TECHNICKÁ UNIVERZITA V LIBERCI
Fakulta přírodovědně-humanitní
a pedagogická

Google PageRank: Relevance webových stránek a problém vlastních čísel

Bakalářská práce

Studijní program: B1101 – Matematika
Studijní obory: 7504R015 – Matematika se zaměřením na vzdělávání
7507R036 – Anglický jazyk se zaměřením na vzdělávání

Autor práce: **Markéta Hejlová**
Vedoucí práce: Martin Plešinger

Liberec 2015

TECHNICAL UNIVERSITY OF LIBEREC
Faculty of Science, Humanities
and Education

Google's PageRank: Ranking of web pages and the eigenvalue problem

Bachelor thesis

Study programme: B1101 – Mathematics
Study branches: 7504R015 – Mathematics for Education
7507R036 – English for Education

Author: **Markéta Hejlová**
Supervisor: Martin Plešinger

Liberec 2015

Tento list nahraďte
originálem zadání.

Prohlášení

Byla jsem seznámena s tím, že na mou bakalářskou práci se plně vztahuje zákon č. 121/2000 Sb., o právu autorském, zejména § 60 – školní dílo.

Beru na vědomí, že Technická univerzita v Liberci (TUL) nezasahuje do mých autorských práv užitím mé bakalářské práce pro vnitřní potřebu TUL.

Užiji-li bakalářskou práci nebo poskytnu-li licenci k jejímu využití, jsem si vědoma povinnosti informovat o této skutečnosti TUL; v tomto případě má TUL právo ode mne požadovat úhradu nákladů, které vynaložila na vytvoření díla, až do jejich skutečné výše.

Bakalářskou práci jsem vypracovala samostatně s použitím uvedené literatury a na základě konzultací s vedoucím mé bakalářské práce a konzultantem.

Současně čestně prohlašuji, že tištěná verze práce se shoduje s elektronickou verzí, vloženou do IS STAG.

Datum:

Podpis:

Anotace

Bakalářská práce se zaměřuje na spektrální vlastnosti některých speciálních matic a jejich souvislosti s Markovovými řetězci a teorií orientovaných grafů. Jedná se zejména o nezáporné, kladné, (sub)stochastické, (ne)rozložitelné a (im)primitivní matice. Výklad je v průběhu celého textu doprovázen praktickým příkladem, čímž je práce motivována. Tento příklad se týká nalezení tzv. PageRanků jednotlivých webových stránek na internetu. PageRank lze reprezentovat jako míru důležitosti webové stránky. Koncept PageRanku využívá například internetový vyhledávač Google. Práce se zabývá způsobem, jak lze PageRank definovat a odhaluje některé obtížnosti, které se objevují v samotné definici a při výpočtu PageRanků. Tyto obtíže lze ovšem snadno obejít s využitím vlastností příslušných matic.

Klíčová slova:

nezáporné matice; kladné matice; (sub)stochastické matice; (ne)rozložitelné matice; (im)primitivní matice; PageRank vektor; googlovská matice; Markovovy řetězce; orientované grafy; mocninná metoda; Perronovo vlastní číslo; Perronův vlastní vektor

Abstract

The bachelor thesis is focused on the spectral properties of some particular matrices and their connections to Markov chains and the theory of digraphs. In particular we concentrate on nonnegative, positive, (sub)stochastic, (ir)reducible, and (im)primitive matrices. The theory is continuously demonstrated on a practical example, which works also as a motivation. This example is to determine so-called PageRanks of individual internet webpages. This PageRank can be interpreted as a measure of imparotance of the given webpage. The PageRank concept is employed for example by Google web search engine. This thesis analyzes the PageRank definition and reveals some difficulties that appear in the definition as well as in the computation of the PageRank. Hovewer, these difficulties can be easily avoided by using properties of the abovementioned matrices.

Key words:

nonnegative matrices; positive matrices; (sub)stochastic matrices; (ir)reducible matrices; (im)primitive matrices; PageRank vector; Google matrix; Markov chain; digraphs; power method; Perron eigenvalue; Perron eigenvector

Poděkování

Ráda bych poděkovala svému vedoucímu bakalářské práce Martinu Plešingerovi za jeho trpělivou pomoc, odborné vedení a ochotu při zpracování této práce.

Obsah

Seznam obrázků

Seznam tabulek

Použité značení

Matice a vektory

Značení	**Význam**
$A \in \mathbb{R}^{m \times n}$	reálná matice s rozměry m krát n a s prvky $a_{j,k}$
$A \in \mathbb{C}^{m \times n}$	komplexní matice s rozměry m krát n a s prvky $a_{j,k}$
$\mathbf{i}$	imaginární jednotka, $\mathbf{i}^2 = -1$
$\overline{A}$	matice komplexně sdružená s prvky $\overline{a}_{j,k}$
A^T	transponovaná matice
A^H	matice transponovaná, komplexně sdružená $A^H = \overline{A}^T$
A^{-1}	inverzní matice ke čtvercové regulární matici A
A^{-T}	matice $(A^{-1})^T = (A^T)^{-1}$, kde A je čtvercové regulární
A^{-H}	matice $(A^{-1})^H = (A^H)^{-1}$, kde A je čtvercové regulární
$\mathrm{sp}(A)$	spektrum čtvercové matice A (množina všech vlastních čísel)
$\varrho(A)$	spektrální poloměr čtvercové matice A
$\mathscr{K}_\odot(A)$	spektrální kružnice matice A $\mathscr{K}_\odot(A) = \{\varrho(A)(\cos(\varphi) + \mathbf{i}\sin(\varphi)) \,:\, 0 \leq \varphi \leq 2\pi\} \subset \mathbb{C}$
$\det(A)$	determinant čtvercové matice A
$A > 0,\ A \geq 0$	reálná matice A, jejíž prvky splňují $a_{j,k} > 0$, $a_{j,k} \geq 0$ tj. kladná, resp. nezáporná matice
$A < 0,\ A \leq 0$	reálná matice A, jejíž prvky splňují $a_{j,k} < 0$, $a_{j,k} \leq 0$
$I \in \mathbb{R}^{n \times n}$	jednotková matice
e_j	j-tý sloupec jednotkové matice I
e	vektor $e = [1, \ldots, 1]^T$
$\lvert x \rvert \in \mathbb{R}^n$	vektor absolutních hodnot $x = [\lvert\xi_1\rvert, \ldots, \lvert\xi_n\rvert]^T$
$\lvert A \rvert \in \mathbb{R}^{m \times n}$	matice absolutních hodnot s prvky $\lvert a_{j,k}\rvert$
$\lVert x \rVert_p$	p-norma vektoru $(\sum_j \lvert\xi_j\rvert^p)^{1/p}$, $p = 1, 2, \ldots$
$\lVert x \rVert_\infty$	maximová norma vektoru $\max_j \lvert\xi_j\rvert$
$\lVert A \rVert_p$	p-norma matice $\max_{\lVert x \rVert_p = 1} \lVert Ax \rVert_p$
$\lVert A \rVert_\infty$	∞-norma matice $\max_j(\sum_k \lvert a_{j,k}\rvert)$

Speciální nezáporné matice a nezáporné vektory

Značení	Význam
$\pi \in \mathbb{R}^n$	levý Perronův vlastní vektor stochastické matice zvaný také PageRank vektor, jedná-li se o (opravenou) matici hyperlinků, resp. googlovskou matici
$d \in \{0,1\}^n$	dangling node vektor
$S \in \mathbb{R}^{n \times n}$	stochastická matice
$H \in \mathbb{R}^{n \times n}$	(substochastická) matice hyperlinků
$D = \frac{1}{n} de^T \in \mathbb{R}^{n \times n}$	(substochastická) oprava matice hyperlinků
$\widetilde{H} = H + D$	(stochastická) opravená matice hyperlinků
$E = \frac{1}{n} ee^T \in \mathbb{R}^{n \times n}$	(stochastická) teleportační matice
$G = \alpha\widetilde{H} + (1-\alpha)E$	(stochastická) googlovská matice, $\alpha \in (0,1)$

Grafy a množiny

Značení	Význam
$\vec{G} = (\mathscr{V}, \mathscr{E})$	orientovaný graf s vrcholy $v_j \in \mathscr{V}$ a hranami $e_{j,k} \in \mathscr{E}$
$\vec{G}(A)$	orientovaný graf matice A
$\mathscr{M} \times \mathscr{W}$	kartézský součin množin $\mathscr{M}$ a $\mathscr{W}$
$\lvert\mathscr{M}\rvert$	počet prvků konečné množiny $\mathscr{M}$

Ostatní

Značení	Význam
$\mathtt{P}_j$	j-tá webová stránka
$\lvert\mathtt{P}_j\rvert$	počet odkazů na stránce $\mathtt{P}_j$
$\mathscr{B}_{\mathtt{P}_j}$	množina všech stránek odkazujících na $\mathtt{P}_j$
$r(\mathtt{P}_j)$	PageRank stránky $\mathtt{P}_j$
$\mathscr{D}$	množina všech stránek, které neobsahují žádný odkaz (dangling nodes)

Úvod

Tato práce pojednává o pojmu *PageRank*, který byl v roce 1996 navržený *Larry Pagem* a *Sergeyem Brinem* na universitě ve Stanfordu jako součást výzkumného projektu. Před tímto datem se touto problematikou zabývalo více vědců, z nichž bychom zmínili Gabriela Pinskiho a Francise Narina, kteří v roce 1976 poprvé definovali problém propojení jako problém vlastních čísel. PageRank je hlavním nástrojem, který dostává Google na tak dobrou pozici mezi internetovými vyhledávači a sama společnost Google tvrdí, že je jejím srdcem. Můžeme nyní říci, že se jedná o algoritmus, který pomáhá v řazení vyhledávaných stránek podle důležitosti a který určuje to, v jakém pořadí se budou uživatelům zobrazovat při zadání určitého pojmu do vyhledávače. Přesnější definicí se budeme zabývat později. Celý text čerpá zejména z knihy [17], která objasňuje problematiku PageRanku. Další informace k tomuto tématu je možné nalézt v [2] nebo v [24]. V matemických částech práce je čerpáno především z [12] a [14].

Text je rozdělen do dvou hlavních částí, Existence PageRank vektoru a Výpočet PageRank vektoru. Ovšem před těmito částmi se v kapitole 1 zabýváme některými základními pojmy z teorie vlastních čísel matic a z teorie grafů, se kterými dále pracujeme, a proto se s nimi čtenář musí seznámit hned na začátku. Také je zde uvedena definice PageRanku, ke které se pak vracíme v závěru práce. Po první kapitole následuje první část, která obsahuje 3 kapitoly a která se zabývá existencí PageRank vektoru. Kapitola 2 definuje nezáporné a stochastické matice a jejich specifika. V kapitole 3 hovoříme o nerozložitelnosti matic a v kapitole 4 řešíme problém, že matice hyperlinků, se kterou pracujeme v průběhu celé práce, obecně není ani stochastická ani nerozložitelná. Tyto dvě vlastnosti, jak zjistíme, jsou pro existenci PageRanku nepostradatelné a jsou spojené s problematikou uživatele internetu, který neprochází internetem tak, jak by nám vyhovovalo. Na konci první části se ale dozvídáme, že problémy se stochasticitou a (ne)rozložitelností, lze vyřešit, a tudíž víme, že PageRank existuje.

V druhé části výkladu se věnujeme již samotnému výpočtu PageRank vektoru. Nejprve v kapitole 5 představujeme mocninnou metodu, která je nástrojem tohoto výpočtu, a pak se v kapitole 6 vracíme k definici PageRanku a nacházíme způsob, jak PageRank vektor neboli také Perronův vlastní vektor spočítat. V této poslední kapitole také uvádíme (im)primitivní matice, které jsou pro použití mocninné metody k výpočtu hledaného vektoru nepostradatelné. Zjistíme, že pro imprimitivní matice tato metoda nefunguje, a proto budeme muset ověřit, že googlovská matice, pro kterou PageRank počítáme, je primitivní. V závěru budeme schopni ověřit (im)primitivitu, upravit definici PageRank vektoru a nalezneme tak způsob, jak ho spočítat. Konkrétní příklad výpočtu PageRanku je možné nalézt v příloze A.

Celý text má za úkol seznámit se s partiemi lineární algebry, teorie grafů a Markovových řetězců a jejich vzájemným propojením v rozsahu, který překračuje běžně vykládanou látku základních kurzů na příkladu praktické úlohy a tím odhalit způsob, kterým nejen Google pracuje. PageRank je také využíván v bibliometrii, v sociálních a informačních sítích, v systémech silničních sítí, biologii, chemii či fyzice. Znalost PageRanku mimo jiné může pomoci majitelům různých webových stránek vylepšovat svou pozici na internetu tak, aby se jejich stránky ukazovaly co nejvýše.

1 Základní pojmy

Hodnocení důležitosti nebo významnosti webových stránek je značně obtížné. Ukazuje se, že vhodným přístupem při takovém hodnocení je podívat se, které (jiné) stránky na hodnocenou stránku odkazují, viz [17], [2, kapitola 7.2]. Odkazují-li na ni stránky důležité, můžeme i stránku hodnocenou považovat za jistým způsobem důležitou. Tedy, stručně řečeno, každá webová stránka je důležitá, pokud je na ni odkázáno jinou důležitou webovou stránkou. To však vede k „zacyklené“ definici, abychom byli schopni hodnotit jednu stránku, musíme být nejprve schopni zhodnotit stránky jiné. Takto definovaná míra důležitosti stránky se nazývá *PageRank*[1]. Přesná definice je sepsána v sekci 1.4. My nejprve začneme základními pojmy z teorie vlastních čísel matic, které se nám budou hodit.

1.1 Vlastní čísla a vlastní vektory

Klíčovým nástrojem v této práci budou tzv. *vlastní čísla* a *vlastní vektory*, což známe ze základního kurzu lineární algebry. Pro úplnost připomeneme defnici.

Definice 1 (Vlastní číslo, vlastní vektor, spektrum, spektrální poloměr)**.** *Nechť je obecná komplexní čtvercová matice $A \in \mathbb{C}^{n\times n}$, $\lambda \in \mathbb{C}$ je skalár a x, $y \in \mathbb{C}^n$, $x \neq 0$, $y \neq 0$, nenulové vektory tak, že platí*

$$Ax = x\lambda, \qquad y^H A = \lambda y^H, \tag{1.1}$$

kde $y^H = \overline{y}^T$. Pak skalár λ nazýváme vlastním (nebo také charakteristickým) číslem matice A, vektor x vlastním, resp. pravým vlastním (nebo také charakteristickým) a vektor y levým vlastním vektorem matice A. Uspořádané dvojici (λ, x) se říká vlastní pár matice A, obdobně (λ, y) je vlastní pár matice A^H. Množina všech vlastních čísel matice A se nazývá spektrum matice A a značí se

$$\mathrm{sp}(A) = \{\lambda \in \mathbb{C} : \exists x \in \mathbb{C}^n,\ x \neq 0,\ Ax = x\lambda\}.$$

Absolutní hodnota vlastního čísla nejvíce vzdáleného od nuly se nazývá spektrální poloměr matice A a značí se

$$\varrho(A) = \max_{\lambda \in \mathrm{sp}(A)} |\lambda|. \tag{1.2}$$

[1] Poznamenejme, že slovo *PageRank* je složeninou slov *Page* a *rank*, kde druhé ze slov můžeme přeložit právě např. jako *hodnost* nebo *hodnocení*, ve smyslu *pozice v žebříčku*. Slovo *Page* však neodkazuje ke slovu *(webová) stránka*, z angl. *(web) page*, jak bychom se mohli mylně domnívat, ale k *Larry Pageovi*, autorovi této koncepce hodnocení stránek, viz [17, str. 32, poznámka 1].

Je to tedy poloměr nejmenšího kruhu v komplexní rovině, který má střed v počátku a obsahuje celé spektrum matice A. Hranici tohoto kruhu

$$\mathscr{K}_{\odot}(A) \equiv \{\varrho(A)\exp(\mathbf{i}\,\varphi) = \varrho(A)(\cos(\varphi) + \mathbf{i}\sin(\varphi)) \,:\, 0 \leq \varphi < 2\pi\} \subset \mathbb{C}$$

budeme nazývat spektrální kružnice matice A.

Připomeňme, že vlastní čísla jsou kořeny tzv. *charakteristického polynomu* $\chi(\lambda) = 0$, kde $\chi(\lambda) \equiv \det(I\lambda - A)$ a $\det(\cdot)$ značí *determinant* matice, viz [11, kapitola 1].

Jedním z úkolů v této práci bude zjistit, zda je určité vlastní číslo jednoduché nebo vícenásobné. K tomu nám poslouží následují tzv. *Shurova pomocná věta* (viz [12, věta 1.49]; nezaměňovat s tzv. *Schurovou větou*, viz [11, kapitola 2]), na kterou se budeme později odkazovat.

Věta 1 (Schurova pomocná věta)**.** *Nechť $A \in \mathbb{C}^{n\times n}$ je čtvercová matice a λ je její vlastní číslo. Pak λ je jednoduché vlastní číslo tehdy a jen tehdy, když existuje jediný lineárně nezávislý pravý vlastní x, (tedy také) jediný lineárně nezávislý levý vlastní vektor y a zároveň platí $y^H x \neq 0$.*

Důkaz. Uvažujme matici A s vlastním číslem λ. Uvažujme dále matici B, která je matici A *podobná*, tj. existuje regulární matice W tak, že platí $B = WAW^{-1}$ (poznamenejme, že podobnost je ekvivalence na množině čtvercových matic daného rozměru; viz [11, kapitola 1.8]). Pokud existuje jediný lineárně nezávislý pravý vlastní vektor x tak, že $Ax = x\lambda$, a (tedy také) jediný lineárně nezávislý levý vlastní vektor y tak, že $y^H A = \lambda y^H$, potom vektor Wx je jediný lineárně nezávislý pravý vlastní vektor matice B, tj.

$$B(Wx) = (WAW^{-1})(Wx) = WAx = (Wx)\lambda,$$

a vektor $W^{-H}y \equiv (W^{-1})^H y$ je jediný lineárně nezávislý levý vlastní vektor matice B, tj.

$$(W^{-H}y)^H B = (y^H W^{-1})(WAW^{-1}) = y^H AW^{-1} = \lambda y^H W^{-1} = \lambda(W^{-H}y)^H.$$

Navíc je pro vektory Wx a $W^{-H}y$ splněno $(W^{-H}y)^H(Wx) \neq 0$ tehdy a jen tehdy, když $y^H x \neq 0$, přesněji řečeno

$$(W^{-H}y)^H(Wx) = (y^H W^{-1})(Wx) = y^H(W^{-1}W)x = y^H x.$$

Místo s maticí A tedy můžeme v důkazu pracovat s maticí B, resp. s libovolným reprezentantem třídy matic podobných matici A. Nejvýhodnější bude pracovat s tzv. *Jordanovým kanonickým tvarem* matice A, viz [11, kapitola 1.8].

Nyní přejdeme k vlastnímu důkazu. Nejprve dokážeme implikaci: když x, y je jediný lineárně nezávislý pravý, resp. levý vlastní vektor a $y^H x \neq 0$, pak λ je jednoduché vlastní číslo.

K libovolnému Jordanovu bloku $J_m(\lambda) \in \mathbb{C}^{m\times m}$ existuje jediný lineárně nezávislý levý vlastní vektor, např. $[1, 0, \ldots, 0, 0]^T \in \mathbb{C}^m$, a jediný lineárně nezávislý pravý vlastní vektor, např. $[0, 0, \ldots, 0, 1]^T \in \mathbb{C}^m$, a jejich skalární součin je nenulový, když $m = 1$, resp. je nulový, když $m > 1$. Navíc pravé (resp. levé) vlastní vektory matice v Jordanově kanonickém tvaru odpovídající různým Jordanovým blokům (tj. vlastní vektory Jordanových bloků doplněné nulami) jsou vždy lineárně nezávislé.

Pokud x (resp. y) je jediný lineárně nezávislý vlastní vektor matice A odpovídající vlastnímu číslu λ, pak v Joradnově kanonickém tvaru matice A existuje právě jeden Jordanův blok $J_m(\lambda)$ odpovídající číslu λ. Pokud navíc $y^H x \neq 0$, pak musí být $m = 1$, a tedy vlastní číslo λ má násobnost rovnou jedné.

Nyní dokážeme opačnou implikaci. Nechť λ je jednoduché vlastní číslo matice A. Pak v Jordanově kanonickém tvaru matice A odpovídá vlastnímu číslu λ jediný Jordanův blok, který je navíc velikosti jedna. Pak také existuje jediný lineárně nezávislý pravý (resp. levý) vlastní vektor odpovídající tomuto vlastnímu číslu. Z Jordanova kanonického tvaru je vidět, že jejich skalární součin je nenulový. □

1.2 Normy vektorů a matic

Pro výklad bude potřeba zavést pojmy normy vektoru a normy matice. Zejména proto, abychom mohli dát do vzájemného vztahu spektrální poloměr a normu čtvercové matice.

Definice 2 (p-norma vektoru)**.** *Nechť* $x = [\xi_1, \xi_2, \ldots, \xi_n]^T \in \mathbb{C}^n$*. Číslo*

$$\|x\|_p \equiv \left(\sum_{i=1}^{n} |\xi_i|^p \right)^{1/p}$$

nazýváme vektorová p-norma nebo p-norma vektoru, $p = 1, 2, \ldots$

Zejména důležitý bude její limitní případ.

Definice 3 (maximová norma vektoru)**.** *Nechť* $x = [\xi_1, \xi_2, \ldots, \xi_n]^T \in \mathbb{C}^n$*. Číslo*

$$\|x\|_\infty \equiv \lim_{p \longrightarrow \infty} \|x\|_p = \max_j |\xi_j|$$

nazýváme maximovou normou vektoru.

Kromě vektorových norem budeme potřebovat zavést i normy matic. Normu matice lze zavést mnoha způsoby. My budeme pracovat s maticovou normou generovanou normou vektorovou. Konkrétně p-normou vektoru.

