Univerzita Karlova v Praze
Matematicko-fyzikální fakulta

BAKALÁŘSKÁ PRÁCE

Jaromír Vaněček

Jak funguje vyhledávač Google

Katedra algebry

Vedoucí bakalářské práce:	doc. RNDr. Jiří Tůma
Studijní program:	Matematika
Studijní obor:	Matematické metody informační bezpečnosti

Praha 2015

Prohlašuji, že jsem tuto bakalářskou práci vypracoval samostatně a výhradně s použitím citovaných pramenů, literatury a dalších odborných zdrojů.

Beru na vědomí, že se na moji práci vztahují práva a povinnosti vyplývající ze zákona č. 121/2000 Sb., autorského zákona v platném znění, zejména skutečnost, že Univerzita Karlova v Praze má právo na uzavření licenční smlouvy o užití této práce jako školního díla podle §60 odst. 1 autorského zákona.

V .... dne .... Podpis autora

Poděkování

Rád bych poděkoval svému vedoucímu bakalářské práce doc. RNDr. Jiřímu Tůmovi za odborné poznatky, předané zkušenosti a příjemné chvíle strávené při konzultacích.

Název práce: Jak funguje vyhledávač Google

Autor: Jaromír Vaněček

Katedra: Katedra algebry

Vedoucí bakalářské práce: doc. RNDr. Jiří Tůma, Katedra algebry

Abstrakt: Tato práce se zabývá vyhledávačem Google, převážně způsobem, jakým jsou vyhledávané stránky řazeny a jeho aplikací v jiných oblastech. Nejprve představíme obecné fungování vyhledávače, vytvoříme Google matici a předvedeme si princip algoritmu PageRank. Následně vše, v čistě matematické části práce, podložíme matematickou teorií zahrnující především Perronovu větu. Další část je věnována použití PageRanku na porovnání týmů ve fotbalové Synot lize. Nakonec ještě uvedeme několik jednoduchých pozorování o tom, jak různé jevy v hypertextové struktuře webu ovlivňují Google matici.

Klíčová slova: Google, PageRank, Perronova věta, stochastická matice

Title: How Google works

Author: Jaromír Vaněček

Department: Department of Algebra

Supervisor: doc. RNDr. Jiří Tůma, Department of Algebra

Abstract: This thesis deals with the web search engine Google, particularly the way how searched pages are ordered and with the application of this process in different areas. First, we briefly introduce how a web search engine works, create the Google matrix and show principle of the PageRank algorithm. Then, in the completely mathematical section of the work, we describe the mathematical theory supporting our statements including Perron's theorem. The next section is concerned with how to use PageRank to compare teams in football Synot league. In the end few simple observations on how different facts in the web's hyperlink structure influence the Google matrix will be described.

Keywords: Google, PageRank, Perron's theorem, stochastic matrix

Obsah

1 Úvod

1.1 Potřeba vyhledávat

Na úvod si představme, že potřebujeme najít nějaký sešit ze základní školy doma v krabici. Není jich tam příliš mnoho, a pokud jsme je uložili s nějakým řádem, nalezneme hledaný sešit za pár minut. Nyní jsme v místní knihovně a potřebujeme se podívat do knihy Harry Potter a Ohnivý pohár. Knihy v knihovně jsou seřazené podle žánrů a podle abecedy, takže toto fantasy máme opět za pár minut. Ale co kdybychom potřebovali najít všechny informace o slově abrakadabra. To se může vyskytovat v mnoha knihách, a tak nám nezbyde nic jiného, než projít většinu knih v knihovně. To už nám zabere spoustu času. A co kdybychom měli prohledávat všechny knihovny v kraji, v republice, nebo dokonce na světě? To už bychom určitě nezvládli. Chtěli bychom nějakou osobu, nebo spíše armádu, která přesně ví, kde co je a během chvilky nám dá seznam těch knih, ve kterých se vyskytuje hledané slovo. Přesně stejná situace nastává i na internetu. Naší armádě se říká vyhledávač.

Pro zajímavost můžeme uvést, že pokud bychom si chtěli prohlédnout celý internet(nejméně 4,54 miliard indexovaných stránek[1]) a zvládli bychom každou stránku prolistovat za 1 sekundu, strávili bychom beze spánku u počítače bezmála 144 let.

