03 \end{pmatrix}$$

Resolvendo o sistema

$$M \cdot x = x$$

encontramos a solução geral

$$x_1 = \frac{1769\alpha}{2058}, x_2 = \frac{1769\alpha}{2058}, x_3 = \frac{190\alpha}{343}, x_4 = \frac{703\alpha}{686} \text{ e } x_5 = \alpha$$

para todo $\alpha \in \mathbb{R}$. Fazendo $x_1 + x_2 + x_3 + x_4 + x_5 = 1$, obtemos a seguinte solução para o sistema: $x_1 = 0,2$, $x_2 = 0,2$, $x_3 \approx 0,13$, $x_4 \approx 0,238$ e $x_5 \approx 0,232$. Dessa forma, x_4 é a página mais importante e x_3 a página menos importante.

3 ROTEIRO PARA ABORDAR O TEMA NO ENSINO MÉDIO

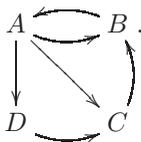
Neste capítulo apresentaremos uma proposta de abordagem, de forma intuitiva, do algoritmo PageRank para turmas de ensino médio. No entanto, necessitamos que os alunos tenham alguns conhecimentos prévios:

1. Matrizes: operações e propriedades.
2. Sistemas Lineares: resolução e classificação através do escalonamento.

3.1 ROTEIRO

Durante o estudo de sistemas lineares, o aprendizado desse assunto pode ser motivado utilizando-se como artifício o algoritmo PageRank. Inicialmente, podemos explanar através de exemplos a ideia intuitiva de se atribuir uma pontuação de importância para cada página de uma internet como feito na Seção 1.1. Depois podemos propor aos alunos que realizem as seguintes atividades:

Questão 1. Considere a internet de 4 páginas representada pelo grafo

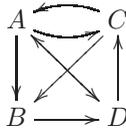


Usando seus conhecimentos sobre o algoritmo PageRank, faça o que se pede:

1. Monte o sistema de equações lineares que indica as relações de importância entre as páginas.
2. Identifique a matriz do grafo A associada a essa internet.
3. Que valor obtemos ao somar todos os elementos de uma coluna da matriz do grafo associado a essa internet?
4. Escalone o sistema obtido.

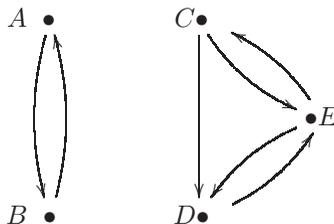
5. Classifique o sistema escalonado quanto à quantidade de soluções: uma única, infinitas ou nenhuma solução.
6. Dê o conjunto solução do sistema.
7. Utilize o conjunto solução do sistema obtido para classificar qual das páginas é a mais importante.
8. Construa um novo grafo acrescentando um link saindo da página B para a página D . Resolva novamente os itens anteriores e determine a nova classificação para as páginas associadas a esse novo grafo.
9. Caso construíssemos um novo grafo acrescentando 3 links saindo da página B para a página D , em vez de apenas 1, como no item anterior, teríamos uma classificação diferente para as páginas dessa internet? Justifique sua resposta.

Questão 2. O Sr. Farias, dono do site correspondente a página C , representada no grafo



deseja alavancar a pontuação de importância de sua página calculada através do algoritmo PageRank de modo a superar a página A que é a página mais importante. Para isso, o Sr. Farias cria uma página E , a qual possui link para a página C da mesma forma que C possui link para E . Será que o Sr. Farias obterá êxito? Justifique sua resposta com suas palavras de forma lógica e/ou com os cálculos necessários.

Questão 3. Considere o grafo a seguir, correspondente a uma internet desconectada.



1. Monte o sistema de equações lineares que indica as relações de importância entre as páginas.

2. Identifique a matriz do grafo A associada a essa internet.
3. Escalone o sistema obtido.
4. Classifique o sistema escalonado quanto à quantidade de soluções: uma única, infinitas ou nenhuma solução.
5. Dê o conjunto solução do sistema.
6. Utilize o conjunto solução do sistema obtido para classificar qual das páginas é a mais importante.
7. Existe mais de uma forma para classificar a importância das páginas dessa internet?
8. Use a matriz $M = (1 - m) \cdot A + m \cdot S$, com $m = 0,15$, no lugar de A para determinar a ordem de importância das páginas dessa internet.
9. O que acontece com a matriz $M = (1 - m) \cdot A + m \cdot S$, ao variarmos m de 0 até 1?
10. O que podemos dizer sobre M quando $m = 0$? E quando $m = 1$?

CONCLUSÃO

Ao desenvolver esse trabalho, mostramos uma parte da matemática envolvida nos mecanismos de busca utilizados pelo Google no algoritmo PageRank. Nos interessamos pelo estudo desse algoritmo por ser um dos principais responsáveis pela qualidade das respostas dadas pelo Google ao efetuarmos uma pesquisa na internet. Como a internet está cada vez mais presente no dia a dia das pessoas, principalmente dos jovens, verificamos que os conceitos matemáticos envolvidos nesse estudo poderiam ser utilizados como motivação aos alunos do ensino médio, no aprendizado de assuntos como: matrizes, determinantes e sistemas lineares.

Desenvolvemos os capítulos desse trabalho em níveis de complexidade e aprofundamento diferentes. No Capítulo 1 nosso objetivo foi simplesmente informar o leitor, numa linguagem acessível, de fácil compreensão e um tanto quanto lúdica, sobre os passos dados pelo Google ao efetuarmos uma pesquisa na internet. Já no segundo capítulo, apresentamos e utilizamos os conceitos matemáticos envolvidos no algoritmo PageRank, com toda a formalidade devida, estipulando condições e modelando algumas das situações a que esse algoritmo é submetido. Por fim, no último capítulo, mostramos uma forma de abordar, de modo intuitivo, os processos descritos e formalizados, no primeiro e no segundo capítulo, respectivamente, através de um roteiro com questões, elaborado de modo acessível aos alunos de ensino médio.

Aos leitores que quiserem conhecer um pouco mais sobre outros modelos que não foram mencionados neste trabalho, indicamos a leitura completa de [2] e [6].

REFERÊNCIAS

- 1 ANTON, Howard. **Álgebra linear:** com aplicações. 8. ed. Porto Alegre: Bookman, 2001-2002.
- 2 BRYAN, Kurt e LEISE, Tanya. **The \$25,000,000,000* eigenvector:** The Linear Álgebra behind Google. Disponível em: <<https://www.rose-hulman.edu/~bryan/googleFinalVersionFixed.pdf>>. Acesso em: 28 de junho de 2015.
- 3 GOOGLE. **Nossa história a fundo.** Disponível em: <<https://www.google.com.br/about/company/history>>. Acesso em: 28 de junho de 2015.
- 4 HEFEZ, Abramo e FERNANDEZ, Cecília S. **Introdução à álgebra linear.** 1. ed. Rio de Janeiro: SBM, 2012.
- 5 LAY, David C. **Álgebra linear e suas aplicações.** 2. ed. Rio de Janeiro: Livros Técnicos e Científicos, 1999.
- 6 PAGE, Lawrence e BRIN, Sergey e MOTWANI, Rajeev e WINOGRAD, Terry. **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: 06 de outubro de 2015.