Definice 4 (p-norma matice (generovaná norma))**.** *Nechť* $A \in \mathbb{C}^{m\times n}$*. Číslo*

$$\|A\|_p \equiv \max_{x \neq 0} \frac{\|Ax\|_p}{\|x\|_p} = \max_{x \neq 0} \left\| A \frac{x}{\|x\|_p} \right\|_p = \max_{\|x\|_p = 1} \|Ax\|_p \tag{1.3}$$

nazýváme maticová p-norma, případně norma generovaná vektorovou normou $\|\cdot\|_p$, $p = 1, 2, \ldots$

Opět pro nás bude zajímavý její limitní případ.

Definice 5 (∞-norma matice (generovaná norma))**.** *Nechť* $A \in \mathbb{C}^{m \times n}$*. Číslo*

$$\|A\|_\infty \equiv \lim_{p \longrightarrow \infty} \|A\|_p = \max_{\|x\|_\infty = 1} \|Ax\|_\infty \tag{1.4}$$

nazýváme maticová ∞*-norma, případně norma generovaná vektorovou normou* $\|\cdot\|_\infty$.

Takto zadefinované maticové normy nelze obecně snadno vypočítat, až na některé případy. Jedná se o normy $\|\cdot\|_1$, $\|\cdot\|_2$ a limitní případ $\|\cdot\|_\infty$. Následující lemma, nám dá návod, jak vypočítat ∞-normu matice, kterou budeme později potřebovat.

Lemma 1. *Nechť* $A \in \mathbb{C}^{m \times n}$*. Pak platí*

$$\|A\|_\infty = \max_{\|x\|_\infty = 1} \|Ax\|_\infty = \max_j \sum_{k=1}^{n} |a_{j,k}|. \tag{1.5}$$

Důkaz je elementární, viz např. [12, str. 167]. Dále bude potřebné následující lemma (viz [11, str. 21]).

Lemma 2. *Je-li* A *čtvercová matice a* $\|\cdot\|$ *libovolná maticová norma generovaná vektorovou normou* $\|\cdot\|$*, pak platí*

$$\|A\| \geq \varrho(A) \tag{1.6}$$

a také $\|I\| = 1$.

Důkaz. Pro každé vlastní číslo λ matice A existuje alespoň jeden vlastní vektor x takový, že $Ax = x\lambda$. Pro tento vlastní pár platí

$$\|A\|\|x\| \geq \|Ax\| = \|\lambda x\| = |\lambda|\|x\|.$$

Nerovnost na začátku vyplývá přímo z definice generované normy (1.3). Protože $x \neq 0$, tj. $\|x\| \neq 0$, dostáváme ihned $\|A\| \geq |\lambda|$ pro libovolné vlastní číslo λ. Z toho plyne $\|A\| \geq \varrho(A)$. □

Poznamenejme, že poslední rovnost vyplývá z tzv. *homogenity* norem (tj. pro libovolný skalár α a vektor v platí $\|\alpha v\| = |\alpha|\|v\|$), která je jednou z vlastností definujícíh pojem norma, viz [11, sekce 1.2, definice 1.10].

1.3 Základní pojmy teorie grafů

Při výkladu budeme potřebovat i některé pojmy z *teorie grafů*. Některé základní zmíníme v následující definici.

Definice 6 (Orientovaný graf, vrchol, hrana, cesta, komponenta)**.** *Uspořádanou dvojici*

$$\vec{G} = (\mathscr{V}, \mathscr{E}), \qquad \textit{kde} \qquad \mathscr{V} = \{v_j\} \quad \textit{a} \quad \mathscr{E} = \{e_{j,k} \equiv (v_j, v_k)\} \subseteq \mathscr{V} \times \mathscr{V},$$

nazýváme orientovaným grafem. Prvky v_j množiny $\mathscr{V}$ nazýváme vrcholy grafu a prvky $e_{j,k} \equiv (v_j, v_k)$ množiny $\mathscr{E}$ (tj. uspořádané dvojice vrcholů) nazýváme hrany. Říkáme, že $e_{j,k}$ je hrana vedoucí z vrcholu v_j do vrcholu v_k.

Posloupnost hran

$$e_{j,j_1}\,,\quad e_{j_1,j_2}\,,\quad e_{j_2,j_3}\,,\quad \ldots\,,\quad e_{j_{\ell-2},j_{\ell-1}}\,,\quad e_{j_{\ell-1},k}\ \in \mathscr{E}$$

vedoucí z vrcholu v_j postupně do v_{j_1}, do v_{j_2}, atd. až konečně do $v_{j_{\ell-1}}$ a v_k se nazývá cesta (případně orientovaná cesta) z vrcholu v_j do vrcholu v_k délky ℓ.

Libovolnou podmnožinu množiny $\mathscr{V}$ nazýváme komponenta grafu $\vec{G}$. Speciálně, nechť $\mathscr{V} = \mathscr{V}_1 \cup \mathscr{V}_2$ a nechť pro každou hranu $e_{j,k} \equiv (v_j, v_k) \in \mathscr{E}$ platí buď $v_j, v_k \in \mathscr{V}_1$, nebo $v_j, v_k \in \mathscr{V}_2$. Pak množiny $\mathscr{V}_1$ a $\mathscr{V}_2$ nazýváme nezávislé komponenty grafu $\vec{G}$.

Komponenta $\mathscr{W} \subseteq \mathscr{V}$ taková, že mezi libovolnými dvěma vrcholy v_j, $v_k \in \mathscr{W}$ existuje cesta procházející pouze přes vrcholy komponenty $\mathscr{W}$, se nazývá silně souvislá komponenta grafu. Pokud $\mathscr{W} = \mathscr{V}$, hovoříme také o silně souvislém grafu.

Pro upřesnění, jsou-li $\mathscr{V}_1$ a $\mathscr{V}_2$ *nezávislé komponenty*, pak v množině $\mathscr{E}$ neexistuje žádná hrana taková, která by vedla z některého vrcholu množiny $\mathscr{V}_1$ do kteréhokoliv vrcholu množiny $\mathscr{V}_2$, ani žádná hrana taková, která by vedla z některého vrcholu množiny $\mathscr{V}_2$ do kteréhokoliv vrcholu množiny $\mathscr{V}_1$. Poznamenejme ještě, že mezi libovolnými vrcholy silně *souvislé (konečné) komponenty* $\mathscr{W}$ (budeme pracovat pouze s grafy, které mají konečný počet vrcholů a hran) vždy existuje cesta délky ℓ, kde $\ell < |\mathscr{W}|$ (přičemž $|\mathscr{M}|$ značí počet prvků konečné množiny $\mathscr{M}$). Pro bližší seznámení se s pojmy teorie grafů doporučujeme např. [16] nebo české učebnice [9], [10], [20], [21], [22], případně rozsáhlejší text [19]. Speciálně budeme pracovat s tzv. *grafem (čtvercové) matice*, viz následující definici.

Definice 7 (Graf matice)**.** *Nechť $A \in \mathbb{C}^{n\times n}$ je obecná čtvercová matice. Orientovaný graf $\vec{G} = (\mathscr{V}, \mathscr{E})$, kde*

$$\mathscr{V} = \{1, 2, 3, \ldots, n\} \qquad \textit{a} \qquad (j,k) \in \mathscr{E} \iff a_{j,k} \neq 0,$$

nazveme grafem matice A. Budeme ho značit $\vec{G}(A)$.

Graf čtvercové matice A obsahuje hranu z vrcholu j do vrcholu k tehdy a jen tehdy, když je prvek $a_{j,k}$ nenulový. Poznamenejme, že graf matice lze zavést i jiným způsobem. Pro podrobnější výklad odkazujeme na [12] nebo [8, kapitola II.], popřípadě na obsáhlejší cizojazyčné učebnice [6], [7].

1.4 Co je to PageRank?

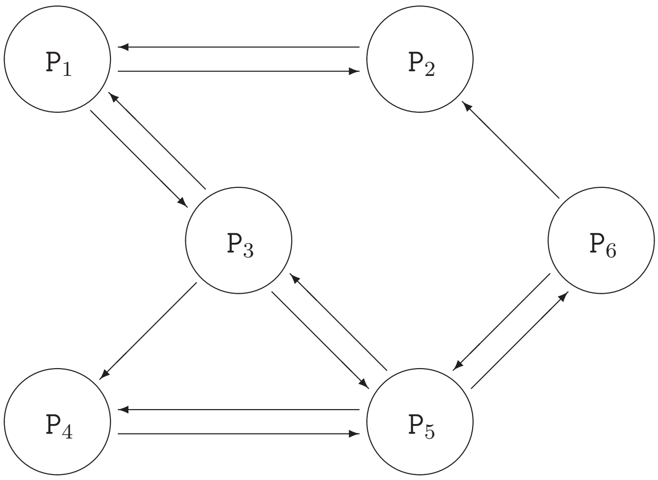
V této části se blíže seznámíme s pojmem PageRank, což je nástroj, který je používán jako míra důležitosti webových stránek.

Definice 8 (PageRank)**.** *PageRank webové stránky* $\mathtt{P}_k$*, značený* $r(\mathtt{P}_k) \in \mathbb{R}$*, je reálné číslo dané jako součet PageRanků všech stránek* $\mathtt{P}_j$*, které na stránku* $\mathtt{P}_k$ *odkazují, dělených celkovým počtem odkazů na stránce* $\mathtt{P}_j$*. Tedy*

$$r(\mathtt{P}_k) = \sum_{\mathtt{P}_j \in \mathscr{B}_{\mathtt{P}_k}} \frac{r(\mathtt{P}_j)}{|\mathtt{P}_j|}, \tag{1.7}$$

kde $\mathscr{B}_{\mathtt{P}_k}$ *je množina všech stránek odkazujících na stránku* $\mathtt{P}_k$ *a* $|\mathtt{P}_j|$ *je počet odkazů na stránce* $\mathtt{P}_j$*.*

Autorem tohoto přístupu k hodnocení webových stránek jsou *Larry Page* a *Sergey Brin*, viz [3], [4] a [5]. Z matematického hlediska je takto zavedené měřítko důležitosti stránek, tj. PageRank, tzv. *Markovovým řetězecem (procesem)*, viz např. [1, kapitola 8], případně [23, kapitoly 3–5] či skriptum [18]. Teorie Markovových řetězců nám také dává nástroje, jak se s výše uvedenou „zacyklenou“ definicí vypořádat.



Obrázek 1.1: Schéma jednoduchého internetu se šesti stránkami. Žádná stránka neodkazuje sama na sebe, na každou stránku odkazuje alespoň jedna jiná stránka a každá stránka obsahuje alespoň jeden odkaz.

Nejjednodušší bude začít s malým pracovním příkladem. V tomto textu budeme využívat služeb malého internetu obsahujícího pouze šest stránek. Na obrázku 1.1 je příklad takového internetu. Všimněme si, že v našem modelovém internetu:

(i) žádná stránka neodkazuje sama na sebe,
(ii) na každou webovou stránku odkazuje alespoň jedna jiná stránka a

(iii) *každá stránka obsahuje alespoň jeden odkaz* a dokonce

(iv) *z každé stránky je možné dostat se pomocí odkazů na libovolnou jinou stránku*, což už vlastnosti (ii) a (iii) implikuje.

Žádná z těchto situací není v reálném prostředí internetu vyloučena. První dvě nejsou z hlediska dalšího vývoje až tak zajímavé. Třetí a čtvrtou situací se budeme muset později zabývat podrobněji, viz kapitoly 4.1, resp. 4.2. Pokusme se nyní pro stránky tohoto internetu vypočítat jednotlivé PageRanky pomocí vztahu (1.7). Dostaneme tak sadu šesti rovnic

$$\begin{aligned}
r(\mathtt{P}_1) &= \tfrac{1}{1}r(\mathtt{P}_2) + \tfrac{1}{3}r(\mathtt{P}_3)\\
r(\mathtt{P}_2) &= \tfrac{1}{2}r(\mathtt{P}_1) + \tfrac{1}{2}r(\mathtt{P}_6)\\
r(\mathtt{P}_3) &= \tfrac{1}{2}r(\mathtt{P}_1) + \tfrac{1}{3}r(\mathtt{P}_5)\\
r(\mathtt{P}_4) &= \tfrac{1}{3}r(\mathtt{P}_3) + \tfrac{1}{3}r(\mathtt{P}_5)\\
r(\mathtt{P}_5) &= \tfrac{1}{3}r(\mathtt{P}_3) + \tfrac{1}{1}r(\mathtt{P}_4) + \tfrac{1}{2}r(\mathtt{P}_6)\\
r(\mathtt{P}_6) &= \tfrac{1}{3}r(\mathtt{P}_5)
\end{aligned},$$

kterou můžeme snadno zapsat maticově jako

$$\underbrace{\begin{bmatrix} r(\mathtt{P}_1)\\ r(\mathtt{P}_2)\\ r(\mathtt{P}_3)\\ r(\mathtt{P}_4)\\ r(\mathtt{P}_5)\\ r(\mathtt{P}_6)\end{bmatrix}}_{\pi} = \underbrace{\begin{bmatrix} 0 & 1/1 & 1/3 & 0 & 0 & 0\\ 1/2 & 0 & 0 & 0 & 0 & 1/2\\ 1/2 & 0 & 0 & 0 & 1/3 & 0\\ 0 & 0 & 1/3 & 0 & 1/3 & 0\\ 0 & 0 & 1/3 & 1/1 & 0 & 1/2\\ 0 & 0 & 0 & 0 & 1/3 & 0\end{bmatrix}}_{H^T} \underbrace{\begin{bmatrix} r(\mathtt{P}_1)\\ r(\mathtt{P}_2)\\ r(\mathtt{P}_3)\\ r(\mathtt{P}_4)\\ r(\mathtt{P}_5)\\ r(\mathtt{P}_6)\end{bmatrix}}_{\pi}, \qquad (1.8)$$

kde matici H nazýváme *maticí weobových odkazů* tzv. *hyperlinků* (anglicky *hyperlink matrix*) a vektor π nazýváme *PageRank vektor*. Všimněme si, že internet na obrázku 1.1 je zároveň grafem $\vec{G}(H)$ své matice hyperlinků H (až na pojmenování vrcholů; vrchol j v grafu $\vec{G}(H)$ odpovídá vrcholu $\mathtt{P}_j$ v obrázku). Porovnáním maticového zápisu (1.8), po záměně levé a pravé strany,

$$H^T\pi = \pi$$

s rovnicí (1.1) vidíme, že PageRank vektor π musí být buď nulový, což by v daném kontextu nebylo příliš přínosné, nebo, má-li být nenulový, musí mít matice hyperlinků H (resp. její transpozice H^T) vlastní číslo $\lambda = 1$. PageRank vektor π je pak vlastním vektorem matice H^T odpovídající tomuto vlastnímu číslu. Důležitou otázkou tedy je, zda má matice hyperlinků vlastní číslo rovné jedné.

V matici H si můžeme všimnout následující zvláštnosti. Součet prvků v řádcích je vždy roven jedné (každá stránka $\mathtt{P}_j$ odkazující na $|\mathtt{P}_j|$ stránek přispívá k jejich hodnocení stejnou měrou rovnou právě číslu $1/|\mathtt{P}_j|$). Kdybychom místo matice H^T

pracovali přímo s maticí H, tak určitě platí

$$\underbrace{\begin{bmatrix} 0 & 1/2 & 1/2 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 1/3 & 0 & 0 & 1/3 & 1/3 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1/3 & 1/3 & 0 & 1/3 \\ 0 & 1/2 & 0 & 0 & 1/2 & 0 \end{bmatrix}}_{H} \underbrace{\begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}}_{e} = \underbrace{\begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}}_{e}. \tag{1.9}$$

Vidíme, že vektor $e = [1, \ldots, 1]^T$ je určitě vlastním vektorem matice H odpovídajícím vlastnímu číslu $\lambda = 1$. Protože $\det(I\lambda - A) = \det(I\lambda - A^T)$, viz [11, kapitola 1], pak $\mathrm{sp}(H) = \mathrm{sp}(H^T)$, a tedy matice H^T musí mít také alespoň jedno vlastní číslo rovné jedné, a tudíž *existuje* nenulový PageRank vektor π.

Poznamenejme, že vlastní vektory matic H a H^T odpovídající stejnému vlastnímu číslu jsou obecně různé, tj. $e \neq \pi$. To je způsobeno tím, že matice H obecně není normální, tedy $HH^T \neq H^T H$, viz [11, kapitola 2]. Pouze pro normální matice platí, že (všechny) jejich levé a pravé vastní vektory jsou stejné; zde je e pravým a π levým vlastním vektorem matice H.

Část I

Existence PageRank vektoru

2 Nezáporné a stochastické matice

V této kapitole se budeme seznámíme se *stochastickými maticemi*, které jsou specifické právě tím, že součet prvků v řádcích je vždy roven jedné. Vycházet budeme z knih [12] a [11].

Zejména zde ukážeme, že pro některé matice A s nezápornými prvky platí, že spektrální poloměr $\varrho(A)$ je (kladné) vlastní číslo nazývané Perronovo vlastní číslo a že mu odpovídá (ne nutně jediný) vlastní vektor s nezápornými prvky nazývaný (levý či pravý) Perronův vlastní vektor.

2.1 Aritmetika nezáporných a kladných matic

Stochastické matice jsou speciálním případem tzv. *nezáporných matic*. Nejprve se tedy seznámíme s nimi.

Definice 9 (Nezáporná matice, kladná matice)**.** *Nechť $A \in \mathbb{R}^{m\times n}$ je reálná matice s prvky $a_{j,k}$. Jestliže pro všechny její prvky platí $a_{j,k} \geq 0$, pak tuto matici A nazýváme nezápornou. Tuto vlastnost budeme zapisovat*

$$A \geq 0.$$

Jestliže navíc platí $a_{j,k} > 0$, pak matici A nazýváme kladnou. Tuto vlastnost budeme zapisovat

$$A > 0.$$

Dále je-li $B \in \mathbb{R}^{m\times n}$ (resp. $B \in \mathbb{C}^{m\times n}$) obecná reálná (resp. komplexní) matice s prvky $b_{j,k}$, pak zápisem

$$|B| \in \mathbb{R}^{m\times n}, \qquad |B| \geq 0$$

budeme značit nezápornou matici, jejíž prvky jsou $|b_{j,k}|$, tj. absolutní hodnoty prvků matice B.

Analogicky zápisem $A \leq 0$ a $A < 0$, budeme značit, že $a_{j,k} \leq 0$, resp. $a_{j,k} < 0$, a budeme takovou matici A nazývat *nekladnou*, resp. *zápornou*. Dále zápisem $A \geq B$, kde A a B jsou reálné matice stejných rozměrů, budeme značit, že $A - B \geq 0$, analogicky pro $A > B$, $A \leq B$, $A < B$. Obdobně zavedeme nerovnosti a značení také pro vektory. Následujících několik lemmat zformuluje některé základní vlastnosti aritmetiky nezáporných matic.

Lemma 3. *Nechť $A \geq 0$ a $B \geq 0$ jsou (reálné) nezáporné matice. Pokud jsou matice A a B stejných rozměrů, pak*

$$A + B \geq 0$$

je nezáporná. Pokud lze matice A a B násobit (v tomto pořadí), pak

$$AB \geq 0 \tag{2.1}$$

je nezáporná. Je-li navíc $A > 0$ kladná matice a $B \neq 0$ nenulová matice, pak

$$AB \geq 0 \qquad \text{a zároveň} \qquad AB \neq 0 \tag{2.2}$$

je nezáporná nenulová matice.

Důkaz tohoto lemmatu je elementární a sestává se pouze z rosepsání maticových operací po prvcích. Další lemma, které pro nás bude důležitější, se zaměří konkrétně na součiny nezáporných matic s vektory.

Lemma 4. *Nechť $A \in \mathbb{R}^{m \times n}$, $A \geq 0$ je nezáporná matice a $x, y \in \mathbb{R}^n$ vektory, pro které platí $x \geq y$. Pak*

$$Ax \geq Ay. \tag{2.3}$$

Je-li navíc $A > 0$ kladná matice a vektory $x \neq y$ jsou různé, pak

$$Ax > Ay. \tag{2.4}$$

Důkaz. Nerovnost (2.3) dostaneme snadno z nerovnosti (2.1), pokud budeme v lemmatu 3 uvažovat matici B ve tvaru $B \equiv x - y \in \mathbb{R}^{n \times 1}$.

Druhou nerovnost (2.4) dokážeme následně. Mějme vektory $x = [\xi_1, \ldots, \xi_n]^T$ a $y = [\nu_1, \ldots, \nu_n]^T$. Podle nerovnosti $x \geq y$ platí

$$\xi_k \geq \nu_k, \qquad k = 1, \ldots, n.$$

Protože $x \neq y$, pak existuje index ℓ takový, že

$$\xi_\ell > \nu_\ell,$$

a bez újmy na obecnosti můžeme předpokládat, že

$$\xi_k = \nu_k, \qquad k = 1, \ldots, \ell - 1, \ell + 1, \ldots, n.$$

Pro j-tý řádek součinů Ax a Ay zřejmě platí

$$e_j^T(Ax) = \left(\sum_{\substack{k=1 \\ k \neq \ell}}^{n} a_{j,k}\xi_k \right) + a_{j\ell}\xi_\ell, \qquad \text{resp.}$$

$$e_j^T(Ay) = \left(\sum_{\substack{k=1 \\ k \neq \ell}}^{n} a_{j,k}\nu_k \right) + a_{j\ell}\nu_\ell = \left(\sum_{\substack{k=1 \\ k \neq \ell}}^{n} a_{j,k}\xi_k \right) + a_{j\ell}\nu_\ell.$$

Protože $A > 0$, tj. $a_{j,\ell} > 0$, a $\xi_\ell > \nu_\ell$, pak

$$e_j^T(Ax - Ay) = a_{j,\ell}(\xi_\ell - \nu_\ell) > 0, \qquad \text{pro} \quad j = 1, \dots, n.$$

Tedy

$$Ax - Ay > 0,$$

z čehož ihned plyne $Ax > Ay$, což jsme chtěli dokázat. $\square$

V úvodním příkladu (1.8)–(1.9) jsme pracovali s nezápornou maticí H (resp. H^T), která byla navíc čtvercová. Následující věta bude užitečná při zkoumání čtvercových nezáporných matic, viz [11, cvičení 2.15].