1.2 Stručný popis vyhledávače

Celý proces vyhledávání lze rozložit na několik částí. Nejprve popíšeme ty, které nejsou závislé na dotazu uživatele. Je tedy možné na nich pracovat nepřetržitě (24 hodin denně). U vyhledávače Google je to většina, ale existují i vyhledávače, které provádějí mnohem více výpočtů až při samotném dotazu.

Určitě je třeba procházet web a získávat informace o jednotlivých stránkách. To obstarávají takzvaní pavouci. Jsou to roboti, kteří shromažďují údaje o stránkách ve velice obsáhlém jednotném úložišti dat. Zde jsou uloženy stránky celé, nijak nezkomprimované a čekají na zařazení do pomyslného katalogu. Tam se dostanou přiřazením indexu. My pro jednoduchost rozlišíme dva druhy indexů, a sice index obsahu a index struktury. První zmíněný se vypočte pouze z textu, který je na stránce uveden. Index struktury pak čerpá ze struktury hypertextových odkazů webu. Zatímco o metodách výpočtu indexu obsahu se zmiňovat nebudeme, hypertextové odkazy pro nás budou velice důležité a dají základ konečnému ohodnocování stránek.

Tím se dostáváme k podstatě úspěchu Googlu. Nestačí, aby z katalogu byly vybrány ty stránky, které odpovídají dotazu uživatele, a tyto nějak náhodně seřazené. Vyhledávač je musí seřadit ve vhodném pořadí. Právě tady nám pomůže index struktury, díky němuž se všechny stránky ohodnotí.

Příklad: Představme si, že chceme najít informace o bakalářském studiu na Matematicko-fyzikální fakultě. Do vyhledávače zadáme dotaz *bakalářské studium mff*. Jaké procesy proběhnou, abychom dostali stránky, které požadujeme?

[1] údaj z www.worldwidewebsize.com, 22.02.2015. Indexovanými stránkami myslíme stránky, se kterými pracuje Google. Celkem jich je přibližně 50 krát více.

Nejprve se slovně zadaný výraz přeloží do podoby, aby s ním uměl program pracovat. Poté se na základě indexů obsahu, které jsou ovšem vypočítány třeba i před týdnem, vybere seznam vhodných stránek. To je typicky prováděno porovnáváním dotazu s četností výskytu na stránce. Najdou se tedy stránky, kde se často vyskytují slova *bakalářské*, *studium* a *mff*. Je ale třeba dát pozor například na synonyma a podobná úskalí, o kterých my ale mluvit nebudeme. Tyto stránky mají nějaké indexy struktury(také vytvořené dlouho dopředu), ze kterých se vypočítá jejich ohodnocení. Pak jsou stránky setříděny dle tohoto ohodnocení a nakonec jsou nám slavnostně předloženy.

Velmi pravděpodobně můžeme očekávat, že na prvním místě bude stránka fakulty, neboť hledané výrazy, především *mff*, se budou nejvíce vyskytovat právě tam. Dále může být zobrazena například stránka s konverzací, kde se někdo ptá jaké je studium na mff. Je ale intuitivně jasné, že tato stránka je méně vhodná co se indexů obsahu i ohodnocení týče.

To vše musí probíhat tak, abychom dostali stránku, která co nejlépe odpovídá našim představám. O to se stará algoritmus zvaný PageRank. Podstatu tohoto algoritmu si popíšeme v následující části.

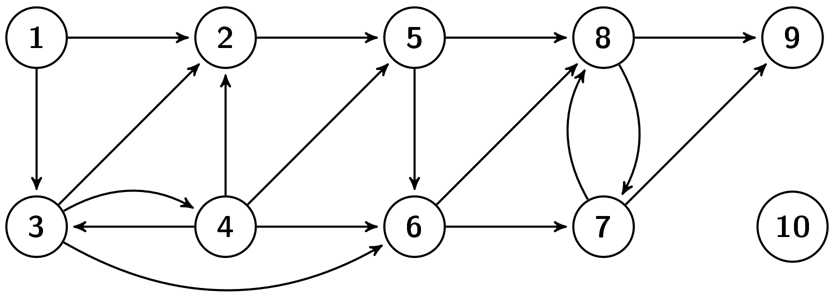
2 Idea algoritmu PageRank

2.1 Graf Internetu

Na Internet se můžeme koukat jako na orientovaný graf, označíme-li stránky jako vrcholy a hypertextové odkazy jako hrany. Hrana vede z vrcholu 1 do vrcholu 2, pokud se na stránce 1 vyskytuje hypertextový odkaz na stránku 2. To, že stránka obsahuje odkaz sama na sebe, nebudeme zakreslovat.

Právě takovýto pohled nám umožňuje seřadit stránky podle důležitosti neboli popularity. Následující formulace přesně říká, jak toho docílíme. *Důležitost stránky je přímoúměrná součtu důležitostí stránek, které na ní odkazují.* Zde se objevuje první problém. Uvažme začátek tohoto ohodnocování. Vybereme náhodně jednu stránku, například 5. Na tuto stránku odkazuje stránka 2, ale ta ještě nemá žádné ohodnocení. Jak tedy začít ohodnocovat? Časem se ukáže, že tato definice lze elegantně popsat pomocí jednoduché teorie matic.

Příklad: Následující graf na obrázku 1 ilustruje naší představu. Náš graf má 10 vrcholů, to znamená, že máme 10 stránek, které na sebe vzájemně odkazují. Všimněme si několika věcí. Na stránku 1 nevede žádný odkaz. Ze stránky 9 nevede žádný odkaz. Z žádné stránky ze skupiny stránek 7, 8, 9 nevede odkaz mimo tuto skupinu. Pokud bychom se na tuto situaci dívali jako na tok v síti, tak se časem vše nahromadí ve vrcholu 9. Vidíme tedy několik dalších problémů.



Obrázek 1: Zjednodušený graf internetu

2.2 Konstrukce ohodnocení

Příklad: Vycházejme z definice důležitosti stránky, jež jsme uvedli výše. Uvažujme dvě stránky A a B. Na první vede pouze jeden odkaz, ale zato z velice populární stránky, například stránky BBC. Na druhou vede odkazů deset, ale pouze z nedávno vytvořených blogů, na kterých není téměř žádný obsah. Je nám jasné, že důležitější by měla být stránka A.

Na druhou stranu si představme, že ze stránky s důležitostí r vede pouze jeden odkaz, a to na stránku A. Na B vede odkaz ze stránky také s důležitostí r, ale z té vede ještě dalších 50 odkazů. Tady bychom přirozeně chtěli, aby stránka A byla důležitější.

Tato dvě pravidla uvedená v příkladu nás vedou k následujícímu. Označme $s_1, s_2, \ldots, s_n$ všechny stránky v katalogu (indexované stránky), $r(s_i)$ důležitost stránky s_i, $|s_i|$ počet odkazů vedoucí ze stránky s_i a O_{s_i} množinu stránek odkazujících na s_i. Nyní můžeme popsat popularitu stránky s_i takto

$$r(s_i) = \sum_{s_j \in O_{s_i}} \frac{r(s_j)}{|s_j|} \ .$$

Na tomto místě můžeme osvětlit, jak se lze vyhnout našemu problému s neznámými počátečními hodnotami. Pro stránku s_i označíme $r_0(s_i)$ její počáteční popularitu. Nakonec se ukáže, že na volbě této počáteční popularity nezáleží, vzhledem ke konečnému hodnocení, a tak budeme nyní volit pro všechny stránky stejnou popularitu (tj. $r_0(s_i) = \frac{1}{n}$). Zajímavých hodnot dosáhneme následujícím iterativním postupem. Pokud definujeme $r_{k+1}(s_i)$ popularitu stránky s_i po $k+1$ krocích, můžeme tomu přizpůsobit i předchozí sumu.

$$r_{k+1}(s_i) = \sum_{s_j \in O_{s_i}} \frac{r_k(s_j)}{|s_j|} \ .$$

Jestliže chceme zjistit popularitu všech stránek, musíme tuto sumu spočítat pro každé $i = 1, 2, \ldots, n$. To lze elegantně popsat maticovým zápisem. Definujeme vektor důležitostí $\boldsymbol{v}_k = (r_k(s_1), r_k(s_2), \ldots, r_k(s_n))^T$ a čtvercovou matici $\boldsymbol{H} = (h_{ij})$ řádu n takovou, že

$$h_{ij} = \begin{cases} \frac{1}{|s_j|}, & \text{stránka } s_j \text{ odkazuje na stránku } s_i \\ 0, & \text{jinak} \ . \end{cases}$$