Věta 2. *Nechť $A \in \mathbb{R}^{n\times n}$, $A \geq 0$ je nezáporná čtvercová matice, $x \in \mathbb{R}^n$, $x \geq 0$ je nezáporný vektor a ς reálné číslo takové, že*

$$Ax > x\,\varsigma.$$

Pak platí

$$\varrho(A) > \varsigma, \tag{2.5}$$

kde $\varrho(A)$ je spektrální poloměr matice A (viz definice 1).

Důkaz. Vzhledem tomu, že A, x a $\varrho(A)$ jsou nezáporná matice, vektor a číslo, viz (1.2), nerovnosti $Ax > x\,\varsigma$ a $\varrho(A) > \varsigma$ jsou splněny triválně pro $\varsigma < 0$. Nechť je tedy $\varsigma \geq 0$ (a tudíž také $A \neq 0$, $x \neq 0$). Z předpokladu $Ax > x\,\varsigma$ vyplývá, že existuje $\varepsilon > 0$ takové, že platí

$$Ax \geq x(\varsigma + \varepsilon).$$

Zadefinujme si nyní matici $B \equiv (\varsigma + \varepsilon)^{-1}A$. Tato matice je zřejmě nezáporná, tj. $B \geq 0$, $B \neq 0$, a navíc platí

$$Bx \geq x.$$

Opakovaným užitím tvrzení lemmatu 4, konkrétně vztahu (2.3), pak dostaneme

$$B^\ell x \geq B^{\ell-1}x \geq \dots \geq x, \qquad \text{pro} \quad \ell = 1, 2, \dots \tag{2.6}$$

Víme, že platí

$$\lim_{\ell \longrightarrow \infty} B^\ell = 0 \qquad \Longleftrightarrow \qquad \varrho(B) < 1,$$

viz [11, cvičení 1.23]. Ze vztahu (2.6) však vidíme, že posloupnost matic B^ℓ nemůže konvergovat k nulové matici, a tudíž musí mít matice B spektrální poloměr větší nebo roven jedné,

$$\varrho(B) \geq 1,$$

Pro matici $A = (\varsigma + \varepsilon)B$ pak ihned dostaneme

$$\varrho(A) \geq \varsigma + \varepsilon > \varsigma,$$

čímž je důkaz hotov. $\square$

Na závěr ještě stručně připomeneme vlastnosti aritmetiky komplexních čísel, resp. jejich absolutních hodnot.

Lemma 5. *Nechť $A \in \mathbb{C}^{m\times n}$ a $B \in \mathbb{C}^{n\times d}$ jsou dvě libovolné komplexní matice, pak*

$$|A|\,|B| \geq |AB|. \tag{2.7}$$

Důkaz. Zřejmě pro $a, b \in \mathbb{C}$ platí $|a|\,|b| = |ab|$, $|a|+|b| \geq |a+b|$. Speciálně pro prvky $a_{j,k}$ a $b_{k,\ell}$ matice A, resp. B platí

$$\sum_{k=1}^{n} |a_{j,k}|\,|b_{k,\ell}| = \sum_{k=1}^{n} |a_{j,k}b_{k,\ell}| \geq \left|\sum_{k=1}^{n} a_{j,k}b_{k,\ell}\right|$$

pro $j = 1,\dots,m$ a $\ell = 1,\dots,d$. □

Toto triviální pozorování má následující důsledek, který bude později užitečný.

Lemma 6. *Nechť $A \in \mathbb{R}^{m\times n}$ je nezáporná matice, $A \geq 0$, a $B \in \mathbb{C}^{n\times d}$ je libovolná komplexní matice taková, že součin $AB \in \mathbb{R}^{m\times d}$ je reálný. Pak*

$$A|B| \geq AB. \tag{2.8}$$

Je-li navíc matice A kladná, $A > 0$, pak

$$A|B| = AB \qquad \Longleftrightarrow \qquad |B| = B, \tag{2.9}$$

tj. v nerovnosti (2.8) nastane rovnost jen tehdy, když B je reálná a nezáporná.

Důkaz. Je-li $A \geq 0$ nezáporná a AB reálná, pak $|A| = A$, resp. $|AB| \geq AB$, a nerovnost (2.8) dostáváme rovnou z nerovnosti (2.7). V druhé části lemmatu je implikace „$\Longleftarrow$“ zřejmá. Když $|B| = B$, pak $A|B| = AB$. Zbývá tedy dokázat implikaci „$\Longrightarrow$“ opačným směrem. Předpokládejme *naopak*, že druhá implikace neplatí, tj. že

$$A|B| = AB \qquad \text{a zároveň} \qquad |B| \neq B.$$

Přeuspořádáním první rovnosti dostaneme

$$A\Big(|B| - B\Big) = AM = 0, \qquad \text{kde} \qquad M \equiv |B| - B.$$

Protože $|B| \neq B$, pak matice M má alespoň jeden nenulový prvek. Protože $A > 0$ je kladná, pak $AM \neq 0$, což je spor. Tím je dokázána i druhá implikace, a tím i celé lemma. □

2.2 Stochastické matice: spektrální poloměr

Důležitou podmnožinou nezáporných matic jsou tzv. *stochastické matice*. Tyto matice jsou nedílnou součástí výpočtu PageRanku, a proto si v této kapitole stochastické matice zadefinujeme a blíže popíšeme jejich vlastnosti.

Definice 10 (Stochastická matice)**.** *Čtvercová nezáporná matice $S \in \mathbb{R}^{n\times n}$ s prvky $s_{j,k}$, pro kterou platí*

$$\sum_{k=1}^{n} s_{j,k} = 1, \qquad \textit{pro} \quad j = 1, \dots, n,$$

se nazývá stochastická matice (viz např. [12, str. 99]).

Prvky stochastické matice můžeme považovat například za pravděpodobnosti. Všimněme si, že v matici hyperlinků H to můžeme interpretovat jako pravděpodobnosti přechodu ze stránky $\mathtt{P}_j$ na $\mathtt{P}_k$. Jak definice tvrdí, stochastická matice S má součet všech prvků v řádku vždy roven jedné. To lze vyjádřit také jako $Se = e$, kde $e = [1, \dots, 1]^T$. Tedy e je (pravý) vlastní vektor stochastické matice S a odpovídá vlastnímu číslu $\lambda = 1$, viz také (1.9). Vidíme, že úloha (1.8), kterou chceme řešit je následující:

> Hledáme levý vlastní vektor π stochastické matice S odpovídající vlastnímu číslu $\lambda = 1$. Tento vektor je (levý) *Perronův vlastní vektor*.

To, co bychom nyní rádi ukázali, je, zda lze Perronův vlastní vektor volit kladný (tj. abychom mohli stránky pomocí komponenty PageRank vektor rozumně seřadit) a zda je určen jednoznačně.

Nyní ukážeme, že pro libovolnou stochastickou matici S dokonce platí $\varrho(S) = 1$, tj. matice S nemá žádné vlastní číslo, které by bylo v absolutní hodnotě větší než jedna.

Věta 3. *Nechť S je čtvercová nezáporná stochastická matice, $s_{j,k} \geq 0$, $\sum_{k=1}^{n} s_{j,k} = 1$ pro $j = 1, \dots, n$. Pak platí*

$$\varrho(S) = 1.$$

Důkaz. Pro $e = [1, \dots, 1]^T$ zřejmě platí

$$Se = \begin{bmatrix} \sum_{k=1}^{n} a_{1,k} \\ \vdots \\ \sum_{k=1}^{n} a_{n,k} \end{bmatrix} = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} = e,$$

tedy $\lambda = 1$ je vlastním číslem stochastické matice, takže

$$\varrho(S) \geq 1. \tag{2.10}$$

Zároveň vidíme, že

$$\|S\|_\infty = 1,$$

viz (1.5). Použijeme-li tvrzení lemmatu 2 na stochastickou matici S a na ∞-normu (1.4), dostaneme

$$\varrho(S) \leq \|S\|_\infty = 1. \tag{2.11}$$

Z nerovností (2.10) a (2.11) zřejmě vyplývá, že $\varrho(S) = 1$. □

2.3 Stochastické matice: (levý) Perronův vlastní vektor

Nyní již víme, že $\varrho(S) = 1$ pro každou stochastickou matici, což znamená, že neexistuje žádné jiné vlastní číslo matice S, které by bylo v absolutní hodnotě větší než jedna. Pojďme se nyní podívat na další část naší úlohy. Hledáme levý vlastní vektor π, který bychom rádi volili kladný. Tato podmínka je pro nás důležitá, protože nám umožní rozumně seřadit webové stránky. V této kapitole se přesvědčíme, že tento kladný vektor exiistuje. Nejprve se ovšem přesvědčíme, že levý vlastní vektor odpovídající vlastnímu číslu $\lambda = 1$ je možné volit nezáporný. Budeme vycházet z [11, cvičení 2.16].

Věta 4. *Nechť S je čtvercová nezáporná stochastická matice, $s_{j,k} \geq 0$, $\sum_{k=1}^{n} s_{j,k} = 1$ pro $j = 1, \ldots, n$. Nechť vlastnímu číslu $\lambda \equiv \varrho(S) = 1$ odpovídá levý vlastní vektor π. Pak lze tento vektor π volit nezáporný, tj.*

$$\pi^T S = \pi^T, \qquad \text{kde} \qquad \pi \geq 0.$$

Důkaz. Mějme čtvercovou nezápornou stochastickou matici S. Z věty 3 víme, že $\varrho(S) = 1$. Označme ϖ_j složky levého vlastního vektoru π, tedy $\pi = [\varpi_1, \ldots, \varpi_n]^T$. Pokud $\varpi_j \geq 0$, $j = 1, \ldots, n$, pak π je právě hledaný nezáporný vektor, pro který platí

$$\pi^T S = \pi^T.$$

Je-li $\varpi_j \leq 0$, $j = 1, \ldots, n$ pak $(-\pi)^T$ je hledaný nezáporný vektor, pro který platí

$$(-\pi)^T S = (-\pi)^T.$$

Předpokládejme nyní, v rozporu s tvrzením věty, že prvky vektoru π mají různá znaménka. Pokud existuje alespoň jeden prvek s odlišným znaménkem než mají ostatní prvky, pak

$$\sum_{j=1}^{n} |\varpi_j| s_{j,k} > \left| \sum_{j=1}^{n} \varpi_j s_{j,k} \right| = |\varpi_k|, \qquad \text{pro} \qquad k = 1, \ldots, n.$$

Z toho vyplývá, že platí

$$|\pi|^T S > |\pi|^T, \qquad \text{kde} \qquad |\pi| \equiv \Big[|\varpi_1|, \ldots, |\varpi_n| \Big]^T.$$

Podle věty 2 pro nezápornou matici $A \equiv S^T$, nezáporný vektor $x \equiv |\pi|$ a reálné číslo $\varsigma \equiv 1$ takové, že $S^T|\pi| \equiv Ax > x\lambda \equiv |\pi|$, platí $\varrho(A) > \varsigma$. Takže v našem případě musí být

$$\varrho(S^T) \equiv \varrho(S) > 1,$$

což je ve sporu s vlastností $\varrho(S) = 1$, viz větu 3. Tím jsme dokázali, že levý vlastní vektor π můžeme volit nezáporný. □

Z kapitoly 2.2 víme, že spektrální poloměr čvercové nezáporné stochastické matice je roven jedné. Navíc víme, že spektrální poloměr je přímo vlastním číslem, tj. $\lambda = 1$. Nyní jsme zjistili, že levý vlastní vektor odpovídající tomuto vlastnímu čísle lze volit nezáporný. To má následující důsledek pro naši úlohu.

> PageRank vektor π splňující $\pi^T H = \pi^T$, viz (1.7) a (1.8), kde H je matice hyperlinků, *existuje* a jeho složky, tj. ranky (hodnocení) jednotlivých stránek $\mathtt{P}_j$, jsou nezáporná čísla. Můžeme je tedy snadno seřadit.

2.4 Kladné matice: spektrální poloměr a (pravý) Perronův vlastní vektor

Větu 4 lze modifikovat pro libovolnou kladnou čtvercovou matici, tj. $A > 0$. Tato modifikace mění předpoklady věty. Na jedné straně své požadavky zesílíme tím, že vyžadujeme nenulovost prvků, na druhé straně je ale oslabíme tím, že nepožadujeme stochasticitu matice. Tvrzení modifikované věty je také nepatrně silnější. Tato modifikace je známá jako *Perronova věta*, případně *Perronovo lemma*. Tato věta se bude také hodit v následujícím textu. Budeme vycházet z výkladu v knize [12, str. 86–87].

Věta 5 (Perronova věta (o spektrálním poloměru a kladném vlastním vektoru kladných matic))**.** *Nechť A je čtvercová matice s kladnými prvky, tj. $a_{j,k} > 0$. Pak $\lambda \equiv \varrho(A)$ je kladné vlastní číslo matice A a odpovídající vlastní vektor x lze volit kladný, tj.*

$$Ax = x\lambda, \qquad \textit{kde} \qquad x > 0.$$

Důkaz. Mějme $A > 0$ a vlastní číslo λ matice A, pro které platí

$$\varrho(A) = |\lambda|. \tag{2.12}$$

Mějme nenulový vlastní vektor $x \neq 0$ odpovídající číslu λ, tj.

$$Ax = x\lambda. \tag{2.13}$$

Poznamenejme, že pro libovovolná komplexní čísla v a w a kladná čísla α a β platí

$$\alpha|v| + \beta|w| \geq |\alpha v + \beta w|, \tag{2.14}$$

přičemž rovnost nastane právě tehdy, když existuje komplexní jednotka η ($|\eta| = 1$, tj. $\eta = \exp(\mathbf{i}\,\varphi) = \cos(\varphi) + \mathbf{i}\sin(\varphi)$) tak, že $v\eta,\, w\eta \in \mathbb{R}$ a navíc $v\eta,\, w\eta \geq 0$. Jednoduchým zobecněním tohoto pozorování dostaneme, že pro komplexní čísla $\xi_1, \ldots, \xi_n$ a kladná čísla $\alpha_1, \ldots, \alpha_n$ platí

$$\sum_{j=1}^{n} \alpha_j |\xi_j| \;\geq\; \left| \sum_{j=1}^{n} \alpha_j \xi_j \right|; \tag{2.15}$$

viz také (2.7). Rovnost při tom nastane pouze tehdy, když existuje komplexní jednotka η taková, že $\xi_j \eta \in \mathbb{R}$ a navíc

$$\xi_j \eta \geq 0, \qquad \text{pro} \quad j = 1, \ldots, n. \tag{2.16}$$

Předpokládejme nyní, že pro vektor $x = [\xi_1, \ldots, \xi_k]^T$, respektive pro jeho složky ξ_j, neexistuje takové η, aby vztah (2.16) platil. Pak z k-tého řádku rovnice (2.13) dostaneme užitím vztahu (2.15) nerovnost

$$\sum_{j=1}^{n} a_{k,j} |\xi_j| \;>\; \left| \sum_{j=1}^{n} a_{k,j} \xi_j \right| = |\lambda| |\xi_k|, \qquad \text{pro} \qquad k = 1, \ldots, n.$$

Z toho vyplývá, že platí

$$A|x| > |x||\lambda|, \qquad \text{kde} \qquad |x| \equiv \Big[|\xi_1|, \ldots, |\xi_n| \Big]^T.$$

Podle věty 2 pro nezápornou matici A, nezáporný vektor $|x|$ a reálné číslo $|\lambda|$ takové, že $A|x| > |x||\lambda|$, platí $\varrho(A) > |\lambda|$. Tedy v našem případě musí být

$$\varrho(A) > |\lambda|,$$

což je ve sporu s vlastností z (2.12). Tím jsme dokázali, že existuje taková komplexní jednotka η, že vztah (2.16) pro složky vektoru x platí. Takže

$$\widetilde{x} = x\eta \in \mathbb{R}, \quad \text{kde} \quad \widetilde{x} \equiv [\widetilde{\xi}_1, \ldots, \widetilde{\xi}_n], \quad \text{tj.} \quad \widetilde{\xi}_j = \xi_j \eta,$$

a platí

$$\widetilde{x} \geq 0, \quad \widetilde{x} \neq 0.$$

Navíc podle (2.13) platí

$$A\widetilde{x} = \widetilde{x}\lambda, \tag{2.17}$$

tedy vlastní vektor odpovídající vlastnímu číslu λ lze volit nezáporný. Nyní zbývá dokázat, že $\lambda > 0$, tj. spektrální poloměr $\varrho(A) = |\lambda| = \lambda$ je přímo vlastním číslem, a dále, že $\widetilde{x} > 0$.

Protože $A > 0$, $\widetilde{x} \geq 0$, $\widetilde{x} \neq 0$, pak podle lemmatu 4 platí $A\widetilde{x} > A \cdot 0$, viz (2.4). Protože $\widetilde{x} \neq 0$, určitě existuje index ℓ takový, že $\widetilde{\xi}_\ell > 0$. Z ℓ-tého řádku rovnice (2.17), tj. ze vztahu $e_\ell^T A\widetilde{x} = \lambda\widetilde{\xi}_\ell$, kde $(e_\ell^T A\widetilde{x}) > 0$, pak plyne $\lambda > 0$. Tudíž

$$\varrho(A) = |\lambda| = \lambda,$$

čímž jsme dokázali, že spektrální poloměr matice A s kladnými prvky je přímo jejím vlastním číslem.

Konečně z k-tého řádku rovnice (2.17), tj. ze vztahu $e_k^T A\widetilde{x} = \lambda\widetilde{\xi}_k$, kde $(e_k^T A\widetilde{x}) > 0$ a $\lambda > 0$, plyne

$$\widetilde{\xi}_k > 0, \qquad \text{pro} \qquad k = 1, \ldots, n. \tag{2.18}$$

Tím jsme dokázali, že vlastní vektor matice A s kladnými prvky odpovídající vlastnímu číslu $\varrho(A)$ vždy můžeme volit kladný. □

3 Nerozložitelnost matic

V předchozí kapitole jsme ukázali, že spektrální poloměr stochastické (případně kladné) matice je přímo vlastním číslem této matice a ukázali jsme, že vlastní vektor, který tomuto vlastnímu číslu odpovídá je nezáporný (případně kladný). Přirozeně nás může dále zajímat, zda je nezáporný (případně kladný) vlastní vektor z věty 4 (resp. 5) určen jednoznačně.

3.1 Stochastické matice: jednoznačnost (levého) Perronova vlastního vektoru

Perronův vektor stochastické matice S, tj. levý vlastní vektor π, $\pi \geq 0$ odpovídající vlastnímu číslu $\lambda = \varrho(S) = 1$ obecně není obecně určen jednoznačně. Pokusíme se to ilustrovat na několika příkladech. Uvažujme nejjednodušší čtvercovou nezápornou stochastickou matici

$$S \equiv I = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix} \in \mathbb{R}^{n \times n}.$$

Zřejmě $\varrho(S) = 1$ a $\lambda = 1$ je n-násobné vlastní číslo matice S. Pro *jakýkoliv* nezáporný (případně kladný) vektor π, resp. x, platí

$$\pi^T S = \pi^T, \quad \text{resp.} \quad Sx = x.$$

Úloha tedy obecně *není* jednoznačná (a to ani v případě, že budeme pracovat s normalizovaným vektorem, tj. když $\|\pi\| = \|x\| = 1$).

Vezměme si nyní méně triviální případ. Mějme čtvercovou nezáporou stochastickou matici S v blokově diagonálním tvaru se čtvercovými bloky S_1 a S_2 na diagonále,

$$S = \begin{bmatrix} S_1 & 0 \\ 0 & S_2 \end{bmatrix}.$$

Zřejmě je matice S nezáporná stochastická matice tehdy a jen tehdy když matice S_1 a S_2 jsou nezáporné stochastické matice. Označme π_1 a π_2 levými nezápornými vlastními vektory matic S_1 a S_2 odpovídající vlastnímu číslu $\lambda = \varrho(S_1) = \varrho(S_2) = 1$, tj.

$$\pi_1^T S_1 = \pi_1^T, \qquad \pi_1 \geq 0,\ \pi_1 \neq 0,$$

$$\pi_2^T S_2 = \pi_2^T, \qquad \pi_2 \geq 0,\ \pi_2 \neq 0.$$

Pak $\varrho(S) = 1$ a $\lambda = 1$ je minimálně dvojnásobné vlastní číslo. Levý vlastní vektor matice S odpovídající vlastnímu číslu $\lambda = 1$, tj. $\pi^T S = \pi^T$, lze volit například následujícími způsoby

$$\pi = \begin{bmatrix} \pi_1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ \pi_2 \end{bmatrix}, \begin{bmatrix} \pi_1 \\ \pi_2 \end{bmatrix}, \quad \text{případně obecně} \quad \pi = \begin{bmatrix} \pi_1\alpha_1 \\ \pi_2\alpha_2 \end{bmatrix},$$

kde $\alpha_1, \alpha_2 \in \mathbb{R}$, $\alpha_1, \alpha_2 \geq 0$, $\alpha_1 + \alpha_2 > 0$. Ve všech těchto případech $\pi \geq 0$, $\pi \neq 0$. Je tedy zřejmé, že ani nyní levý vlastní vektor π matice S není dán jednoznačně. Výše zmíněnou blokovou strukturu budeme ilustorvat na příkladu našeho modelového internetu, viz obrázek 1.1, resp. rovnici (1.8). Příklad internetu s takovou blokou strukturou získáme vynecháním několika odkazů (hran v grafu), viz obrázek 3.1. Internetu z obrázku 3.1 pak odpovídá rovnice

$$\underbrace{\begin{bmatrix} r(\mathtt{P}_1) \\ r(\mathtt{P}_2) \\ r(\mathtt{P}_3) \\ r(\mathtt{P}_4) \\ r(\mathtt{P}_5) \\ r(\mathtt{P}_6) \end{bmatrix}}_{\pi} = \underbrace{\left[\begin{array}{ccc|ccc} 0 & 1/1 & 1/1 & 0 & 0 & 0 \\ 1/2 & 0 & 0 & 0 & 0 & 0 \\ 1/2 & 0 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & 1/2 & 0 \\ 0 & 0 & 0 & 1/1 & 0 & 1/1 \\ 0 & 0 & 0 & 0 & 1/2 & 0 \end{array}\right]}_{H^T} \underbrace{\begin{bmatrix} r(\mathtt{P}_1) \\ r(\mathtt{P}_2) \\ r(\mathtt{P}_3) \\ r(\mathtt{P}_4) \\ r(\mathtt{P}_5) \\ r(\mathtt{P}_6) \end{bmatrix}}_{\pi}, \tag{3.1}$$

všimněme si, že tato matice H má právě dvě vlastní čísla rovna jedné.