Poznamenejme, že v j-tém sloupci jsou na i-té pozici nenulové prvky právě tehdy, když ze stránky s_j vede odkaz na stránku s_i. Na druhou stranu v i-tém řádku jsou na j-té pozici nenulové prvky právě tehdy, když ze stránky s_j vede odkaz na stránku s_i. Nyní už k slibovanému maticovému zápisu našich n rovnic popularity.

$$\boldsymbol{v}_{k+1} = \boldsymbol{H}\boldsymbol{v}_k$$

Je snadné ověřit, že tato rovnost platí. Na pravé straně totiž násobíme matici řádu n s vektorem $n \times 1$ a dostaneme tedy opět vektor $n \times 1$. V něm je na i-té pozici suma

$$\sum_{j=1}^{n} h_{ij} r_k(s_j) \ .$$

Prvek h_{ij} je nenulový právě pro taková j, že $s_j \in O_{s_i}$ a je rovný přesně $\frac{1}{|s_j|}$. Z toho už vidíme

$$\sum_{j=1}^{n} h_{ij} r_k(s_j) = \sum_{s_j \in O_{s_i}} \frac{1}{|s_j|} r_k(s_j) = r_{k+1}(s_i) \ .$$

Ještě udělejme jedno pozorování. Vektor $\boldsymbol{v}_{k+1}$ můžeme vyjádřit pouze pomocí počátečního vektoru $\boldsymbol{v}_0$ a matice $\boldsymbol{H}$ následovně

$$\boldsymbol{v}_{k+1} = \boldsymbol{H}\boldsymbol{v}_k = \boldsymbol{H}(\boldsymbol{H}\boldsymbol{v}_{k-1}) = \ldots = \boldsymbol{H}^{k+1}\boldsymbol{v}_0 \ .$$

Příklad: Konkrétní matici $\boldsymbol{H}$ můžeme uvést pro síť z našeho příkladu s deseti stránkami. Co znamenají jednotlivé hodnoty v matici víme.

$$\boldsymbol{H} = \begin{array}{c} \begin{array}{cccccccccc} s_1 & s_2 & s_3 & s_4 & s_5 & s_6 & s_7 & s_8 & s_9 & s_{10} \end{array} \\ \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1/2 & 0 & 1/3 & 1/4 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1/2 & 0 & 0 & 1/4 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1/3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1/4 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1/3 & 1/4 & 1/2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1/2 & 0 & 1/2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1/2 & 1/2 & 1/2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1/2 & 1/2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} \end{array} \begin{array}{c} \\ s_1 \\ s_2 \\ s_3 \\ s_4 \\ s_5 \\ s_6 \\ s_7 \\ s_8 \\ s_9 \\ s_{10} \end{array}$$

Podívejme se na různé jevy, které se v matici vyskytují a na jejich význam. Nulový i-tý řádek nám říká, že na s_i nevede žádný odkaz. Naopak nulový i-tý sloupec nám říká, že ze stránky s_i nevede žádný odkaz. Také je zajímavý počet nul v pravém horním rohu matice, říká nám to, že z žádné ze stránek s_5, s_6, s_7, s_8 a s_9 už nevede odkaz zpět na některou ze stránek s_1, s_2, s_3 a s_4.

Počáteční vektor popularity vypadá takto $\boldsymbol{v}_0 = (1/10,\ 1/10,\ \dots\ ,\ 1/10)^T$. Zkusme několikrát aplikovat výpočet $\boldsymbol{v}_{k+1} = \boldsymbol{H}\boldsymbol{v}_k$. Dostáváme tyto výsledky

$$\boldsymbol{v}_0 = \frac{1}{10}\begin{pmatrix}1\\1\\1\\1\\1\\1\\1\\1\\1\\1\end{pmatrix},\ \boldsymbol{v}_1 = \frac{1}{120}\begin{pmatrix}0\\13\\9\\4\\15\\13\\12\\18\\12\\0\end{pmatrix},\ \boldsymbol{v}_2 = \frac{1}{240}\begin{pmatrix}0\\8\\2\\6\\28\\23\\31\\40\\30\\0\end{pmatrix},\ \cdots,\ \boldsymbol{v}_5 \approx \begin{pmatrix}0\\0\\0\\0\\0\\0\\1/48\\1/46\\1/25\\0\end{pmatrix}.$$

Všimněme si, že už ve vektoru $\boldsymbol{v}_1$ má stránka s_1 nulovou popularitu. To je dáno tím, že na ní neodkazuje žádná stránka, takže $O_{s_1} = \emptyset$. To ale neznamená, jak už jsme si dříve uvědomili, že by tato stránka byla zcela bezvýznamná. Zpočátku také vidíme, jak postupně popularita mizí ze stránek s_1 - s_4. Nakonec nezanedbatelné hodnoty zůstávají pouze u stránek s_7, s_8, a s_9, ale i tyto hodnoty jsou dost malé oproti počáteční 1/10. Zdá se, že vektor popularity bude konvergovat k nulovému vektoru, což nás rozhodně neuspokojí.