3.2 (Ne)rozložitelné matice

Při zjednoznačnění úlohy bude hrát důležitou roli tzv. *nerozložitelnost matic*. Začneme proto definicí (ne)rozložitelných matic. Vycházet budeme z knihy [12, kapitola 3, str. 71].

Definice 11 ((Ne)rozložitelná matice)**.** *Mějme čtvercovou matici $A \in \mathbb{R}^{n\times n}$. Pokud existuje permutace taková, že matice $\Pi A \Pi^T$, kde Π je odpovídající permutační matice, je v blokově (horním) trojúhelníkovém tvaru s alespoň dvěma čtvercovými bloky na diagonále, tj.*

$$\Pi A \Pi^T = \begin{bmatrix} A_{1,1} & A_{1,2} \\ 0 & A_{2,2} \end{bmatrix},$$

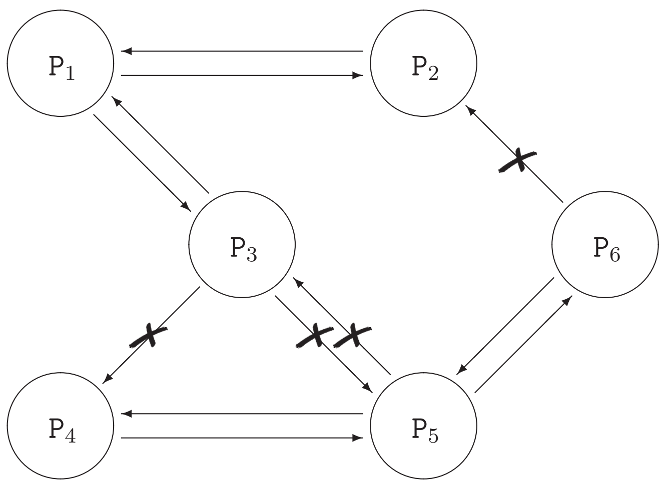
pak se matice A nazývá rozložitelná. Matice, která není rozložitelná, se nazývá nerozložitelná.

Než budeme pokračovat ve výkladu, pokusíme se pojem rozložitelnosti ilustrovat opět na příkladu našeho modelového internetu. Příklad internetu s rozložitelnou maticí opět získáme vynecháním několika odkazů (hran v grafu), viz obrázek 3.2. Internetu z obrázku 3.2 pak odpovídá rovnice

$$\underbrace{\begin{bmatrix} r(\mathtt{P}_1) \\ r(\mathtt{P}_2) \\ r(\mathtt{P}_3) \\ r(\mathtt{P}_4) \\ r(\mathtt{P}_5) \\ r(\mathtt{P}_6) \end{bmatrix}}_{\pi} = \underbrace{\left[\begin{array}{ccc|ccc} 0 & 1/1 & 1/2 & 0 & 0 & 0 \\ 1/2 & 0 & 0 & 0 & 0 & 0 \\ 1/2 & 0 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & 1/2 & 0 \\ 0 & 0 & 1/2 & 1/1 & 0 & 1/1 \\ 0 & 0 & 0 & 0 & 1/2 & 0 \end{array}\right]}_{H^T} \underbrace{\begin{bmatrix} r(\mathtt{P}_1) \\ r(\mathtt{P}_2) \\ r(\mathtt{P}_3) \\ r(\mathtt{P}_4) \\ r(\mathtt{P}_5) \\ r(\mathtt{P}_6) \end{bmatrix}}_{\pi}. \tag{3.2}$$

Připomeňme, že schémata na obrázcích 1.1, 3.1 a 3.2 jsou grafy $\vec{G}(H)$ matice H z rovnic (1.8), (3.1) a (3.2). Všimněme si, že graf na obrázku 1.1 je *silně souvislý* (tj. existuje *cesta* mezi libovolnými dvěma vrcholy; viz sekce 1.3) a matice H, která mu odpovídá, je *nerozložitelná*. Podobně obě nezávislé komponenty $\mathscr{S}_1 = \{\mathtt{P}_1, \mathtt{P}_2, \mathtt{P}_3\}$ a $\mathscr{S}_2 = \{\mathtt{P}_4, \mathtt{P}_5, \mathtt{P}_6\}$ na obrázku 3.1 jsou *silně souvislé* a jim odpovídající bloky na diagonále matice H jsou opět *nerozložitelné*. Toto pozorování se dá snadno zobecnit. Zřejmě platí následující lemma.

Lemma 7. *Matice A je nerozložitelná tehdy a jen tehdy, když je její graf $\vec{G}(A)$ silně souvislý.*



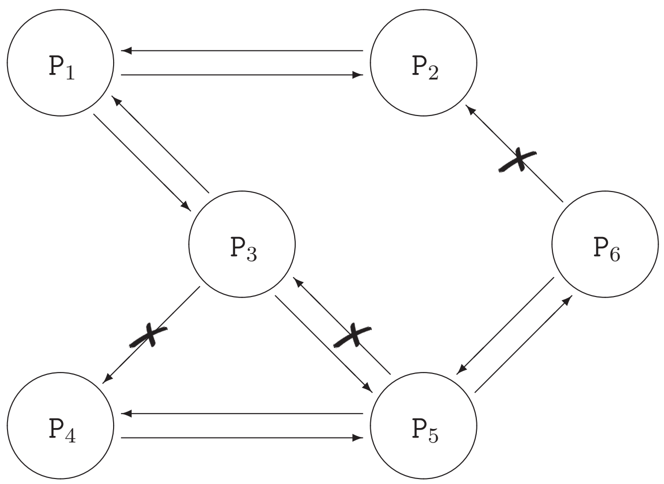
Obrázek 3.1: Jednoduchý model internetu z obrázku 1.1 po vyškrtání některých odkazů (hran); symbol „✗“ značí vyškrtnutou hranu. Vidíme, že internet se nyní skládá ze dvou nezávislých částí (graf obsahuje dvě *nezávislé komponenty*), sítí $\mathscr{S}_1 \equiv \{\mathtt{P}_1, \mathtt{P}_2, \mathtt{P}_3\}$ a $\mathscr{S}_2 \equiv \{\mathtt{P}_4, \mathtt{P}_5, \mathtt{P}_6\}$. Masový výskyt takové struktury ve skutečném internetu (tj. výskyt dvou nebo více rozsálých oblastí zcela nepropojených webovými odkazy) je značně nepravděpodobný.

Důkaz. Důkaz je možné provést snadno. Předpokládáme-li rozložitelnost matice, snadno ukážeme, že jsou v grafu takové vrcholy, mezi nimiž neexistuje cesta (tedy *silná souvislost* implikuje *nerozložitelnost*).

Předpokládáme-li naopak, že graf není silně souvislý, můžeme množinu vrcholů rozložit na systém (alespoň dvou) navzájem disjuktních podmnožin tvořících silně souvislé komponenty *maximální* velikosti. Mezi libovolnými dvěma takovými komponentami buď žádné hrany nevedou, nebo vedou jen jedním směrem (pokud by vedli oběma směry, sjednocení příslušných komponent tvoří opět silně souvislou komponentu, což je ve sporu s maximální velikostí silně souvislých komponent). Vhodným seřazením těchto komponent (tj. vhodným očíslováním vrcholů grafu, tj. vhodnou volbou permutační matice P z definice 11) dostaneme matici v blokově (horním) trojúhelníkovém tvaru se čtvercovými bloky (odpovídajícími silně souvislým komponentám) na diagonále (tedy *nerozložitelnost* implikuje *silnou souvislost*). Pro podrobnější výklad viz [12, věta 3.6]. □

Pro další výklad bude důležité následující lemma a jeho důsledek, viz [12, věty 4.4 a 4.5].

Lemma 8. *Nechť A je čtvercová nezáporná matice, tj. $a_{j,k} \geq 0$, a ℓ přirozené číslo, potom pro matici $B = A^\ell$ platí, že $b_{j,k} > 0$ tehdy a jen tehdy, když v orientovaném grafu $\vec{G}(A)$ existuje cesta z vrcholu j do vrcholu k délky ℓ.*



Obrázek 3.2: Jednoduchý model internetu z obrázku 1.1 po vyškrtání některých odkazů (hran); symbol „✗“ značí vyškrtnutou hranu. Na rozdíl od internetu z obrázku 3.1, tento internet neobsahuje nezávislé komponenty. Vidíme ale, že stále ze sítě $\mathscr{S}_2 = \{\mathtt{P}_4, \mathtt{P}_5, \mathtt{P}_6\}$ nevede žádný odkaz do sítě $\mathscr{S}_1 = \{\mathtt{P}_1, \mathtt{P}_2, \mathtt{P}_3\}$ (graf neobsahuje cestu mezi libovolnými dvěma vrcholy). Matice hyperlinků odpovídající takovému internetu je (stejně jako v případě obrázku 3.1) rozložitelná; matice hyperlinků odpovídající sítím $\mathscr{S}_1$ a $\mathscr{S}_2$ už rozložitelné nejsou (stejně jako matice celého internetu z obrázku 1.1).

Důkaz. Tvrzení je zřejmě pravdivé pro $\ell = 1$. Předpokládejme tedy, že tvrzení platí až do ℓ. Z tohoto předpokladu dokážeme, že platí i pro $\ell + 1$. Označme $B = A^\ell$ a $C = BA = A^{\ell+1}$. Pak

$$c_{j,k} = \sum_i b_{j,i} a_{i,k}. \tag{3.3}$$

Nechť existuje v $\vec{G}(A)$ cesta délky $\ell + 1$ z j do k, tj.

$$(j, p_1), \quad (p_1, p_2), \quad (p_2, p_3), \quad \ldots, \quad (p_{\ell-1}, p_\ell), \quad (p_\ell, k).$$

Pak podle indukčního předpokladu platí $b_{j,p_\ell} > 0$. Zároveň platí $a_{p_\ell,k} > 0$. Všechny sčítance v 3.3 jsou nezáporné, přičemž alespoň jeden z nich je kladný, konkrétně $b_{j,p_\ell} a_{p_\ell,k} > 0$. Tedy $c_{j,k} > 0$. V matici $A^{\ell+1}$ se na pozici (j, k) nachází nenulový prvek.

Předpokládejme nyní obráceně, že je v matici $A^{\ell+1}$ na pozic (j, k) nenulový prvek. Potom v součtu 3.3 musí být některý člen kladný. Nechť např. $b_{j,t} a_{t,k} > 0$. Pak nutně $b_{j,t} > 0$ a $a_{t,k} > 0$. Podle indukčního předpokladu existuje v grafu $\vec{G}(A)$ cesta délky ℓ z j do t. Protože $a_{t,k} > 0$ exituje v grafu hrana vedoucí z t do k. Potom také existuje cesta délky $\ell + 1$ z j do k. □

Protože v grafu nerozložitelné matice $A \in \mathbb{C}^{n \times n}$ vždy existuje cesta mezi libovolnými dvěma vrcholy délky ℓ, $\ell < n$ (graf $\vec{G}(A)$ má právě n různých vrcholů; viz sekce 1.3), má předchozí lemma následující důležitý důsledek.

Lemma 9. *Nechť A je nezáporná nerozložitelná matice, tj. $a_{j,k} \geq 0$, a nechť $\alpha_0, \alpha_1, \ldots, \alpha_{n-1}$ jsou kladná čísla, pak matice $\alpha_0 I + \alpha_1 A + \alpha_2 A^2 + \cdots + \alpha_{n-1} A^{n-1}$ je kladná. Speciálně platí $(I + A)^{n-1} > 0$.*

Důkaz. $\sum_{p=0}^{n-1} \alpha_p A^p$ je součet nezáporných matic. Musíme tedy dokázat, že pro libovolná pevná j a k je u některé mocniny A^p, $0 \leq p \leq n - 1$, na místě (j, k) kladný prvek. Pokud $j = k$, pak je to pravda vždy, protože I má na tomto místě kladný prvek. V případě, že $j \neq k$, pak z nerozložitelnosti matice plyne (viz lemma 7), že v grafu $\vec{G}(A)$ existuje cesta z vrcholu j do vrcholu k délky ℓ, kde $\ell < n$. Podle lemmatu 8 má právě matice A^ℓ kladný prvek na pozici (j, k).

První tvrzení je tudíž dokázáno a druhé trvzení je triviálním důsledkem. Podle binomické věty totiž platí

$$(A + I)^{n-1} = \sum_{p=0}^{n-1} \binom{n-1}{p} A^p, \tag{3.4}$$

kde $A^0 \equiv I$; stačí proto položit $\alpha_p = \binom{n-1}{p}$. □

Nyní budeme pokračovat v analýze jednoznačnosti vlastního vektoru, který odpovídá vlastnímu číslu rovnému spektrálnímu poloměru.

3.3 Kladné matice: jednoznačnost (pravého) Perronova vlastního vektoru

Začneme rozšířením Perronovy věty o spektrálním poloměru a kladném vlastním vektoru kladných matic, viz větu 5.

Věta 6 (Perronova věta (o jednoznačnosti Perronova vlastního páru kladných matic))**.** *Nechť A je čtvercová matice s kladnými prvky, tj. $a_{j,k} > 0$. Pak vlastní číslo $\lambda = \varrho(A)$ matice A je jednoduché a odpovídající vlatní vektor x, $x > 0$, $Ax = x\varrho(A)$, je určen jednoznačně.*

Důkaz. Poznamenejme nejprve, že vlastní číslo splňující $\lambda = \varrho(A)$ podle první Perronovy věty (viz větu 5) vždy existuje. Předpokládejme nyní, že k vlastnímu číslu λ existují dva lineárně nezávislé vektory $x = [\xi_1, \ldots, \xi_n]^T$ a $v = [\nu_1, \ldots, \nu_n]^T$. Protože $x \neq 0$, tak existuje index ℓ takový, že $\xi_\ell \neq 0$. Vektor $w = v - x(\xi_\ell^{-1}\nu_\ell)$ je netriviální lineární kombinací lineárně nezávislých vektorů, je tedy nenulový, tj. $w \neq 0$. Zřejmě je také vlastním vektorem matice A odpovídající vlastnímu číslu λ,

$$Aw = Av - Ax(\xi_\ell^{-1}\nu_\ell) = v\lambda - x(\xi_\ell^{-1}\nu_\ell)\lambda = (v - x(\xi_\ell^{-1}\nu_\ell))\lambda = w\lambda.$$

Jeho ℓ-tá složka je však nulová, tj.

$$e_\ell^T w = \nu_\ell - \xi_\ell(\xi_\ell^{-1}\nu_\ell) = 0.$$

Z první Perronovy věty (viz větu 5 a její důkaz) ovšem vyplývá, že je-li w vlastní vektor odpovídající vlastnímu číslu $\lambda = \varrho(A)$, pak existuje komplexní jednota η ($\eta = \exp(\mathbf{i}\,\varphi)$) taková, že $\widetilde{w} = w\eta \geq 0$. Z ℓ-tého řádku rovnice $A\widetilde{w} = \widetilde{w}\lambda$, tj. $(e_\ell^T A)\widetilde{w} = \lambda(e_\ell^T \widetilde{w})$, kde $(e_\ell^T A) > 0$ a $\lambda > 0$, pak dostaneme $(e_\ell^T A)\widetilde{w} > 0$, tedy také $\lambda(e_\ell^T \widetilde{w}) > 0$, což je ovšem ve *sporu* s $e_\ell^T \widetilde{w}\,\overline{\eta} = e_\ell^T w = 0$. Tím jsme dokázali, že je-li $A > 0$, pak vlastnímu číslu $\lambda = \varrho(A)$ odpovídá jediný lineárně nezávislý (pravý) vlastní vektor.

Protože levý vlastní vektor y matice A odpovídající vlastnímu číslu $\lambda = \varrho(A)$ je pravým vlastním vektorem matice A^T a ta je kladná, y je jediným lineárně nezávislým levým vlastním vektorem matice A, který lze volit kladný $y > 0$. Nechť tedy $x > 0$, $y > 0$. Speciálně pak $y^T x \neq 0$ a z pomocné Schurovy věty (viz větu 1) vyplývá, že $\lambda = \varrho(A)$ je jednoduché vlastní číslo matice A. □

3.4 Nezáporné nerozložitelné matice

Nyní už jsme připraveni vyslovit tzv. *Perronovu–Frobeniovu větu* o nezáporných maticích (obě předchozí Perronovy věty hovořily o maticích kladných).

Věta 7 (Perronova–Frobeniova věta (o nezáporných nerozložitelných maticích))**.** *Nechť $A \in \mathbb{R}^{n\times n}$ je čtvercová nezáporná nerozložitelná matice, tj. $a_{j,k} \geq 0$. Pak spektrální poloměr $\varrho(A)$ je kladné jednoduché vlastní číslo matice A a tomuto vlastnímu číslu odpovídá kladný vlastní vektor.*

Důkaz. Uvažujme čtvercovou nezápornou nerozložitelnou matici $A \in \mathbb{R}^{n\times n}$. Podle lemmatu 9 je matice $(I+A)^{n-1}$ kladná; zřejmě je tedy i její transpozice, tj. matice $((I+A)^{n-1})^T = (I+A^T)^{n-1}$, kladná. Podle Perronovy věty o spektrálním poloměru kladných matic (viz větu 5) existuje kladný vlastní vektor $y > 0$ tak, že

$$(I + A^T)^{n-1}\, y = y\, \varrho\Big((I + A^T)^{n-1}\Big)$$

neboli

$$y^T(I + A)^{n-1} = \varrho\Big((I + A)^{n-1}\Big)\, y^T. \tag{3.5}$$

Mějme dále vlastní číslo λ matice A takové, že

$$|\lambda| = \varrho(A).$$

Nechť $x \neq 0$ je nějaký vlastní vektor matice A odpovídající tomuto vlastnímu číslu, tj. $Ax = x\lambda$. Pak platí

$$A|x| = |A||x| \geq |Ax| = |x\lambda| = |x||\lambda|,$$

kde nerovnost získáme obdobně jako ve vztahu (2.14), použijeme-li ho na jednotlivé řádky levé a pravé strany. Tudíž platí

$$A|x| \geq |x|\varrho(A).$$

Vynásobíme-li předchozí nerovnost maticí A zleva, dostaneme

$$A^2|x| = A(A|x|) \geq A(|x|\varrho(A)) = (A|x|)\varrho(A) \geq (|x|\varrho(A))\varrho(A) = |x|\varrho^2(A).$$

Obdobně dostaneme obecný vztah

$$A^p|x| \geq |x|\varrho^p(A), \quad \text{pro} \quad p = 1, \ldots, n-1. \tag{3.6}$$

Jestliže vynásobíme nerovnost ze vztahu (3.6) číslem $\binom{n-1}{p}$ (viz (3.4)) a sečteme pro $p = 1, \ldots, n-1$, dostaneme, spolu s rovností $|x| = I|x|$, vztah

$$(I + A)^{n-1}|x| \geq |x|\Big(1 + \varrho(A)\Big)^{n-1}. \tag{3.7}$$

Vynásobíme-li tuto nerovnost zleva kladným vektorem $y^T > 0$, dostaneme

$$y^T(I + A)^{n-1}|x| \geq (y^T|x|)\Big(1 + \varrho(A)\Big)^{n-1},$$

kde levá strana je podle (3.5) rovna $\varrho((I+A)^{(n-1)})(y^T|x|)$. Protože $y^T|x|$ je kladné číslo, dostáváme

$$\varrho\Big((1 + A)^{n-1}\Big) \geq \Big(1 + \varrho(A)\Big)^{n-1}. \tag{3.8}$$

Matice $(I + A)^{n-1}$ má vlastní čísla ve tvaru $(1 + \lambda_j)^{n-1}$, kde λ_j jsou vlastní čísla matice A. Protože $(I + A)^{n-1}$ je kladná matice, její spektrální poloměr je přímo (kladným) vlastním číslem. To znamená, že existuje vlastní číslo μ matice A tak, že

$$(1 + \mu)^{n-1} = \varrho\Big((I + A)^{n-1}\Big). \tag{3.9}$$

Kombinací vztahů (3.8)–(3.9) dostaneme

$$(1+\mu)^{n-1} \geq \Big(1+\varrho(A)\Big)^{n-1},$$

neboli (po odmocnění)

$$|1+\mu| \geq 1+\varrho(A).$$

Protože platí $\varrho(A) \geq |\mu|$, pak

$$1+\varrho(A) \geq 1+|\mu| \geq |1+\mu| \geq 1+\varrho(A).$$

Vidíme, že levá strana nerovnosti je shodná s pravou, a proto i ve všech mezilehlých nerovnostech musí nastat rovnost. Speciálně odtud plyne, že

$$\varrho(A) = \mu, \quad \mu \geq 0.$$

Rovnost také platí v nerovnosti (3.8), v nerovnosti (3.7) a ve všech nerovnostech v (3.6). Zde speciálně pro $k = 1$ dostáváme

$$A|x| = |x|\varrho(A), \qquad \text{neboli} \qquad A|x| = |x|\mu.$$

Specálně také z (3.7) a (3.9) platí

$$(I+A)^{n-1}|x| = |x|(1+\mu)^{n-1} = |x|\varrho\Big((I+A)^{n-1}\Big). \tag{3.10}$$

Podle Perronovy věty (viz větu 5) je $|x|$ ve vztahu (3.10) kladný, tj. $|x| > 0$. Z věty 6 pak plyne, že k μ existuje jediný lineárně nezávislý vlastní vektor. Navíc je $\varrho(A) > 0$, protože A je nerozložitelná, a tedy nenulová matice. Protože levý vlastní vektor y matice A je pravým vlastním vektorem matice A^T a ta je nezáporná nerozložitelná, je y jediným lineárně nezávislým levým vlastním vektorem matice A a lze volit kladný. Speciálně $y^T x \neq 0$, tedy ze Schurovy pomocné věty (viz větu 1) vyplývá, že $\varrho(A)$ je jednoduché vlastní číslo matice A. □

Předchozí věta má následující důležitý důsledek.

Vidíme, že bude-li (čtvercová a nezáporná) matice hyperlinků H nejen *stochastická*, ale také *nerozložitelná*, tj. bude-li možné dostat se z každé stránky $\mathtt{P}_j$ na libovolnou stránku $\mathtt{P}_k$, pak bude (kladný) PageRank vektor π určen jednoznačně.

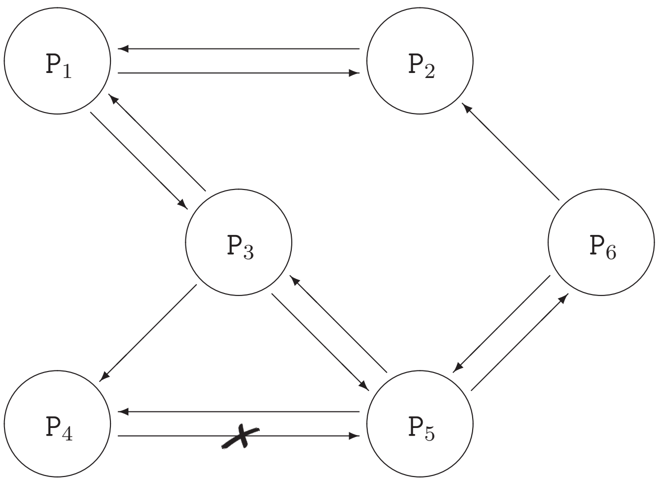
Všimněme si, že tyto vlastnosti (stochasticitu a nerozložitelnost) má právě náš modelový intenet z obrázku 1.1. Jeho (kladný) PageRank vektor bude určen jednoznačně. Vztahy mezi jednotlivými druhy matic jsou schématicky znázorněny na obrázku 6.3 na str. 62.

4 Matice hyperlinků obecně není stochastická ani nerozložitelná. Co s tím?

Ačkoliv našemu modelovému internetu z obrázku 1.1 odpovídá matice hyperlinků, která je stochastická a nerozložitelná, obecně to tak být nemusí. Tento deficit souvisí s vlastnostmi (i)–(iv), které jsme zmínili na str. 21–22. Stochasticita je důležitá z hlediska vlastnosti (iii) a nerozložitelnost požadujeme pro vlastnost (iv).

4.1 Matice hyperlinků je obecně substochastická.

Již jsme zjistili, že stochasticita souvisí s vlastností (iii), tj. každá stránka musí obsahovat alespoň jeden odkaz na jinou stránku. Podívejme se na internet, který tuto vlastnost nesplňuje, viz obrázek 4.1. Internetu z tohoto obrázku odpovídá matice H, která má nulový řádek, tudíž součet prvků v řádku není vždy roven jedné, přesněji



Obrázek 4.1: Jednoduchý model internetu z obrázku 1.1 po vyškrtání některých odkazů (hran); symbol „✗" značí vyškrtnutou hranu. Vidíme, že nyní nevede žádný odkaz ze stránky $\mathtt{P}_4$. Odpovídající matice hyperlinků je substochastická. Řádek matice H, jehož součet není roven jedné, je zvýrazněn.

řečeno, součet prvků v řádku je v intervalu $[0,1]$,

$$\underbrace{\begin{bmatrix} r(\mathtt{P}_1) \\ r(\mathtt{P}_2) \\ r(\mathtt{P}_3) \\ r(\mathtt{P}_4) \\ r(\mathtt{P}_5) \\ r(\mathtt{P}_6) \end{bmatrix}}_{\pi} = \underbrace{\left[\begin{array}{ccc|c|cc} 0 & 1/1 & 1/3 & 0 & 0 & 0 \\ 1/2 & 0 & 0 & 0 & 0 & 1/2 \\ 1/2 & 0 & 0 & 0 & 1/3 & 0 \\ 0 & 0 & 1/3 & 0 & 1/3 & 0 \\ 0 & 0 & 1/3 & 0 & 0 & 1/2 \\ 0 & 0 & 0 & 0 & 1/3 & 0 \end{array}\right]}_{H^T} \underbrace{\begin{bmatrix} r(\mathtt{P}_1) \\ r(\mathtt{P}_2) \\ r(\mathtt{P}_3) \\ r(\mathtt{P}_4) \\ r(\mathtt{P}_5) \\ r(\mathtt{P}_6) \end{bmatrix}}_{\pi}, \tag{4.1}$$

taková matice se nazývá *substochastická*. Uvědomme si, co představuje matice hyperlinků H. V podstatě říká, že uživatel internetu ze stánky $\mathtt{P}_j$ přejde na $\mathtt{P}_k$ s pravděpodobností $1/|\mathtt{P}_j|$, když $\mathtt{P}_j$ na $\mathtt{P}_k$ odkazuje, respektive s pravděpodobností 0, když $\mathtt{P}_j$ na $\mathtt{P}_k$ neodkazuje. To je konzistentní s tím, že na řádku jsou samé nuly. Problém nalezení PageRanku postavený pouze na hyperlinkové matici modeluje uživatele, který náhodně přechází ze stránky na stránku (tzv. *random surfer*) s pravděpodobností $1/|\mathtt{P}_j|$, když to jde. Celková doba, kterou tento uživatel stráví na jedné konkrétní stránce v průběhu delšího časového úseku, je meřítkem relativní důležitosti této stránky. Jestliže na jedné konkrétní stránce stráví celkově více času než na ostatních, pak je zřejmé, že se k této stránce vrací pravidelně. Tato stránka (a také všechny další, na které se uživatel vrací pravidelně) musí být důležitá a musí na ní být odkazováno dalšími jinými důležitými stránkami. Bohužel uživatel, který náhodně prochází internet, také občas narazí na stránky, které již neodkazují dále. Běžným příkladem jsou pdf soubory nebo obrázky. To jsou vrcholy grafu (tzv. *dangling node*), do kterých vede hrana, ale nevede už z nich.

Zakladatelé PageRanku tento problém vyřešili následovně. Všechny nulové řádky matice hyperlinků H nahradili řádkem e^T/n, kde $e = [1, \dots, 1]^T \in \mathbb{R}$ a n je počet stránek celého internetu. PageRank vektor π je pak levým vlastním vektorem nově vzniklé matice

$$\widetilde{H} \equiv H + D, \qquad \text{kde} \qquad D \equiv \frac{1}{n}\, de^T,$$

která už je stochastická. Vektor d (anglicky nazývaný *dangling node vector*) nabývá hodnot jedna nebo nula, konkrétně $d_j = 1$, jestliže j-tá stránka je právě tím vrcholem, z něhož nevede hrana ven. V ostatních případech je $d_j = 0$. Spektrální poloměr $\varrho(\widetilde{H}) = 1$ a vlastní vektor π je nezáporný, tj. $\pi \geq 0$. V tuto chvíli už je splněna vlastnost (iii), tzn. uživatel může, i po tom co se dostane do vrcholu, ze kterého nevede hrana ven, přecházet mezi stránkami nahodile. Nyní je matice $\widetilde{H}$ stochastická, čímž jsme vyřešili problém vrcholů bez zpětných hran (dangling node). Ovšem nemusí být ještě nerozložitelná.

4.2 Stochastická matice hyperlinků je obecně rozložitelná.

Nerozložitelnost je ekvivalentní s vlastností (iv), tj. s požadavkem, aby bylo možné dostat se z každé stránky pomocí odkazů na libovolnou jinou stránku. To ovšem obecně neplatí, viz např. intenet na obrázcích 3.1 a 3.2, resp. odpvídající matice hyperlinků (3.1) a (3.2). V závěru předchozí sekce model přisoudil uživateli schopnost přejít kamkoliv, a to dokonce z tzv. dangling node. Proč by tedy uživatel nemohl přejít kamkoliv odkudkoliv? Samozřejmě to lze. Představme si uživatele, který nejen přechází mezi stránkami pomocí odkazů, ale občas také napíše přímo adresu webové stránky. Pokud toto uživatel udělá, pak se tzv. *teleportuje* přímo na požadovanou stránku bez použití odkazů, odkud pokračuje v nahodilém přecházení mezi stránkami pomocí odkazů až do té doby, než se opět teleportuje jinam. Tuto teleportaci, tj. schopnost uživatele přecházet odkudkoliv kamkoliv, vyřešili Page a Brin matematicky tak, že stochastickou matici $\widetilde{H}$ opět upravili a získali tzv. *googlovskou matici*

$$G \equiv \alpha\widetilde{H} + (1-\alpha)E, \qquad \text{kde} \qquad E \equiv \frac{1}{n}ee^T$$

je tzv. *teleportační matice* a

$$0 < \alpha < 1.$$

Číslo α zde udává poměr mezi časem, který běžný uživatel stráví přecházením mezi stránkami pomocí odkazů, a časem, který věnuje teleportaci. Bude-li $\alpha = 0,8$, pak 80 % času uživatel přechází náhodně mezi stránkami pomocí odkazů a 20 % času uživatel tráví přechodem přímo na nové stránky zadáním jejich webové adresy. Teleportace je náhodná, protože uživatel může kdykoliv přejít náhodně na kteroukoliv jinou stránku tak, že napíše její webovou adresu. Googlovská matice G je *konvexní kombinací* dvou stochastických matic, opravené matice hyperlinků $\widetilde{H}$ a teleportační matice E, která je navíc kladná,

$$\widetilde{H} = H + \frac{1}{n}de^T, \qquad E = \frac{1}{n}ee^T.$$

Protože $\alpha < 1$, pak $(1-\alpha) > 0$, a tedy googlovská matice G je určitě *kladná*, $G \geq 0$. Podle následujícího lemmatu je googlovská matice navíc stochastická.

Lemma 10. *Nechť S_1 a S_2 jsou stochastické matice a $\alpha \in [0,1]$. Pak konvexní kombinací těchto dvou stochastických matic vznikne stochastická matice S, tj.*

$$S = \alpha S_1 + (1-\alpha)S_2.$$

Důkaz. Důkaz je elementární. Stačí si uvědomit, že v j-tém řádku matice S je součet j-tého řádku matice S_1 (který je roven jedné) vynásobený α a j-tého řádku matice S_2 (který je také roven jedné) vynásobený $(1-\alpha)$. □

Tím jsme se dostali k nejdůležitějšímu pozorování.

Googlovská matice

$$G = \alpha\Big(H + \frac{1}{n}de^T\Big) + (1-\alpha)\frac{1}{n}ee^T, \quad 0 < \alpha < 1$$

je stochastická a nerozložitelná (přesněji řečeno kladná, všechny její prvky jsou nenulové), takže jsou splněny vlastnosti (iii) a (iv), viz stranu 22. Tudíž je možné dostat se z každé stránky pomocí *odkazů* nebo *teleportací* na libovolnou jinou stránku, tj. *kladný PageRank vektor pro googlovskou matici existuje a je jednoznačný.*

Pro vyjasnění vztahů mezi jednotlivými druhy matic viz schéma na obrázku 6.3 na str. 62.

Část II

Výpočet PageRank vektoru

5 Mocninná metoda

Když už víme, že PageRank vektor existuje a je jednoznačný, rádi bychom ho uměli spočítat. Výpočet budeme provádět pomocí tzv. *mocninné metody* (*power method*). Ta slouží k nalezení vlastního čísla matice A, které je v absolutní hodnotě největší (tzv. dominantní vlastní číslo), resp. k nalezení vlastního vektoru, který mu odpovídá (tj. k nalezení dominantního vlastního páru).

Nejprve si ukážeme, kdy mocninná metoda funguje a zjistíme, že pro hyperlinkovou matici tato metoda obecně platit nemusí. Nakonec ovšem budeme schopni upravit matici tak, aby se tato metoda dala použít.

5.1 Diagonalizovatelné matice

Pro jednoduchost se v následujícím výkladu omezíme jen na tzv. diagonalizovatelné matice, začneme definicí, viz např. [11, kapitola 2].

Definice 12 (Diagonalizovatelná matice)**.** *Matice $A \in \mathbb{C}^{n \times n}$ se nazývá diagonalizovatelná matice, pokud existuje regulární matice $X \in \mathbb{C}^{n \times n}$ taková, že*

$$D \equiv X^{-1} A X \in \mathbb{C}^{n \times n}$$

je diagonální.

Diagonalizovatelná matice je tedy taková čtvercová matice, která neobsahuje žádné *netriviální Jordanovy bloky* (tj. dimenze 2×2 a větší), viz [11, kapitoly 1.8, 2.2]. Bez újmy na obecnosti můžeme předpokládat, že sloupce x_j matice X jsou normalizované, tedy

$$X = [x_1, \ldots, x_n] \in \mathbb{C}^{n \times n} \qquad \text{kde} \qquad \|x_j\| = 1, \quad j = 1, \ldots, n.$$

Označme dále λ_j diagonální prvky matice D, tj.

$$D = \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix} \in \mathbb{C}^{n \times n}.$$

Matici A pak můžeme psát ve tvaru rozkladu

$$A = X D X^{-1}. \tag{5.1}$$

Označme dále

$$Y \equiv X^{-H} = (X^H)^{-1} = (\overline{X}^T)^{-1}, \qquad Y = [y_1, \ldots, y_n],$$

tj.

$$Y^H X = X^{-1} X = I \qquad \text{a} \qquad y_k^H x_j = \begin{cases} 1, & \text{když } k = j \\ 0, & \text{když } k \neq 0 \end{cases}.$$

Pomocí (regulární) matice Y můžeme rozklad $A = XDX^{-1}$ (5.1) přepsat ve tvaru $A = Y^{-H} D Y^H$ a získat tak rozklad matice A^H v analogickém tvaru, tj.

$$A^H = Y \overline{D} Y^{-1}. \tag{5.2}$$

Maticové rovnosti (5.1) a (5.2) lze také přepsat do tvaru

$$\begin{aligned} AX &= XD, \quad \text{resp.} \quad A^H Y = Y\overline{D}, \quad Y^H A = DY^H, \quad \text{tj.} \\ Ax_j &= x_j \lambda_j, \quad \text{resp.} \quad A^H y_j = y_j \overline{\lambda_j}, \quad y_j^H A = \lambda_j y_j^H, \quad \text{kde} \quad j = 1, \ldots, n. \end{aligned}$$

Vidíme, že x_j jsou pravé vlastní vektory, y_j levé vlastní vektory a λ_j vlastní čísla matice A. Pravé vlastní vektory jsou normalizovány $\|x_j\| = 1$ pomocí nějaké normy $\| \cdot \|$ (záměrně nespecifikujeme jaké). Levé vlastní vektory jsou pak normalizovány tak, aby platilo $y_j^H x_j = 1$. Užitím obou sad vlastních vektorů můžeme matici A zapsat pomocí tzv. *dyadického rozvoje* následujícím způsobem. Zřejmě platí

$$\begin{aligned} A = XDY^H &= [x_1, \ldots, x_n] \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix} \begin{bmatrix} y_1^H \\ \vdots \\ y_n^H \end{bmatrix} \\ &= [x_1\lambda_1, \ldots, x_n\lambda_n] \begin{bmatrix} y_1^H \\ \vdots \\ y_n^H \end{bmatrix} = \sum_{j=1}^{n} x_j \lambda_j y_j^H. \end{aligned} \tag{5.3}$$

Rozklad $A = XDX^{-1}$, resp. $A = XDY^H$, se někdy nazývá *spektrální rozklad* matice A.

5.2 Mocninná metoda pro diagonalizovatelné matice

Uvažujme diagonalizovatelnou matici $A \in \mathbb{C}^{n \times n}$ a její spektrální rozklad, resp. dyadický rozvoj $A = XDX^{-1} = \sum_{j=1}^{n} x_j \lambda_j y_j^H$. Protože matice A je diagonalizovatelná, její pravé (a také levé) vlastní vektory x_j (resp y_j) tvoří bázi celého $\mathbb{C}^n$. Nechť $v \in \mathbb{C}^n$ je libovolný vektor a ν_j, $j = 1, \ldots, n$, jeho souřadnice v bázi x_j, tj.

$$v = \sum_{j=1}^{n} x_j \nu_j = X \begin{bmatrix} \nu_1 \\ \vdots \\ \nu_n \end{bmatrix}.$$

Zřejmě platí

$$Av = (XDX^{-1})v = XDX^{-1}X \begin{bmatrix} \nu_1 \\ \vdots \\ \nu_n \end{bmatrix} = X \begin{bmatrix} \lambda_1\nu_1 \\ \vdots \\ \lambda_n\nu_n \end{bmatrix} = \sum_{j=1}^{n} x_j(\lambda_j\nu_j).$$

Obdobně, pro libovolné přirozené k platí

$$A^\ell v = (XDX^{-1})(A^{\ell-1}v) = XDX^{-1}X \begin{bmatrix} \lambda_1^{\ell-1}\nu_1 \\ \vdots \\ \lambda_n^{\ell-1}\nu_n \end{bmatrix} = \sum_{j=1}^{n} x_j(\lambda_j^\ell\nu_j).$$

Označme λ_1 vlastní číslo s největší absolutní hodnotou a předpokládejme, že je nenulové (tedy, že matice A je nenulová), tj.

$$|\lambda_1| = \varrho(A), \quad \lambda_1 \neq 0; \quad \text{předpokládejme také} \quad \nu_1 \neq 0.$$

Předpoklad $\nu_1 \neq 0$ není při praktických výpočtech (tj. v aritmetice s konečnou přesností) příliš důležitý a je téměř vždy vynucen zaokrouhlovacími chybami, viz [15, kapitola 7.3.1]. Pak zřejmě

$$Av = x_1(\lambda_1\nu_1) + \sum_{j=2}^{n} x_j(\lambda_j\nu_j) = \left(x_1 + \sum_{j=2}^{n} x_j \left(\frac{\lambda_j}{\lambda_1}\right) \left(\frac{\nu_j}{\nu_1}\right) \right) (\lambda_1\nu_1),$$

resp.

$$A^\ell v = \left(x_1 + \sum_{j=2}^{n} x_j \left(\frac{\lambda_j}{\lambda_1}\right)^\ell \left(\frac{\nu_j}{\nu_1}\right) \right) (\lambda_1^\ell\nu_1), \quad \ell = 1, 2, 3, \ldots.$$

Protože vlevo je mocnina matice, mohlo by se stát, že prvky vektoru $A^\ell v$, resp. jejich absolutní hodnoty, porostou do nekonečna. Abychom tomu zabránili, budeme vektor normalizovat, neboli

$$\frac{A^\ell v}{\|A^\ell v\|} = \left(x_1 + \sum_{j=2}^{n} x_j \left(\frac{\lambda_j}{\lambda_1}\right)^\ell \left(\frac{\nu_j}{\nu_1}\right) \right) \frac{\lambda_1^\ell\nu_1}{\|A^\ell v\|}, \tag{5.4}$$

kde zlomek nalevo je vektor délky (resp. normy) jedna a zlomek napravo je skalár.

Pokud platí

$$|\lambda_1| > |\lambda_j| \qquad \text{pro} \qquad j = 2, \ldots, n,$$

pak zřejmě

$$\left|\frac{\lambda_j}{\lambda_1}\right| < 1,$$

a tedy

$$\lim_{\ell\longrightarrow\infty} \left(\frac{\lambda_j}{\lambda_1}\right)^\ell = 0 \qquad \text{a také} \qquad \lim_{\ell\longrightarrow\infty} \sum_{j=2}^{n} x_j \left(\frac{\lambda_j}{\lambda_1}\right)^\ell \left(\frac{\nu_j}{\nu_1}\right) = 0.$$

A proto za předpokladu, že matice A má jediné (a jednoduché) vlastní číslo, pro které platí $|\lambda_1| = \varrho(A)$, pak

$$\lim_{\ell \longrightarrow \infty} \frac{A^\ell v}{\|A^\ell v\|} = \lim_{\ell \longrightarrow \infty} x_1 \frac{\lambda_1^\ell \nu_1}{\|A^\ell v\|}.$$

Posloupnost

$$\frac{\lambda_1^\ell \nu_1}{\|A^\ell v\|}$$

obecně limitu nemá. Stačí si uvědomit, co se děje, když je λ_1 záporné. Protože jsou ale velikosti (resp. normy) obou vektorů $\frac{A^\ell v}{\|A^\ell v\|}$ a x_1 rovny jedné, platí

$$\lim_{\ell \longrightarrow \infty} \left| \frac{\lambda_1^\ell \nu_1}{\|A^\ell v\|} \right| = 1.$$

Pokud budeme každý vektor $\frac{A^\ell v}{\|A^\ell v\|}$ navíc násobit nějakou vhodně zvolenou komplexní jednotkou η_ℓ ($|\eta_\ell| = 1$, $\eta_\ell = \exp(\mathbf{i}\,\varphi_\ell)$), např. tak aby první nenulová složka vektoru zůstávala kladná, pak

$$\lim_{\ell \longrightarrow \infty} \frac{A^\ell v}{\|A^\ell v\|}\,\eta_\ell = x_1.$$

Pokud libovolný vektor v násobíme stále dokola maticí A a mezivýsledky (šikovně) normalizujeme, proces bude konvergovat k (pravému) vlastnímu vektoru x_1, který odpovídá vlastnímu číslu λ_1.