2.3 Google matice

V této části si ukážeme, jak nastíněné problémy vyřešit. Potřebujeme zabránit ztrátě důležitosti (tj. pokud začneme iteraci s vektorem, který má součet složek roven 1, budeme chtít, aby se tento součet po provedení několika kroků nezměnil). Jinak řečeno, musíme vyřešit problém s koncovými body z předchozí části. Také bychom přirozeně chtěli, abychom po několika krocích dostali jeden vektor, který bude určovat pořadí stránek podle důležitosti. Budeme tedy směřovat k tomu, aby proces $\boldsymbol{v}_{k+1} = \boldsymbol{H}\boldsymbol{v}_k$ konvergoval k nějakému jednoznačně určenému vektoru $\boldsymbol{v} = (v_1, v_2, \ldots, v_n)^T$ takovému, že $\sum v_i = 1$. To nás vede k tomu, že budeme muset upravit naší matici $\boldsymbol{H}$. To, jak jí máme modifikovat nám říká matematická teorie, kterou si popíšeme později, a nebo přístup "otců" PageRanku Larryho Page a Sergeye Brina. Tento postup je více názorný, a tak si ho předvedeme jako první. Následně pak ukážeme, že skutečně funguje vzhledem k oné teorii.

K popisu chování uživatele na internetu můžeme využít náhodnou procházku. Tento termín má svou matematickou definici, kterou uvádět nebudeme, ale budeme ho používat intuitivně. Představa je následující. Uživatel brouzdá ze stránky na stránku pomocí odkazů. Z jedné stránky se dostane na jinou tím, že náhodně klikne na jeden z odkazů, které jsou na ní umístěny. Nakonec bude množství času strávené na jednotlivých stránkách udávat jejich důležitost, protože čím důležitější stránka, tím více odkazů na ní vede z jiných důležitých stránek. Je nám jasné, že takovýto model nám určuje stejnou situaci, jako v předchozí části.

Nyní můžeme vyřešit problém se ztrácením důležitosti. To nastává v případě, že se uživatel dostane na stránku, ze které už nevede žádný odkaz. Takové stránky jsou například dokumenty a obrázky.

Řešení je jednoduché. Pokud se uživatel na takovou stránku dostane, tak náhodně přejde na jakoukoliv jinou. Na konkrétní stránku se tedy dostane s pravděpodobností $1/n$. Na tuto úpravu se lze také koukat tak, že do orientovaného grafu přidáme hrany z vrcholu, ze kterého žádné hrany nevedou, do všech vrcholů v grafu. Můžeme tedy matici $\boldsymbol{H}$ upravit tak, že nulové sloupce nahradíme sloupci $1/n\ \boldsymbol{e}$, kde $\boldsymbol{e}$ je vektor $n \times 1$, který má na každé pozici 1. Definujme vektor $\boldsymbol{a} = (a_1, a_2, \ldots, a_n)$ tak, že $a_i = 1$, pokud ze stránky s_i nevede žádný odkaz a $a_i = 0$ jinak. Dostáváme matici $\boldsymbol{S}$, pro kterou platí