5.3 Mocninná metoda pro obecné čtvercové matice

Předpoklad diagonalizovatelnosti v předchozí sekci byl pouze z důvodu zjednodušení výkladu. Mocninnou metodu zformulujeme v následující větě pro obecné čtvercové matice.

Věta 8. *Nechť $A \in \mathbb{C}^{n\times n}$ je libovolná čtvercová matice, $\lambda_1, \ldots, \lambda_n$ její vlastní čísla, přičemž*

$$|\lambda_1| = \varrho(A),$$

a x_1 je odpovídající vlastní vektor, tj. $Ax_1 = x_1\lambda_1$. Nechť $v \in \mathbb{C}^n$ je libovolný vektor. Pokud platí

$$|\lambda_1| > |\lambda_j|, \quad j = 2, \ldots, n \qquad \text{a zároveň} \qquad \nu_1 \neq 0,$$

pak

$$\lim_{\ell \longrightarrow \infty} \frac{A^\ell v}{\|A^\ell v\|}\,\eta_\ell = x_1$$

pro nějaká vhodně volená η_ℓ, $|\eta_\ell| = 1$.

Pro diagonalizovatelné matice je věta dokázána již v předchozí sekci. V důkazu pro obecné matice by bylo potřeba zavést tzv. *zobecněné vlastní vektory* a uvažovat *netriviální Jordanovy bloky* (tj. dimenze 2×2 a větší) a sledovat, jak se chovají jejich mocniny, viz např. [11, kapitoly 1.8, 2.4]. Postup je jinak stejný jako v předchozí sekci, jen techničtější.

Věta opět obsahuje formální přepoklad

$$\nu_1 \neq 0,$$

který není při praktických výpočtech (tj. v aritmetice s konečnou přesností) příliš důležitý ani v případě obecných matic. Opět je téměř vždy vynucen vlivem zaokrouhlovacích chyb, viz [15, kapitola 7.3.1]. Z tohoto důvodu se také příliš nezabýváme významem čísla ν_1 pro obecnou matici A (v předchozí sekci představovala čísla ν_j souřadnice v bázi x_j vlastních vektorů matice A; vlastní vektory obecné matice ale obecně netvoří bázi $\mathbb{C}^n$, museli bychom si vzít na pomoc právě výše zmíněné zobecněné vlastní vektory).

Obecně můžeme algoritmus mocninné metody shrnout do následujících schématických kroků:

- Dáno: obecná čtvercová matice $A \in \mathbb{C}^{n \times n}$.
- Vyber nebo vytvoř náhodný vektor $v \in \mathbb{C}^n$.
- Opakuj kroky:
- $\quad v \leftarrow Av$,
- $\quad v \leftarrow v/\|v\|$, kde $\|\cdot\|$ je nějaká vhodně zvolená norma,
- dokud nejsi s vektorem v spokojen.

Ze vztahu (5.4) je zřejmé, že mocninná metoda bude konvergovat k vlastnímu vektoru x_1 tak rychle, jak rychle půjde

$$\max_{j=2,\ldots,n} \left| \frac{\lambda_j}{\lambda_1} \right|^{\ell}, \quad \ell = 1, 2, 3, \ldots$$

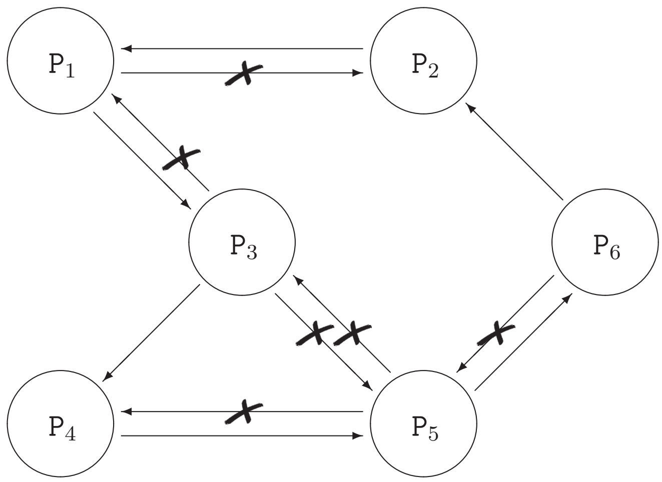
k nule. Pro další detaily, viz [15, kapitola 7.3.1], [17, kapitola 15].

6 Výpočet Perronova vlastního vektoru

Z výkladu o mocinné metodě vidíme, že na to, abychom mohli PageRank vektor spočítat (touto metodou), musí být jemu odpovídající vlastní číslo $\lambda_1 = 1$ jednoduché (což v případě googlovské matice G, $G > 0$, je, viz větu 6) a v absolutní hodnotě ostře větší než všechna ostatní vlastní čísla. Zda tomu tak je, budeme nejprve ilustrovat na příkladech.

6.1 (Im)primitivní matice

Vezměme si ku pomoci opět náš modelový internet, ze kterého vyškrtáme všechny hrany tak, aby matice jemu odpovídající zůstala stochastická a nerozložitelná, viz obrázek 6.1. Tomuto internetu odpovídá matice hyperlinků H, která má v každém řádku a v každém sloupci jediný nenulový prvek. Rovnice pro výpočet PageRank



Obrázek 6.1: Jednoduchý model internetu z obrázku 1.1 po vyškrtání některých odkazů (hran); symbol „✗" značí vyškrtnutou hranu. Matice hyperlinků, která mu odpovídá, je stochastická (každá stránka obsahuje alespoň jeden odkaz) a nerozložitelná (z každé stránky je možné se dostat na kteroukoliv jinou stránku).

vektoru z této matice hyperlinků pak vypadá následovně

$$\underbrace{\begin{bmatrix} r(\mathtt{P}_1) \\ r(\mathtt{P}_2) \\ r(\mathtt{P}_3) \\ r(\mathtt{P}_4) \\ r(\mathtt{P}_5) \\ r(\mathtt{P}_6) \end{bmatrix}}_{\pi} = \underbrace{\begin{bmatrix} 0 & 1/1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1/1 \\ 1/1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1/1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1/1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1/1 & 0 \end{bmatrix}}_{H^T} \underbrace{\begin{bmatrix} r(\mathtt{P}_1) \\ r(\mathtt{P}_2) \\ r(\mathtt{P}_3) \\ r(\mathtt{P}_4) \\ r(\mathtt{P}_5) \\ r(\mathtt{P}_6) \end{bmatrix}}_{\pi}. \tag{6.1}$$

Po permutaci vrcholů grafu (přečíslování stránek) permutací

$$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 1 & 6 & 2 & 3 & 4 & 5 \end{pmatrix}$$

tak, jak je naznačeno na obrázku 6.2, dostaneme matici

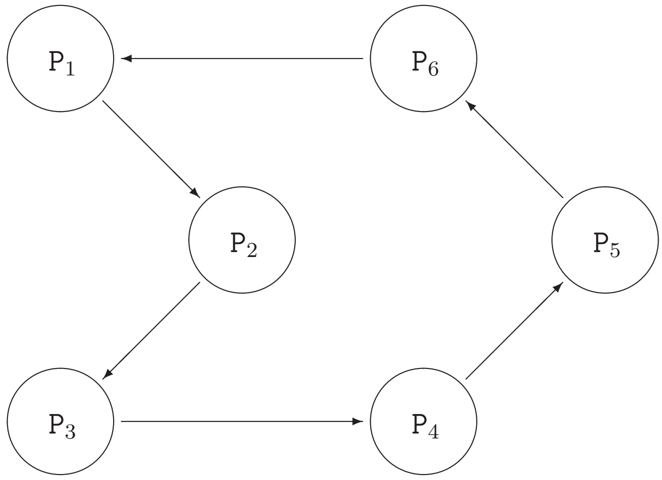
$$S \equiv \Pi H \Pi^T = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}, \quad \text{kde} \quad \Pi = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 & 0 \end{bmatrix}$$

je permutační matice. Zřejmě je tato matice S stochastická a nerozložitelná (z každého vrcholu odpovídajícího grafu je možné dostat se do všech ostatních vrcholů). Podívejme se, jak budou vypadat její vlastní čísla. Zřejmě

$$\begin{aligned} \det(S - \lambda I) &= 0, \\ \det\left(\begin{bmatrix} -\lambda & 1 & 0 & 0 & 0 & 0 \\ 0 & -\lambda & 1 & 0 & 0 & 0 \\ 0 & 0 & -\lambda & 1 & 0 & 0 \\ 0 & 0 & 0 & -\lambda & 1 & 0 \\ 0 & 0 & 0 & 0 & -\lambda & 1 \\ 1 & 0 & 0 & 0 & 0 & -\lambda \end{bmatrix}\right) &= 0, \\ -\lambda \det\left(\begin{bmatrix} -\lambda & 1 & 0 & 0 & 0 \\ 0 & -\lambda & 1 & 0 & 0 \\ 0 & 0 & -\lambda & 1 & 0 \\ 0 & 0 & 0 & -\lambda & 1 \\ 0 & 0 & 0 & 0 & -\lambda \end{bmatrix}\right) - 1 \det\left(\begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ -\lambda & 1 & 0 & 0 & 0 \\ 0 & -\lambda & 1 & 0 & 0 \\ 0 & 0 & -\lambda & 1 & 0 \\ 0 & 0 & 0 & -\lambda & 1 \end{bmatrix}\right) &= 0, \\ \lambda^6 - 1 &= 0, \\ \lambda^6 &= 1, \end{aligned}$$

tedy

$$\lambda_{1,2} = \pm 1, \qquad \lambda_{3,4,5,6} = \pm\frac{1}{2} \pm \frac{\sqrt{3}}{2}\,\mathbf{i}\,.$$



Obrázek 6.2: Internet z obrázku 6.1 po permutaci očíslování stránek.

Vidíme, že $|\lambda_1| = |\lambda_2| = |\lambda_3| = |\lambda_4| = |\lambda_5| = |\lambda_6|$, a proto v tomto případě nebude mocninná metoda fungovat, resp. konvergovat. To je způsobeno tím, že matice hyperlinků je tzv. *imprimitivní*. Z obrázku 6.2 je zřejmé, že graf tohoto internetu je cyklický (přesněji bychom řekli, že tvoří tzv. *kružnici*; v terminologii Markových řetězeců bychom řekli, že se jedná o *periodický* řetězec). Právě (a)cykličnost grafu spolu s jeho silnou souvislostí (nerozložitelností odpovídající nezáporné matice) definuje *(im)primitivní* matice. Přesná definice následuje. Definici i následující věty lze nalézt např. v [17, str. 172–175].

Definice 13 ((Im)primitivní matice, index imprimitivity)**.** *Nezáporná nerozložitelná matice* $A \in \mathbb{R}^{n\times n}$ *se nazývá primitivní, pokud na spektrální kružnici* $\mathscr{K}_\odot(A) \equiv \{\varrho(A)\exp(\mathbf{i}\,\varphi) \;:\; 0 \leq \varphi < 2\pi\}$*, tj. na kružnici v komplexní rovině se středem v počátku a o poloměru* $\varrho(A)$*, leží jediné vlastní číslo* $\lambda = \varrho(A)$*. Naopak nezáporná nerozložitelná matice* $A \in \mathbb{R}^{n\times n}$*, která není primitivní, se nazývá imprimitivní.*

Počet vlastních čísel ležících na spektrální kružnici $\mathscr{K}_\odot(A)$ *se nazývá index imprimitivity matice* A*.*

6.2 Spektrální vlastnosti (im)primitivních matic

Primitivita je z hlediska PageRanku velmi významná. Pokud by matice byla imprimitivní, nefungovala by mocninná metoda a nebyli bychom schopni PageRank spočítat, a proto si uvedeme ještě několik vět, z nichž některé uvedeme bez úplných důkazů. První a nejdůležitější z těchto vět popisuje, jak jsou vlastní čísla rozmístěna na spektrální kružnici, viz např. [17, str. 173] nebo [14, Vol. 2, kapitola XIII, §2, str. 53–66] (případně původní ruské vydání [13, kapitola XIII, §2, str. 354–365]).

Věta 9. *Nechť $A \in \mathbb{R}^{n\times n}$ je nezáporná nerozložitelná matice a nechť h je její index imprimitivity. Pak*

$$\lambda_1, \quad \lambda_2 \equiv \lambda_1\omega, \quad \lambda_3 \equiv \lambda_1\omega^2, \quad \ldots, \quad \lambda_h \equiv \lambda_1\omega^{h-1},$$

kde

$$\lambda_1 \equiv \varrho(A) \in \mathbb{R}, \quad \omega \equiv \exp(2\pi\mathbf{i}/h) \in \mathbb{C},$$

jsou jednoduchá vlastní čísla matice A.

Pokud navíc $h > 1$, existuje permutační matice Π taková, že

$$\Pi A \Pi^T = \begin{array}{l} \left[\begin{array}{cccccc} 0 & A_{1,2} & & & & \\ & 0 & A_{2,3} & & & \\ & & \ddots & \ddots & & \\ & & & & 0 & A_{h-1,h} \\ A_{h,1} & & & & & 0 \end{array}\right] \begin{array}{l} \}n_1 \\ \}n_2 \\ \vdots \\ \}n_{h-1} \\ \}n_h \end{array} \\ \quad\underbrace{\quad}_{n_1}\ \underbrace{\quad}_{n_2}\ \underbrace{\quad}_{n_3}\ \cdots\ \underbrace{\quad}_{n_h} \end{array}, \quad \sum_{j=1}^{h} n_j = n, \tag{6.2}$$

tj. diagonální (vyznačené nulové) bloky jsou čtvercové a jediné nenulové bloky jsou nezáporné (obecně obdélníkové) bloky $A_{j,j+1} \in \mathbb{R}^{n_j \times n_{j+1}}$, $j = 1, \ldots, h-1$, a $A_{h,1} \in \mathbb{R}^{n_h \times n_1}$.

Důkaz. Důkaz jen naznačíme. Úplný důkaz lze nalézt např. v knize [14]. Nezáporná nerozložitelná matice s indexem imprimitivity h má právě h vlastních čísel na spektrální kružnici. Nechť jsou to

$$\lambda_\ell \equiv \varrho(A)\exp(\mathbf{i}\,\varphi_\ell), \qquad \ell = 1, \ldots, h.$$

Podle věty 7 je jedním z nich přímo spektrální poloměr. Toto vlastní číslo je navíc jednoduché. Nechť tedy např.

$$\lambda_1 = \varrho(A) \qquad \text{neboli} \qquad \varphi_1 \equiv 0,$$

a nechť dále

$$\varphi_1 < \varphi_2 \leq \varphi_3 \leq \ldots \leq \varphi_h < 2\pi.$$

Uvažujme dále vlastní vektor $x_\ell \in \mathbb{C}^n$ příslušný vlastnímu číslu λ_ℓ, tj. $Ax_\ell = x_\ell\lambda_\ell$, $x_\ell = [\xi_1, \ldots, \xi_n] \neq 0$, kde $\xi_j = |\xi_j|\exp(\mathbf{i}\,\psi_j)$, pak

$$x_\ell = D_\ell|x_\ell|, \qquad \text{kde} \qquad D_\ell = \begin{bmatrix} \exp(\mathbf{i}\,\psi_1) & & \\ & \ddots & \\ & & \exp(\mathbf{i}\,\psi_n) \end{bmatrix}. \tag{6.3}$$

Zřejmě $D_\ell\overline{D_\ell} = |D_\ell| = I$. Takže

$$\begin{aligned} Ax_\ell &= x_\ell\lambda_\ell, \\ AD_\ell|x_\ell| &= D_\ell|x_\ell|\varrho(A)\exp(\mathbf{i}\,\varphi_\ell), \\ \overline{D_\ell}AD_\ell|x_\ell| &= |x_\ell|\varrho(A)\exp(\mathbf{i}\,\varphi_\ell), \\ \exp(-\mathbf{i}\,\varphi_\ell)\overline{D_\ell}AD_\ell|x_\ell| &= |x_\ell|\varrho(A), \\ M|x_\ell| &= |x_\ell|\varrho(A), \end{aligned} \tag{6.4}$$

kde

$$M \equiv \exp(-\mathbf{i}\,\varphi_\ell)\overline{D_\ell}AD_\ell \in \mathbb{C}^{n\times n}, \qquad M|x_\ell| \in \mathbb{R}^n \tag{6.5}$$

je obecně komplexní matice, resp. reálný vektor. Splňují tedy předpoklady prvního tvrzení lemmatu 6. Užitím nerovnosti (2.8) z lemmatu 6 (spolu s (6.4)) dostáváme

$$|M|\,|x_\ell| \geq M|x_\ell| = |x_\ell|\varrho(A). \tag{6.6}$$

Protože matice D_ℓ je diagonální s komplexními jednotkami na diagonále, platí zřejmě

$$|M| = A, \tag{6.7}$$

viz (6.3) a (6.5). Nerovnost (6.6) tak může být dále upravena

$$|M|\,|x_\ell| = A|x_\ell| \geq |x_\ell|\varrho(A).$$

Protože ale nezáporná matice A nemůže žádný nezáporný vektor $|x_\ell|$ *zesílit po komponentách* více, než zesílí vlastní vektor $x_1 > 0$ odpovídající největšímu (kladnému) vlastnímu číslu $\lambda_1 = \varrho(A)$, musí v předchozích vztazích nastat rovnosti. Speciálně

$$|M|\,|x_\ell| = |x_\ell|\varrho(A).$$

Z předchozí rovnice a z rovnice (6.4) pak plyne, že $|x_\ell|$ je vlastním vektorem matice $|M|$ (připomeňme, že $|M| = A$) a zároveň matice M, který odpovídá v obou případech stejnému vlastnímu číslu. Tedy

$$|M|\,|x_\ell| = M|x_\ell|. \tag{6.8}$$

Díky tomu, že $|x_\ell|$ je také vlastním vektorem matice A odpovídající vlastnímu číslu $\lambda_1 = \varrho(A)$, je, podle věty 7, kladný, tj. $|x_\ell| > 0$. Podle druhého tvrzení lemmatu 6 (viz (2.9), kde $A \equiv |x_\ell|^T > 0$, $B \equiv M^T$) dostáváme, že rovnost (6.8) může být splněna jen tehdy, když

$$|M| = M.$$

Tato rovnice spolu s rovnicemi (6.5) a (6.7) dává nejdůležitější identitu

$$A = \exp(-\mathbf{i}\,\varphi_\ell)\overline{D_\ell}AD_\ell, \qquad \ell = 1, \ldots, h.$$

Uvažujme tyto dvě rovnosti, např.

$$\begin{aligned} &A = \exp(-\mathbf{i}\,\varphi_{\ell_1})\overline{D_{\ell_1}}AD_{\ell_1} && \text{a} && \\ &A = \exp(-\mathbf{i}\,\varphi_{\ell_2})\overline{D_{\ell_2}}AD_{\ell_2} && \Longleftrightarrow && \exp(\mathbf{i}\,\varphi_{\ell_2})D_{\ell_2}A\overline{D_{\ell_2}} = A \end{aligned}$$

pro $\ell_1 = 1, \ldots, h$, $\ell_2 = 1, \ldots, h$. Jejich kombinací, přesněji dosazením druhé (v obou variantách) do první, dostáváme

$$\begin{aligned} A &= \exp(-\mathbf{i}\,(\varphi_{\ell_1} + \varphi_{\ell_2}))\overline{D_{\ell_1}D_{\ell_2}}AD_{\ell_2}D_{\ell_1}, \quad \text{resp.} \\ A &= \exp(-\mathbf{i}\,(\varphi_{\ell_1} - \varphi_{\ell_2}))\overline{D_{\ell_1}}D_{\ell_2}A\overline{D_{\ell_2}}D_{\ell_1}. \end{aligned}$$

Zde již důkaz nepatrně zjednodušíme. Lze ukázat, že

$$\varphi_{\ell_1} \pm \varphi_{\ell_2} \in \{\varphi_1, \dots, \varphi_h\} \qquad \text{a} \qquad D_{\ell_2} D_{\ell_1}, \overline{D_{\ell_2}} D_{\ell_1}, \in \{D_1, \dots, D_h\}$$

a také, že

$$\varrho(A) \exp(\mathbf{i}\,(\varphi_{\ell_1} \pm \varphi_{\ell_2})) \in \{\lambda_1, \dots, \lambda_h\}.$$

Úhly φ_ℓ tudíž tvoří konečnou aditivní Abelovu (komutativní) grupu řádu h. Platí proto $h\varphi_\ell = \varphi_1 \equiv 0$. Vlastní čísla λ_ℓ, $\ell = 1, \dots, h$ leží na vrhcholech pravidelného h-úhelníku, tedy

$$\varphi_2 = 2\pi/h, \qquad \varphi_\ell = (\ell - 1)\varphi_2 = 2\pi(\ell - 1)/h, \qquad \ell = 3, \dots, h.$$

Takže platí

$$\lambda_\ell = \varrho(A) \exp(\mathbf{i}\,\varphi_\ell) = \varrho(A)\omega^{(\ell-1)}, \qquad \omega \equiv 2\pi\mathbf{i}/h, \qquad \ell = 1, \dots, h,$$

přičemž tato vlastní čísla jsou různá, takže jednoduchá, což jsme chtěli dokázat. Podobně se dá ukázat, že dokonce celé spektrum (všech n vlastních čísel) se nezmění při pootočení komplexní roviny o libovolný celočíselný násobek úhlu $2\pi/h$.

Podobně také zjistíme, že matice D_ℓ tvoří konečnou multiplikativní Abelovu (komutativní) grupu řádu h. Platí tedy $D_\ell^h = D_1 \equiv I$. Srovnáním diagonálních prvků matice D_2 podle jednotlivých úhlů ψ_j získáme permutaci, která je zmiňována v druhé části věty. Pro podrobnosti viz [14, Vol. 2, str. 61–62]. □

6.3 Testování (im)primitivity matice

Nezáporná nerozložitelná matice má obecně h *jednoduchých* vlastních čísel, jejichž absolutní hodnota je rovna spektrálnímu poloměru. Tato čísla jsou rozmístěna pravidelně na spektrální kružnici $\mathscr{K}_\odot(A)$, přičemž jedno z nich je kladné. Jsou to tedy kořeny binomické rovice

$$\lambda^h = \Big(\varrho(A)\Big)^h.$$

Aby mocninná metoda fungovala, potřebujeme, aby na spektrální kružnici leželo jediné vlastní číslo. Jestliže máme nezápornou nerozložitelnou matici, je pro nás velice důležité zjistit, zda je primitivní, tj. zda $h = 1$. K tomuto účelu slouží dva základní testy, které jsou zformulované v následujících větách. První je velmi snadný, viz např. [17, str. 174].

Věta 10. *Nechť $A \in \mathbb{R}^{n\times n}$ je nezáporná nerozložitelná matice. Pokud existuje alespoň jeden pozitivní prvek na diagonále, pak je matice A primitivní.*

Důkaz. Důkaz je jednoduchý. Pokud je matice imprimitivní, čili je její index imprimitivity h větší než jedna, tj. $h > 1$, pak podle věty 9 existuje permutační matice Π taková, že simultánní permutace řádků a sloupců $\Pi A \Pi^T$ vede na matici, která má na diagonále právě h nulových čtvercových bloků. Speciálně všechny diagonální prvky permutované matice $\Pi A \Pi^T$ jsou nulové.

Při simultánní permutaci řádků a sloupců zůstávají ale diagonální prvky stále na diagonále. Pokud má matice A na diagonále alespoň jeden nenulový prvek, pak matice $\Pi A \Pi^T$ má na diagonále také alespoň jeden nenulový prvek, pro libovolnou permutační matici Π. Matice A tedy nemůže být imprimitivní. □

Tento test je velmi snadný, avšak často nepoužitelný. Podmínka nezáporných prvků na diagonále je pouze *postačující, nikoliv nutná*. Příkladem může být právě naše matice hyperlinků (1.9), tj.

$$H = \begin{bmatrix} 0 & 1/2 & 1/2 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 1/3 & 0 & 0 & 1/3 & 1/3 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1/3 & 1/3 & 0 & 1/3 \\ 0 & 1/2 & 0 & 0 & 1/2 & 0 \end{bmatrix}. \tag{6.9}$$

Jak uvidíme za malou chvíli, tato matice hyperlinků skutečně primitivní je, má však všechny diagonální prvky nulové, tudíž test primitivity, který je založen na větě 10, použít nemůžeme. Poznamenejme ještě, že pro matici hyperlinků platí následující postačující podmínka. Je-li matice hyperlinků nerozložitelná (nezáporná je vždy), pak je primitivní právě tehdy, když v internetu existuje alespoň jedna stránka, která odkazuje sama na sebe. Nakonec uvedeme větu, která formuluje *nutnou a postačující podmínku* primitivity, viz např. [17, str. 174].

Věta 11. *Nechť $A \in \mathbb{R}^{n \times n}$ je nezáporná matice. Matice A je primitivní tehdy a jen tehdy, když existuje takové ℓ, že $A^\ell > 0$, tj. pokud v grafu $\vec{G}(A)$ existuje cesta mezi libovolnými dvěma uzly stejné délky ℓ.*

Důkaz. Důkaz opět jen naznačíme. Podle věty 9 lze imprimitivní matice s indexem primitivity $h > 1$ simultánně přepermutovat do tvaru (6.2). Simultánní permutace řádků a sloupců matice představuje jen přečíslování vrcholů grafu. Můžeme tedy bez újmy na obecnosti uvažovat matici přímo ve tvaru (6.2). Vrcholy množiny $\mathscr{I}_1 \equiv \{1, \ldots, n_1\}$ mohou být pospojovány pouze cestami délek h, $2h$, $3h$, atd. Z vrcholů množiny $\mathscr{I}_1$ do vrcholů množiny $\mathscr{I}_2 \equiv \{n_1 + 1, \ldots, n_1 + n_2\}$ však mohou existovat pouze cesty délek 1, $1 + h$, $1 + 2h$, $1 + 3h$, atd.

Cesty stejných délek tak mohou existovat jen tehdy, když $h = 1$, tj. když je matice primitivní. □

Všimněme si, že věta 11 nevyžaduje nerozložitelnost matice. Pokud ale platí $A^\ell > 0$ pro nějaké ℓ, pak je A nerozložitelná, a naopak pokud je matice rozložitelná, tj. v jejím grafu existuje dvojice bodů, mezi kterými nevede žádná cesta, pak bude příslušná komponenta matice A^ℓ nulová pro libovolné ℓ, viz lemmata 7 a 8 a také viz definici 6.

Poznamenejme, že pro výše uvedenou matici hyperlinků H platí $H^\ell > 0$ pro $\ell = 5$. Zřejmě

$$H = \frac{1}{6}\begin{bmatrix} 0 & 3 & 3 & 0 & 0 & 0 \\ 6 & 0 & 0 & 0 & 0 & 0 \\ 2 & 0 & 0 & 2 & 2 & 0 \\ 0 & 0 & 0 & 0 & 6 & 0 \\ 0 & 0 & 2 & 2 & 0 & 2 \\ 0 & 3 & 0 & 0 & 3 & 0 \end{bmatrix},$$

$$H^2 = \frac{1}{18}\begin{bmatrix} 12 & 0 & 0 & 3 & 3 & 0 \\ 0 & 9 & 9 & 0 & 0 & 0 \\ 0 & 3 & 5 & 2 & 6 & 2 \\ 0 & 0 & 6 & 6 & 0 & 6 \\ 2 & 3 & 0 & 2 & 11 & 0 \\ 9 & 0 & 3 & 3 & 0 & 3 \end{bmatrix},$$

$$H^3 = \frac{1}{108}\begin{bmatrix} 0 & 36 & 42 & 6 & 18 & 6 \\ 72 & 0 & 0 & 18 & 18 & 0 \\ 28 & 6 & 12 & 22 & 28 & 12 \\ 12 & 18 & 0 & 12 & 66 & 0 \\ 18 & 6 & 28 & 22 & 12 & 22 \\ 6 & 36 & 27 & 6 & 33 & 0 \end{bmatrix},$$

$$H^4 = \frac{1}{324}\begin{bmatrix} 150 & 9 & 18 & 60 & 69 & 18 \\ 0 & 108 & 126 & 18 & 54 & 18 \\ 30 & 60 & 70 & 40 & 96 & 28 \\ 54 & 18 & 84 & 66 & 36 & 66 \\ 46 & 60 & 39 & 40 & 127 & 12 \\ 135 & 9 & 42 & 60 & 45 & 33 \end{bmatrix},$$

$$H^5 = \frac{1}{1944}\begin{bmatrix} 90 & 504 & 588 & 174 & 450 & 138 \\ 900 & 54 & 108 & 360 & 414 & 108 \\ 500 & 174 & 282 & 332 & 464 & 192 \\ 276 & 360 & 234 & 240 & 762 & 72 \\ 438 & 174 & 392 & 332 & 354 & 254 \\ 138 & 504 & 495 & 174 & 543 & 90 \end{bmatrix} > 0\,;$$

viz také tabulku 6.1, kde jsou vypsány cesty délky pět mezi libovolnými dvěma vrcholy grafu. Reálně ale pracujeme s googlovskou maticí $G = \alpha\widetilde{H} + (1-\alpha)E$, $0 < \alpha < 1$, pro kterou platí $G^\ell > 0$ pro $\ell = 1$ a která je tedy *vždy primitivní.*

Googlovská matice

$$G = \alpha\Big(H + \frac{1}{n}de^T\Big) + (1-\alpha)\frac{1}{n}ee^T, \quad 0 < \alpha < 1 \tag{6.10}$$

je nezáporná (přesněji řečeno kladná), nerozložitelná a primitivní. Mocninná metoda aplikovaná na matici G^T bude pro libovolný startovací vektor π_0, $\pi_0 > 0$, vždy konvergovat k jednoznačně danému Perronovu vlastnímu vektoru π, $\pi > 0$, který je zároveň hledaným PageRank vektorem.

Pro vyjasnění vztahů mezi jednotlivými druhy matic opět odkazujeme na schéma na obrázku 6.3 na konci textu na str. 62.

6.4 Mocninná metoda pro googlovskou matici

V sekci, kde jsme vysvětlili mocninnou metodu pro obecné čtvercové matice, jsme zmínili, že je potřeba použít nějakou vhodně zvolenou normu. Nyní mocninnou metodu specifikujeme pro googlovskou matici, a proto potřebujeme vybrat už konkrétní normu.

Lemma 11. *Nechť vektor $v \geq 0$, $\|v\|_1 = 1$ a S je stochastická matice. Pak pro vektor w, $w = S^T v$, platí $w \geq 0$, $\|w\|_1 = 1$.*

Důkaz. Pro dokázání tohoto lemmatu si stačí uvědomit, že pokud existuje vektor $v \geq 0$, jehož součet prvků je roven jedné, pak určitě vektor w, který vznikne vynásobením stochastické matice S^T (která má součet prvků ve sloupci roven jedné) vektorem v, je také kladný a jeho jedničková norma je také jedna. □

Vezměme si nyní googlovskou matici G ($G > 0$) a zvolme si kladný (protože PageRank vektor předpokládáme kladný) startovací vektor $\pi_0 > 0$, $\|\pi_0\|_1 = 1$. Pro vektor

$$\pi_\ell \equiv (G^T)^\ell \pi_0 \equiv \begin{bmatrix} r_\ell(\mathtt{P}_1) \\ \vdots \\ r_\ell(\mathtt{P}_n) \end{bmatrix}, \quad \text{kde} \quad \ell = 1, 2, 3, \ldots$$

platí

$$\pi_\ell > 0, \quad \|\pi_\ell\|_1 = 1,$$

podle lemmatu 11 a také platí

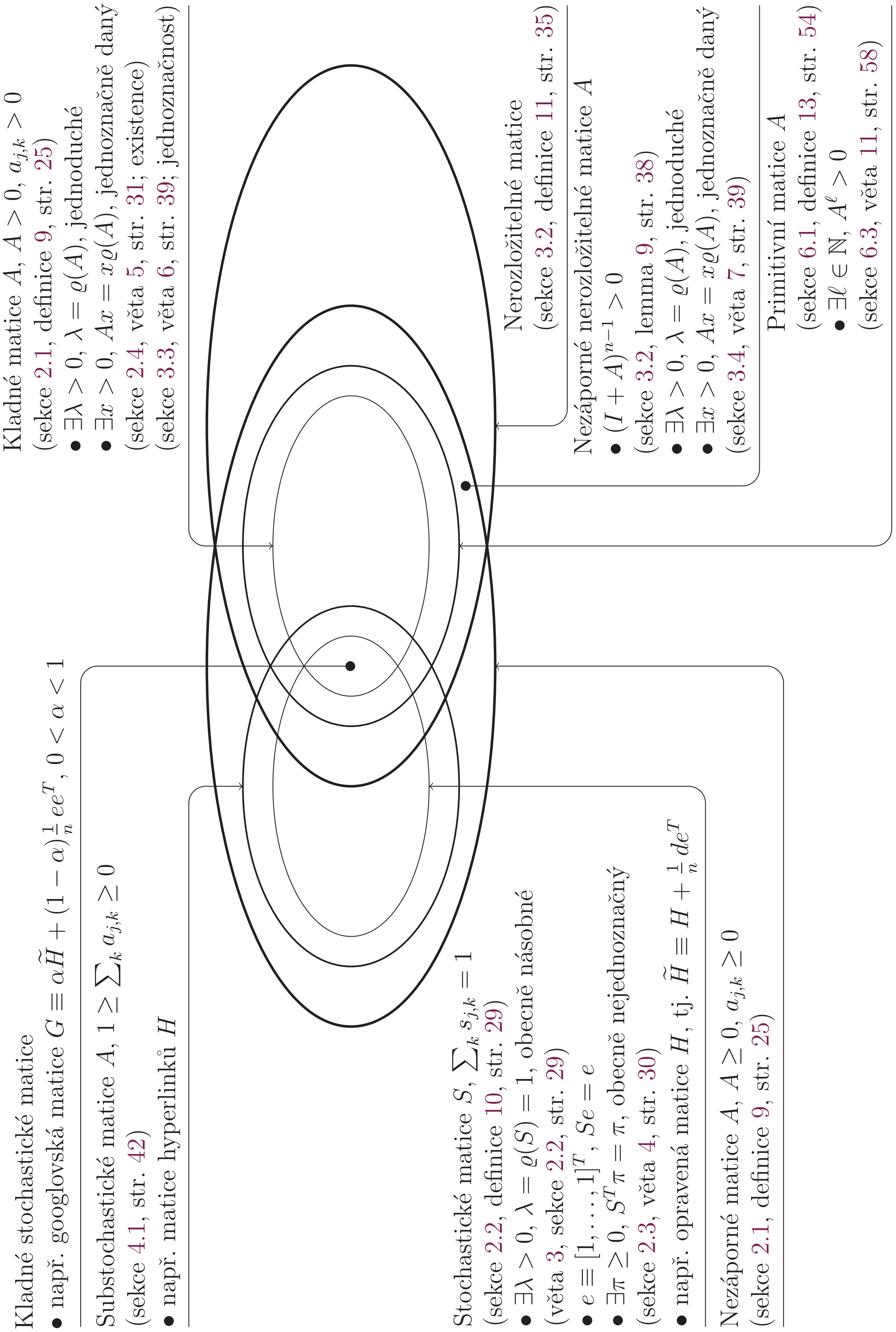
$$\lim_{\ell \longrightarrow \infty} \pi_\ell = (G^T)^\ell \pi_0 = \pi, \tag{6.11}$$

kde π je PageRank vektor podle věty 8, a tedy

$$\lim_{\ell \longrightarrow \infty} r_\ell(\mathtt{P}_k) = r(\mathtt{P}_k), \quad k = 1, \ldots, n. \tag{6.12}$$

Tabulka 6.1: V grafu (internetu) na obrázku 1.1 existuje cesta délky pět mezi libovolnými dvěma vrcholy (stránkami).

Odkud	1		2		3		4		5	Kam
P_1	$\longrightarrow$	P_3	$\longrightarrow$	P_5	$\longrightarrow$	P_6	$\longrightarrow$	P_2	$\longrightarrow$	P_1
P_1	$\longrightarrow$	P_2	$\longrightarrow$	P_1	$\longrightarrow$	P_2	$\longrightarrow$	P_1	$\longrightarrow$	P_2
P_1	$\longrightarrow$	P_3	$\longrightarrow$	P_1	$\longrightarrow$	P_3	$\longrightarrow$	P_1	$\longrightarrow$	P_3
P_1	$\longrightarrow$	P_3	$\longrightarrow$	P_5	$\longrightarrow$	P_4	$\longrightarrow$	P_5	$\longrightarrow$	P_4
P_1	$\longrightarrow$	P_3	$\longrightarrow$	P_4	$\longrightarrow$	P_5	$\longrightarrow$	P_4	$\longrightarrow$	P_5
P_1	$\longrightarrow$	P_3	$\longrightarrow$	P_5	$\longrightarrow$	P_6	$\longrightarrow$	P_5	$\longrightarrow$	P_6
P_2	$\longrightarrow$	P_1	$\longrightarrow$	P_2	$\longrightarrow$	P_1	$\longrightarrow$	P_2	$\longrightarrow$	P_1
P_2	$\longrightarrow$	P_1	$\longrightarrow$	P_3	$\longrightarrow$	P_5	$\longrightarrow$	P_6	$\longrightarrow$	P_2
P_2	$\longrightarrow$	P_1	$\longrightarrow$	P_3	$\longrightarrow$	P_4	$\longrightarrow$	P_5	$\longrightarrow$	P_3
P_2	$\longrightarrow$	P_1	$\longrightarrow$	P_3	$\longrightarrow$	P_4	$\longrightarrow$	P_5	$\longrightarrow$	P_4
P_2	$\longrightarrow$	P_1	$\longrightarrow$	P_3	$\longrightarrow$	P_5	$\longrightarrow$	P_3	$\longrightarrow$	P_5
P_2	$\longrightarrow$	P_1	$\longrightarrow$	P_3	$\longrightarrow$	P_4	$\longrightarrow$	P_5	$\longrightarrow$	P_6
P_3	$\longrightarrow$	P_1	$\longrightarrow$	P_3	$\longrightarrow$	P_1	$\longrightarrow$	P_3	$\longrightarrow$	P_1
P_3	$\longrightarrow$	P_5	$\longrightarrow$	P_6	$\longrightarrow$	P_2	$\longrightarrow$	P_1	$\longrightarrow$	P_2
P_3	$\longrightarrow$	P_5	$\longrightarrow$	P_6	$\longrightarrow$	P_2	$\longrightarrow$	P_1	$\longrightarrow$	P_3
P_3	$\longrightarrow$	P_4	$\longrightarrow$	P_5	$\longrightarrow$	P_4	$\longrightarrow$	P_5	$\longrightarrow$	P_4
P_3	$\longrightarrow$	P_5	$\longrightarrow$	P_3	$\longrightarrow$	P_5	$\longrightarrow$	P_3	$\longrightarrow$	P_5
P_3	$\longrightarrow$	P_4	$\longrightarrow$	P_5	$\longrightarrow$	P_6	$\longrightarrow$	P_5	$\longrightarrow$	P_6
P_4	$\longrightarrow$	P_5	$\longrightarrow$	P_3	$\longrightarrow$	P_1	$\longrightarrow$	P_3	$\longrightarrow$	P_1
P_4	$\longrightarrow$	P_5	$\longrightarrow$	P_6	$\longrightarrow$	P_2	$\longrightarrow$	P_1	$\longrightarrow$	P_2
P_4	$\longrightarrow$	P_5	$\longrightarrow$	P_6	$\longrightarrow$	P_2	$\longrightarrow$	P_1	$\longrightarrow$	P_3
P_4	$\longrightarrow$	P_5	$\longrightarrow$	P_3	$\longrightarrow$	P_4	$\longrightarrow$	P_5	$\longrightarrow$	P_4
P_4	$\longrightarrow$	P_5	$\longrightarrow$	P_4	$\longrightarrow$	P_5	$\longrightarrow$	P_4	$\longrightarrow$	P_5
P_4	$\longrightarrow$	P_5	$\longrightarrow$	P_3	$\longrightarrow$	P_4	$\longrightarrow$	P_5	$\longrightarrow$	P_6
P_5	$\longrightarrow$	P_6	$\longrightarrow$	P_2	$\longrightarrow$	P_1	$\longrightarrow$	P_3	$\longrightarrow$	P_1
P_5	$\longrightarrow$	P_3	$\longrightarrow$	P_1	$\longrightarrow$	P_2	$\longrightarrow$	P_1	$\longrightarrow$	P_2
P_5	$\longrightarrow$	P_3	$\longrightarrow$	P_5	$\longrightarrow$	P_3	$\longrightarrow$	P_5	$\longrightarrow$	P_3
P_5	$\longrightarrow$	P_4	$\longrightarrow$	P_5	$\longrightarrow$	P_4	$\longrightarrow$	P_5	$\longrightarrow$	P_4
P_5	$\longrightarrow$	P_6	$\longrightarrow$	P_2	$\longrightarrow$	P_1	$\longrightarrow$	P_3	$\longrightarrow$	P_5
P_5	$\longrightarrow$	P_6	$\longrightarrow$	P_5	$\longrightarrow$	P_6	$\longrightarrow$	P_5	$\longrightarrow$	P_6
P_6	$\longrightarrow$	P_5	$\longrightarrow$	P_3	$\longrightarrow$	P_1	$\longrightarrow$	P_3	$\longrightarrow$	P_1
P_6	$\longrightarrow$	P_2	$\longrightarrow$	P_1	$\longrightarrow$	P_2	$\longrightarrow$	P_1	$\longrightarrow$	P_2
P_6	$\longrightarrow$	P_2	$\longrightarrow$	P_1	$\longrightarrow$	P_3	$\longrightarrow$	P_5	$\longrightarrow$	P_3
P_6	$\longrightarrow$	P_5	$\longrightarrow$	P_3	$\longrightarrow$	P_4	$\longrightarrow$	P_5	$\longrightarrow$	P_4
P_6	$\longrightarrow$	P_5	$\longrightarrow$	P_6	$\longrightarrow$	P_5	$\longrightarrow$	P_6	$\longrightarrow$	P_5
P_6	$\longrightarrow$	P_2	$\longrightarrow$	P_1	$\longrightarrow$	P_3	$\longrightarrow$	P_5	$\longrightarrow$	P_6



Obrázek 6.3: Přehled vztahů mezi jednotlivými druhy čtvercových matic

Vraťme se nyní k definici PageRanku, viz rovnici (1.7). K určení hodnot $r(\mathtt{P}_k)$ je potřeba provést limitní přechod (6.12), což není v praxi možné. Mocninná metoda je v tento okamžik vhodný iterační postup, jak nalézt alespoň nějakou aproximaci hodnot $r(\mathtt{P}_k)$. Tento přístup také navrhli zakladatelé PageRank, Brin a Page, protože si byli vědomi, že hodnoty $r(\mathtt{P}_k)$, tedy PageRank odkazujících stránek na $\mathtt{P}_j$, jsou neznámé. Na začátku všem stránkám na webu přiřadíme například stejné hodnoty PageRank. Tyto hodnoty jsou $1/n$, kde n je počet všech stran webu. Tento postup pak počítá $r_\ell(\mathtt{P}_k)$, tj. aproximaci $r(\mathtt{P}_k)$ v ℓ-tém kroku pro každou stránku $\mathtt{P}_j$. Dojdeme proto k modifikaci vzorce (1.7)

$$r_{\ell+1}(\mathtt{P}_k) = \sum_{\mathtt{P}_j \in \mathscr{B}_{\mathtt{P}_k}} \frac{r_\ell(\mathtt{P}_j)}{|\mathtt{P}_j|}. \tag{6.13}$$

O tomto vztahu musíme však ještě chvíli uvažovat. První problém s touto rovnicí je, že nepočítá s modelem uživatele, který náhodně přechází ze stránky na stránku (s modelem random surfer) a občas narazí na dangling node, tj. vrchol grafu, do kterého vede cesta, ale už nevede cesta z něj. Je proto potřeba zaprvé zahrnout do rovnice (1.7) dangling node vektor. Druhým problémem je, že vztah nepočítá ani s uživatelem, který nejenom že prochází internetem pomocí odkazů, ale občas také napíše přímo adresu webové stránky a teleportuje se jinam, a proto je nutné zahrnout do definice teleportační matici.