$$\boldsymbol{S} = \boldsymbol{H} + (1/n\ \boldsymbol{e})\boldsymbol{a}^T \ .$$

Pro náš známý příklad by matice $\boldsymbol{S}$ vypadala takto

$$\boldsymbol{S} = \begin{array}{c} \begin{array}{cccccccccc} s_1 & s_2 & s_3 & s_4 & s_5 & s_6 & s_7 & s_8 & s_9 & s_{10} \end{array} \\ \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1/10 & 1/10 \\ 1/2 & 0 & 1/3 & 1/4 & 0 & 0 & 0 & 0 & 1/10 & 1/10 \\ 1/2 & 0 & 0 & 1/4 & 0 & 0 & 0 & 0 & 1/10 & 1/10 \\ 0 & 0 & 1/3 & 0 & 0 & 0 & 0 & 0 & 1/10 & 1/10 \\ 0 & 1 & 0 & 1/4 & 0 & 0 & 0 & 0 & 1/10 & 1/10 \\ 0 & 0 & 1/3 & 1/4 & 1/2 & 0 & 0 & 0 & 1/10 & 1/10 \\ 0 & 0 & 0 & 0 & 0 & 1/2 & 0 & 1/2 & 1/10 & 1/10 \\ 0 & 0 & 0 & 0 & 1/2 & 1/2 & 1/2 & 0 & 1/10 & 1/10 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1/2 & 1/2 & 1/10 & 1/10 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1/10 & 1/10 \end{pmatrix} \end{array} \begin{array}{c} \\ s_1 \\ s_2 \\ s_3 \\ s_4 \\ s_5 \\ s_6 \\ s_7 \\ s_8 \\ s_9 \\ s_{10} \end{array}$$

Ještě je třeba vyřešit přelévání důležitosti z jednoho bloku stránek do jiného, takže všechny stránky z prvního bloku budou mít nakonec nulovou důležitost. To si můžeme představit tak, že se uživatel dostane do bloku stránek, ze kterého už se nemůže dostat jinam. Tráví čas pouze zde, a tak jsou nakonec důležité jen tyto stránky. Vzhledem k počtu stránek na internetu to uživatel samozřejmě neví, takže se nijak nevzrušuje.

Řešení je opět velice jednoduché. Potřebujeme, aby se uživatel dostal na stránku, ke které se ale nemůže přímo proklikat přes odkazy. Představa je taková, že se po čase začne nudit, nebo neuvidí žádný odkaz, na který by klikl, nebo se zkrátka bude chtít podívat na něco jiného. Přejde tedy náhodně na libovolnou stránku na internetu. Za předpokladu, že by uživatel vždy přešel na náhodnou stránku, by graf internetu vypadal tak, že by mezi každými dvěma stránkami vedla hrana. Reprezentovat bychom ho mohli maticí $1/n\ \boldsymbol{e}\boldsymbol{e}^T$.

Označme ještě parametrem $\alpha \in [0,1]$ pravděpodobnost, že se uživatel drží struktury hypertextových odkazů webu. Hned vidíme, že s pravděpodobností $1-\alpha$ přejde na náhodnou stránku. Tím dostáváme matici $\boldsymbol{G}$, kterou nazýváme Google matice a pro kterou platí

$$\boldsymbol{G} = \alpha \boldsymbol{S} + (1-\alpha) 1/n\ \boldsymbol{e}\boldsymbol{e}^T$$

Můžeme zkusit upravit matici $\boldsymbol{S}$ z předchozího příkladu na matici $\boldsymbol{G}$.

$$\boldsymbol{G} = \frac{1}{10}\begin{pmatrix}
1-\alpha & 1-\alpha & 1-\alpha & 1-\alpha & 1-\alpha & 1-\alpha & 1-\alpha & 1-\alpha & 1 & 1\\
4\alpha+1 & 1-\alpha & \frac{7\alpha+3}{3} & \frac{3\alpha+2}{2} & 1-\alpha & 1-\alpha & 1-\alpha & 1-\alpha & 1 & 1\\
4\alpha+1 & 1-\alpha & 1-\alpha & \frac{3\alpha+2}{2} & 1-\alpha & 1-\alpha & 1-\alpha & 1-\alpha & 1 & 1\\
1-\alpha & 1-\alpha & \frac{7\alpha+3}{3} & 1-\alpha & 1-\alpha & 1-\alpha & 1-\alpha & 1-\alpha & 1 & 1\\
1-\alpha & 9\alpha+1 & 1-\alpha & \frac{3\alpha+2}{2} & 1-\alpha & 1-\alpha & 1-\alpha & 1-\alpha & 1 & 1\\
1-\alpha & 1-\alpha & \frac{7\alpha+3}{3} & \frac{3\alpha+2}{2} & 4\alpha+1 & 1-\alpha & 1-\alpha & 1-\alpha & 1 & 1\\
1-\alpha & 1-\alpha & 1-\alpha & 1-\alpha & 1-\alpha & 4\alpha+1 & 1-\alpha & 4\alpha+1 & 1 & 1\\
1-\alpha & 1-\alpha & 1-\alpha & 1-\alpha & 4\alpha+1 & 4\alpha+1 & 4\alpha+1 & 1-\alpha & 1 & 1\\
1-\alpha & 1-\alpha & 1-\alpha & 1-\alpha & 1-\alpha & 1-\alpha & 4\alpha+1 & 4\alpha+1 & 1 & 1\\
1-\alpha & 1-\alpha & 1-\alpha & 1-\alpha & 1-\alpha & 1-\alpha & 1-\alpha & 1-\alpha & 1 & 1
\end{pmatrix}$$