Definici PageRanku, která zahrnuje již všechny výše zmíněné modifikace je

$$r_{\ell+1}(\mathtt{P}_k) = \alpha \underbrace{\sum_{\mathtt{P}_j \in \mathscr{B}_{\mathtt{P}_k}} \frac{r_\ell(\mathtt{P}_j)}{|\mathtt{P}_j|}}_{H^T\pi_\ell} + \frac{\alpha}{n} \underbrace{\sum_{\mathtt{P}_j \in \mathscr{D}} r_\ell(\mathtt{P}_j)}_{(de^T)^T\pi_\ell} + \frac{1-\alpha}{n} \underbrace{\sum_{j=1}^{n} r_\ell(\mathtt{P}_j)}_{(ee^T)\pi_\ell}, \tag{6.14}$$

kde $\mathscr{D}$ je množina všech stránek, které neobsahují žádný odkaz (dangling nodes) a kde jednotlivé součty představují výše zmíněné modifikace, viz (6.10). Po krátké úvaze zjistíme, že vrozec můžeme ještě upravit. Víme totiž, že platí

$$\sum_{j=1}^{n} r_\ell(\mathtt{P}_j) = \|\pi_\ell\|_1 = 1,$$

a proto

$$r_{\ell+1}(\mathtt{P}_k) = \alpha \sum_{\mathtt{P}_j \in \mathscr{B}_{\mathtt{P}_k}} \frac{r_\ell(\mathtt{P}_j)}{|\mathtt{P}_j|} + \frac{\alpha}{n} \sum_{\mathtt{P}_j \in \mathscr{D}} r_\ell(\mathtt{P}_j) + \frac{1-\alpha}{n}. \tag{6.15}$$

Poznamenejme, že původní definice před úpravami je platná a naprosto správná pro model, ve kterém by nebyl problém s dangling nodes, s nerozložitelností a také s teleportací. Výpočet PageRank vektoru pro náš modelový internet z obrázku 1.1 je možné nalézt v příloze A.

Závěr

V této práci jsme se zabývali PageRank vektorem, který je využíván v mnoha odvětvích, například silniční doprava nebo fyzika. My jsme hovořili o PageRank vektoru neboli také o Perronovu vlastním vektoru v souvislosti s internetovým vyhledávačem Google. PageRank byl také navržen pro potřeby internetu, a to už v roce 1976 Larry Pagem a Sergeyem Brinem na univerzitě ve Stanfordu. Pro společnost Google se stal jejím srdcem, které napomáhá společnosti si držet přední pozici mezi internetovými vyhledávači.

V kapitole 1 jsme se věnovali nejprve některým základním pojmům z teorie matic a jejich vlastních čísel, z teorie grafů a také jsme si uvedli původní definici PageRank vektoru. Dále následovala první část práce, která se zabývala existencí PageRank vektoru. Zadefinovali jsme si stochastické matice, bez kterých by PageRank neexistoval. Dále jsme se věnovali problematice (ne)rozložitelnosti, protože jsme zjistili, že v případě, že je matice rozložitelná, nebude PageRank vektor jednoznačný. Museli jsme tedy upravit matici hyperlinků, která obecně není ani stochastická ani nerozložitelná a dostali jsme googlovskou matici, pro kterou již PageRank vektor existuje a je jednoznačný.

V druhé části textu jsme se věnovali samotnému výpočtu PageRank vektoru. K výpočtu byla použita mocninná metoda, kterou jsme si představili v kapitole 5. V závěrečné kapitole 6 jsme ještě narazili na problém spojený s mocninnou metodou. Zjistili jsme, že proto aby se PageRank vektor ustálil k nějaké konkrétní hodnotě, tj. aby mocninná metoda fungovala, potřebujeme pracovat s primitivní maticí. Proto jsme si ještě definovali (im)primitivní matice a ukázali si testy (im)primitivity. Po té nám již nic nebránilo k úpravě definice PageRanku a k samotnému výpočtu. V průběhu celé práce jsme pracovali s modelovým internetem o šesti stránkách. Pro tento model jsme také PageRank vektor spočítali, viz přílohu A.

A PageRank vektor modelového internetu z obrázku 1.1

Zde uvádíme příklad výpočtu PageRanku. PageRank je vypočítán pro náš modelový internet z obrázku 1.1. Zvolíme startovací vektor

$$\pi_0 = [1/6, 1/6, 1/6, 1/6, 1/6, 1/6]^T$$

a zkusíme mocninnou metodu. Ta nám ukáže, že nejdůležitější stránkou je $\mathtt{P}_5$ a nejméně důležitou je $\mathtt{P}_6$.

Zkusíme nyní mocninnou metodu znovu, a to zaprvé pro startovací vektor

$$\pi_0 = [1/10, 1/10, 1/10, 1/10, 1/2, 1/10]^T,$$

tzn. té nejdůležitější stránce $\mathtt{P}_5$ přiřadíme již na zažátku největší důležitost a vidíme, že nic zvláštního se neděje. Hodnoty PageRanku se ustálily po několika málo iteracích a stránka $\mathtt{P}_5$ zůstala nejdůležitější. Zkusme se podívat na to, co se stane, když si jako startovací vektor zvolíme

$$\pi_0 = [1/10, 1/10, 1/10, 1/10, 1/10, 1/2]^T,$$

tzv. stránce $\mathtt{P}_6$ přiřadíme na zažátku největší důležitost. Zadáme tedy falešnou stopu a nejvyšší důležitost přiřadíme ve skutečnosti nejméně důležité stránce. Můžeme vidět, že proces konvergence trvá o něco déle, ale nakonec dojdeme ke stejnému závěru, a to že vypočtené PageRanky jsou

$$r(\mathtt{P}_5) > r(\mathtt{P}_1) > r(\mathtt{P}_3) > r(\mathtt{P}_2) \approx r(\mathtt{P}_4) > r(\mathtt{P}_6),$$

viz obrázek A.1. Poznamenejme, že zde pracujeme pouze s maticí hyperlinků H, viz (1.8), parametr $\alpha = 1$, a proto se teleportační matice výpočtu neúčastní.

Na obrázku A.2 již studujeme vliv parametru α pro různé hodnoty a pracujeme již s googlovskou maticí

$$G = \alpha H + (1 - \alpha)E.$$

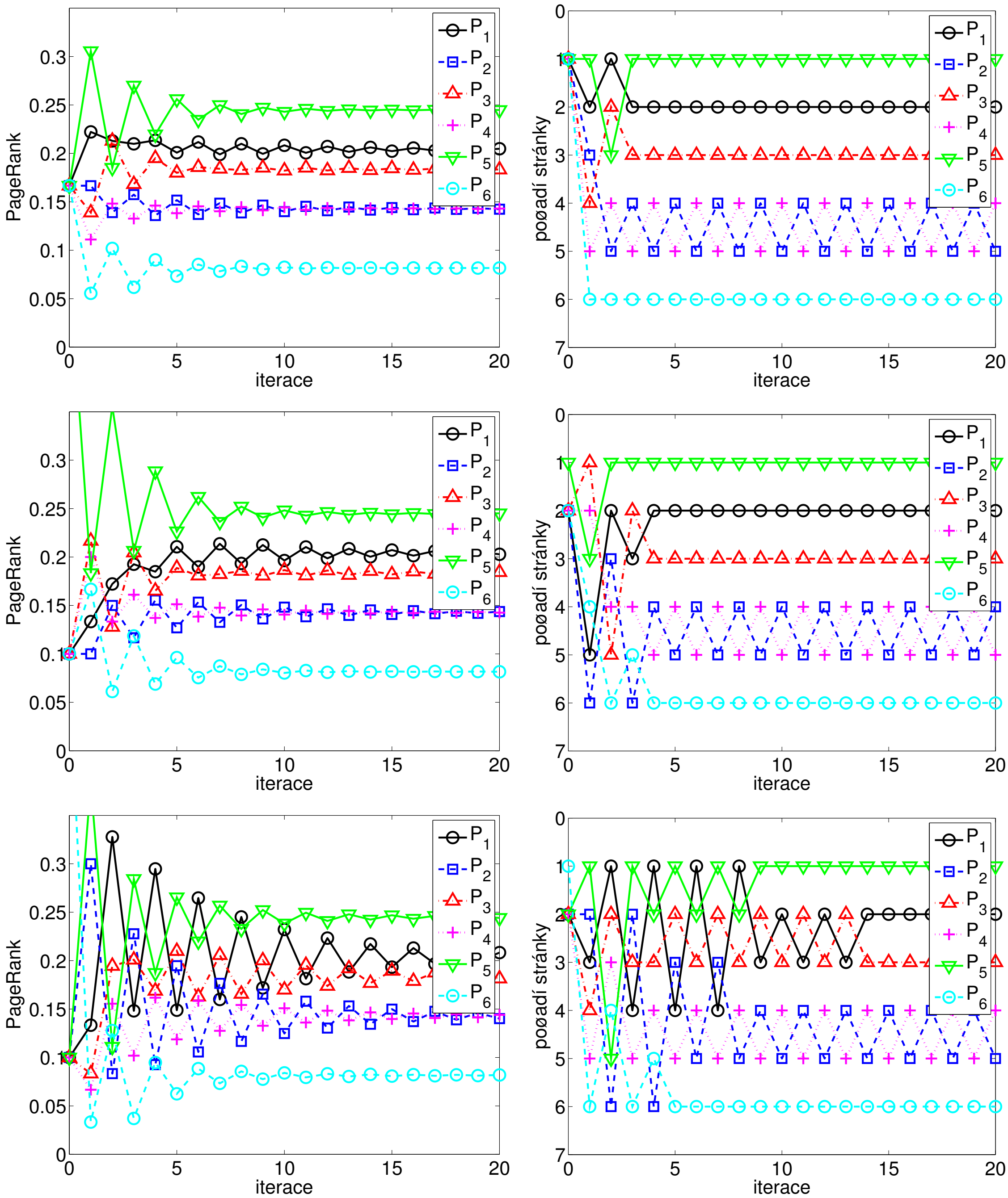
V případě, kdy je α velké, má teleportační matice méně významný vliv a rozdíly mezi PageRanky jsou výraznější. Zjednoznační se tak pořadí stránek $\mathtt{P}_2$ a $\mathtt{P}_4$, takže vidíme, že

$$r(\mathtt{P}_5) > r(\mathtt{P}_1) > r(\mathtt{P}_3) > r(\mathtt{P}_2) > r(\mathtt{P}_4) > r(\mathtt{P}_6).$$

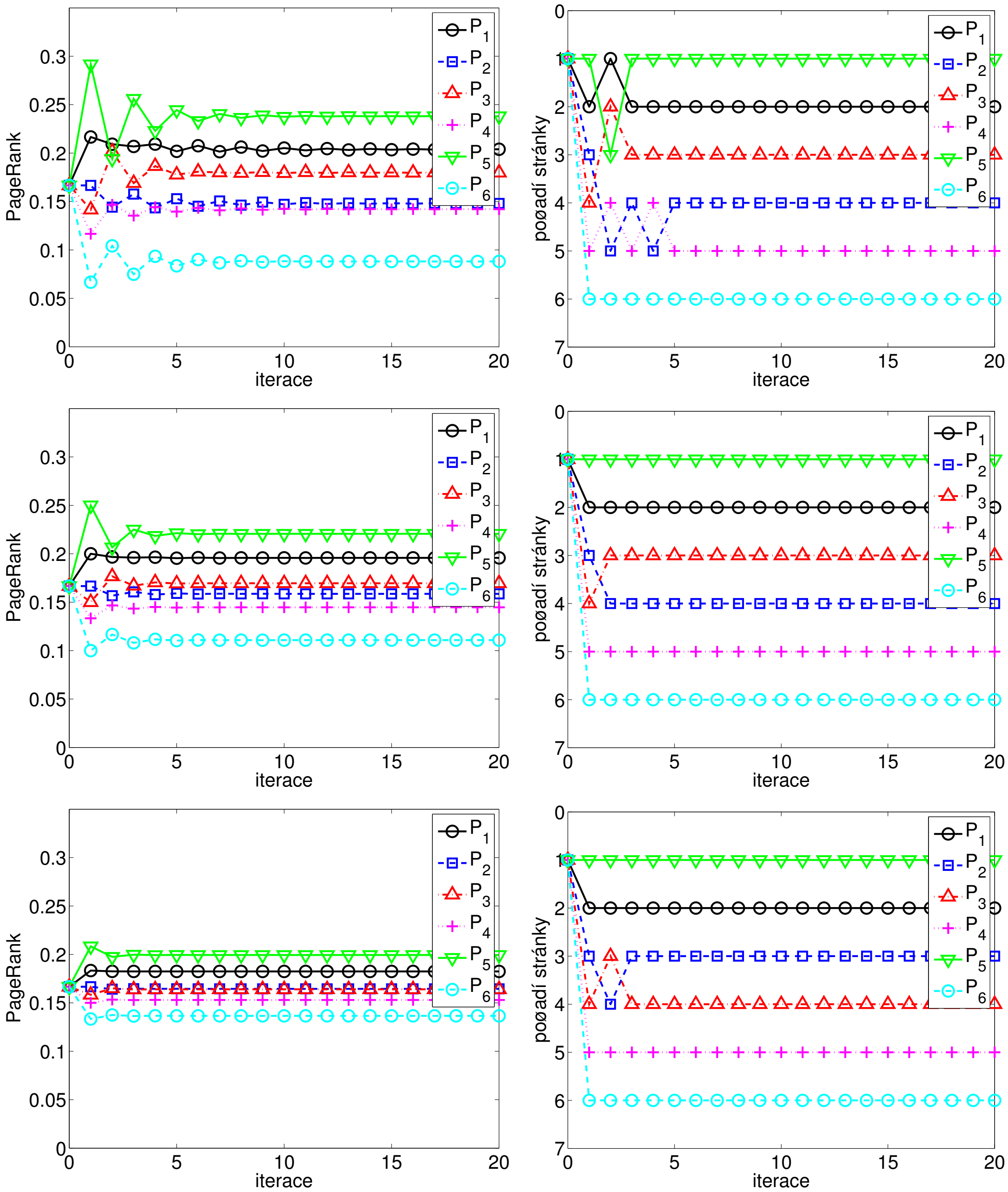
V opačném případě, tj. pro malá α, se PageRanky liší méně a navíc dojde k záměně pořadí stránek $\mathtt{P}_3$ a $\mathtt{P}_2$ na

$$r(\mathtt{P}_5) > r(\mathtt{P}_1) > r(\mathtt{P}_2) > r(\mathtt{P}_3) > r(\mathtt{P}_4) > r(\mathtt{P}_6).$$

Poznamenejme, že výpočet je proveden v programu MATLAB® 8.1.0.604 (R2013a) na počítači Dell Latitude 6430U s procesorem Intel® Core™ i7-3687U, 2.10 GHz s 8.00 GB RAM a se 64 bitovým operačním systémem Windows® 7 Professional, SP1.



Obrázek A.1: Konvergence PageRank vektoru internetu z obrázku 1.1. Googlovská matice $G = \alpha H + (1-\alpha)E$, kde matice hyperlinků H je (1.8) a $\alpha = 1$ (teleportační matice E se tedy výpočtu neúčastní). Vlevo vývoj jednotlivých PageRanků, vpravo vývoj pořadí stránek. V jednotlivých řadách byla mocninná metoda nastartována, odshora, vektorem $\frac{1}{6}[1,1,1,1,1,1]^T$, $\frac{1}{2}[\frac{1}{5},\frac{1}{5},\frac{1}{5},\frac{1}{5},1,\frac{1}{5}]^T$, resp. $\frac{1}{2}[\frac{1}{5},\frac{1}{5},\frac{1}{5},\frac{1}{5},\frac{1}{5},1]^T$. Vektor je normalizován tak, aby součet prvků byl roven jedné. Vykresleno je prvních dvacet iterací. Vypočtené PageRanky jsou $r(\mathtt{P}_5) > r(\mathtt{P}_1) > r(\mathtt{P}_3) > r(\mathtt{P}_2) \approx r(\mathtt{P}_4) > r(\mathtt{P}_6)$.



Obrázek A.2: Konvergence PageRank vektoru internetu z obrázku 1.1. Googlovská matice $G = \alpha H + (1-\alpha)E$, kde matice hyperlinků H je (1.8) a $\alpha = 0.9, 0.6$, resp. 0.3, v jednotlivých řadách. Vlevo vývoj jednotlivých PageRanků, vpravo vývoj pořadí stránek. Mocninná metoda byla nastartována vektorem $\frac{1}{6}[1, 1, 1, 1, 1, 1]^T$. Vektor je normalizován tak, aby součet prvků byl roven jedné. Vykresleno je prvních dvacet iterací. Pro velká α se vypočtené pořadí stránek $\mathtt{P}_2$ a $\mathtt{P}_4$ zjednoznační tak, že $r(\mathtt{P}_5) > r(\mathtt{P}_1) > r(\mathtt{P}_3) > r(\mathtt{P}_2) > r(\mathtt{P}_4) > r(\mathtt{P}_6)$. Pro malá α (tj. při velkém vlivu teleportace) se zmenšují rozdíly mezi PageRanky. Navíc dojde k záměně pořadí stránek $\mathtt{P}_3$ a $\mathtt{P}_2$ na $r(\mathtt{P}_5) > r(\mathtt{P}_1) > r(\mathtt{P}_2) > r(\mathtt{P}_3) > r(\mathtt{P}_4) > r(\mathtt{P}_6)$.

Literatura

[1] A. Berman, R. J. Plemmons: *Nonnegative matrices in the mathematical sciences*, SIAM Publications, Philadelphia, 1994.

[2] M. W. Berry, M. Browne: *Understanding search engines. Mathematical modeling and text retrieval*, 2nd ed., SIAM Publications, Philadelphia, 2005.

[3] S. Brin, L. Page: *The anatomy of a large-scale hypertextual Web search engine*, Computer Networks and ISDN Systems 30(1–7) (1998), str. 107–117.

[4] S. Brin, R. Motwani, L. Page, T. Winograd: *What can you do with a Web in your pocket?*, Data Engineering Bulletin 21(2) (1998), str. 37–47.

[5] S. Brin, R. Motwani, L. Page, T. Winograd: *The PageRank citation ranking: Bringing order to the Web*, Technical Report 1999-0120, Computer Science Department, Stanford University, Stanford, 1999.

[6] R. A. Brualdi, H. J. Ryser: *Combinatorial matrix theory*, Cambridge University Press (edice Encyclopedia of Mathematics and its Applications, vol. 39), Cambridge, 1991 (reprint 2003).

[7] D. Cvektović, P. Rowlinson, S. Simić: *Eigenspaces of graphs*, Cambridge University Press (edice Encyclopedia of Mathematics and its Applications, vol. 66), Cambridge, 1997 (reprint 2004).

[8] K. Čulík, V. Doležal, M. Fiedler: *Kombinatorikcá analýza v praxi*, SNTL, Státní nakladatelství technické literatury, Praha, 1967.

[9] J. Demel: *Grafy a jejich aplikace*, Academia, Praha 2002.

[10] J. Demel: *Grafy*, SNTL, Státní nakladatelství technické literatury (edice MVŠT, Matematika pro vysoké školy technické, sešit XXXIV), Praha, 1988, resp. 1989 (dotisk).

[11] J. Duintjer Tebbens, I. Hnětynková, M. Plešinger, Z. Strakoš, P. Tichý: *Analýza metod pro maticové výpočty. Základní metody*, Matfyzpress, Praha, 2012.

[12] M. Fiedler: *Speciální matice a jejich použití v numerické matematice*, SNTL, Státní nakladatelství technické literatury (edice TKI, Teoretická knižnice inženýra), Praha, 1981.

[13] Ф. Р. Гантмачер: *Теория матриц*, Издательство Наука, Moskva, 1966.

[14] F. R. Gantmacher: *The theory of matrices*, AMS Chelsea Publishing, Providence, 1998.

[15] G. H. Golub, C. F. Van Loan: *Matrix computations* (čtvrté vydání), The Johns Hopkins University Press, Baltimore, 2013.

[16] G. Chartrand: *Introductory graph theory*, Dover Publications, New York, 1985.

[17] A. N. Langville, C. D. Meyer: *Google's PageRank and beyond. The science of search engine rankings*, Princeton University Press, Princeton, 2006.

[18] L. Maixner: *Makovovy procesy a jejich aplikace*, Univerzita Palackého v Olomouci, Olomouc, 1991.

[19] J. Matoušek, J. Nešetřil: *Kapitoly z diskrétní matematiky*, Karolinum, Praha, 2009.

[20] J. Nešetřil: *Teorie grafů*, SNTL, Státní nakladatelství technické literatury (edice MS, Matematický seminář, č. 13), Praha, 1979.

[21] J. Sedláček: *Kombinatorika v teorii a praxi. Úvod do teorie grafů*, Academia (edice CV, Cesta k vědění, č. 7), Praha, 1964.

[22] J. Sedláček: *Úvod do teorie graů*, Academia (edice CV, Cesta k vědění, č. 25, resp. 29), Praha, 1977 (2. vydání), resp. 1981 (3. vydání).

[23] L. Unčonvský: *Stochastické modely operačnej analýzy*, Alfa/SNTL, Bratislava, 1980.

[24] J. Vraný: *Lokalizace algoritmů pro řazení výsledků vyhledávání informací na webu*, Disertační práce, TU v Liberci, Liberec, 2009.