To není příliš přehledné, a tak si zkusíme zvolit konkrétní α a uvidíme, jaké hodnoty dostaneme. Udává se, že tento parametr se pohybuje okolo hodnoty $0,85$, takže my budeme počítat matici $\boldsymbol{G}$ pro $\alpha = 0,85$. Dostaneme takovouto matici.

$$\boldsymbol{G} = \frac{1}{1200}\begin{array}{c@{}c}
\begin{array}{cccccccccc} s_1 & s_2 & s_3 & s_4 & s_5 & s_6 & s_7 & s_8 & s_9 & s_{10} \end{array} & \\
\begin{pmatrix}
18 & 18 & 18 & 18 & 18 & 18 & 18 & 18 & 120 & 120\\
528 & 18 & 358 & 273 & 18 & 18 & 18 & 18 & 120 & 120\\
528 & 18 & 18 & 273 & 18 & 18 & 18 & 18 & 120 & 120\\
18 & 18 & 358 & 18 & 18 & 18 & 18 & 18 & 120 & 120\\
18 & 1038 & 18 & 273 & 18 & 18 & 18 & 18 & 120 & 120\\
18 & 18 & 358 & 273 & 528 & 18 & 18 & 18 & 120 & 120\\
18 & 18 & 18 & 18 & 18 & 528 & 18 & 528 & 120 & 120\\
18 & 18 & 18 & 18 & 528 & 528 & 528 & 18 & 120 & 120\\
18 & 18 & 18 & 18 & 18 & 18 & 528 & 528 & 120 & 120\\
18 & 18 & 18 & 18 & 18 & 18 & 18 & 18 & 120 & 120
\end{pmatrix} & \begin{array}{c} s_1\\ s_2\\ s_3\\ s_4\\ s_5\\ s_6\\ s_7\\ s_8\\ s_9\\ s_{10} \end{array}
\end{array}$$

Zatím jsme téměř výlučně vycházeli pouze z knihy Google's PageRank and Beyond (viz Langeville a Meyer, 2006). Pro doplňující informace je tedy možné nahlédnout do počátečních kapitol této knihy. Dále už to vzhledem k matematické povaze textu nebude možné a budeme muset sáhnout i po jiných publikacích.

3 Matematika za PageRankem

3.1 Připomenutí důležitých pojmů

Na úvod této části připomeneme nejdůležitější fakta a zavedeme značení, která budeme používat.

Budeme rozlišovat kladné a nezáporné matice (případně vektory). Kladná matice má každý prvek kladný a nezáporná matice má každý prvek kladný nebo rovný nule. Tedy pro $\boldsymbol{A} = (a_{ij})$ budeme značit $\boldsymbol{A} > \boldsymbol{0}$, pro $a_{ij} > 0$ a $\boldsymbol{A} \geq \boldsymbol{0}$, pro $a_{ij} \geq 0$.

Dále značením $|\boldsymbol{A}|$ budeme rozumět matici (nebo vektor), která má vstupy $|a_{ij}|$. Takováto matice je tedy jistě nezáporná.

V dalším textu budeme využívat i normu vektorů. Nebude to klasická euklidovská norma, ale norma, které se někdy říká oktaedrická, nebo manhattanská. Není to nic jiného než obyčejný součet absolutních hodnot složek vektoru. Pro vektor $\boldsymbol{v} = (v_1, v_2, \ldots, n)^T$ to tedy znamená $\|\boldsymbol{v}\|_1 = \sum_{i=1}^{n} |v_i|$.

Pokud někde nebude uvedeno jinak, budeme vždy značit složky matice $\boldsymbol{A}$ jako a_{ij} a složky vektoru $\boldsymbol{v}$ jako v_i.

Na některých místech se budou vyskytovat vektory $\boldsymbol{e}$ a $\boldsymbol{e}_i$. Takovýmto zápisem budeme mít na mysli vektory $(1, 1, \ldots, 1)^T$ a $(0, \ldots, 0, 1, 0, \ldots, 0)^T$, kde ve druhém je 1 na pozici i.

K označování pravděpodobnosti budeme používat zápis $P(X)$, kde X značí zkoumaný jev.

Na konec si ještě připomeneme nějaká fakta o Jordanově kanonickém tvaru. Jordanova buňka řádu k s vlastním číslem λ je čtvercová matice řádu k, která vypadá následovně

$$\boldsymbol{J}_{\lambda,k} = \begin{pmatrix} \lambda & 1 & & & \boldsymbol{0} \\ & \lambda & 1 & & \\ & & \ddots & \ddots & \\ & & & \ddots & 1 \\ \boldsymbol{0} & & & & \lambda \end{pmatrix}.$$

Jordanův kanonický tvar je blokově diagonální matice, kde každý blok je Jordanova buňka. Pro matici $\boldsymbol{A}$ s vlastními čísly $\{\lambda_1, \lambda_2, \ldots, \lambda_l\}$ vypadá Jordanův kanonický tvar takto

$$\boldsymbol{J}_{\boldsymbol{A}} = \begin{pmatrix} \boldsymbol{J}_{\lambda_1,k_1} & & & \boldsymbol{0} \\ & \boldsymbol{J}_{\lambda_2,k_2} & & \\ & & \ddots & \\ \boldsymbol{0} & & & \boldsymbol{J}_{\lambda_l,k_l} \end{pmatrix}.$$

Poznamenejme, že nad komplexními čísly je každá matice $\boldsymbol{A}$ podobná matici v Jordanově kanonickém tvaru. To znamená, že existuje regulární matice $\boldsymbol{T}$ taková, že $\boldsymbol{A} = \boldsymbol{T}\boldsymbol{J}_{\boldsymbol{A}}\boldsymbol{T}^{-1}$. Sloupce matice $\boldsymbol{T}$ jsou tvořeny vektory z Jordanových řetízků příslušných vlastním číslům matice $\boldsymbol{A}$. Pro detailnější popis Jordanova kanonického tvaru se můžeme podívat například do zápisek z přednášky z lineární algebry na Matematicko-fyzikální fakultě Univerzity Karlovi v Praze (Barto a Tůma, 2014, kap. 9.4.).

Na konci této kapitoly se nám bude hodit vědět, jak se mocní matice v Jordanově kanonickém tvaru. Jednoduše - umocníme zvlášť každou Jordanovu buňku. To, jak se mocní Jordanova buňka, si zde přímo napíšeme.

$$\boldsymbol{J}_{\lambda,k}^{m} = \begin{pmatrix} \lambda^m & \binom{m}{1}\lambda^{m-1} & \binom{m}{2}\lambda^{m-2} & \cdots \\ & \lambda^m & \binom{m}{1}\lambda^{m-1} & \cdots \\ & \mathbf{0} & \lambda^m & \ddots \\ & & & \ddots \end{pmatrix}$$

3.2 Úvod do Markovových řetězců

Nyní můžeme začít budovat teorii pro ospravedlnění uvedeného ohodnocování v předchozích sekcích. Nejprve se zaměříme na získání původní matice $\boldsymbol{H}$. K tomu nám poslouží teorie okolo Markovových řetězců a čerpat budeme většinou z knihy Google's PageRank and Beyond (Langeville a Meyer, 2006, kap. 15.3).

Definice 3.1. Systém náhodných veličin $\{X_t, t \in T\}$ se stejným prostorem stavů Ω nazýváme *náhodný proces.*

V našem případě bude Ω množinou webových stránek, bude tedy konečná. Dále také budeme uvažovat $T = \mathbb{N}$, neboť prvky T pro nás budou představovat jednotlivé kroky v náhodné procházce. Stránka, na kterou se dostaneme kliknutím na náhodný odkaz, představuje náhodnou veličinu.

Definice 3.2. Buď $\{X_t, t \in \mathbb{N}\}$ náhodný proces a $\Omega = \{S_1, S_2, \ldots, S_n\}$. Tento proces nazveme *Markovův řetězec*, pokud pro každé $t = 0, 1, 2, \ldots$ splňuje $P(X_{t+1} = S_j | X_t = S_{i_t}, X_{t-1} = S_{i_{t-1}}, \ldots, X_0 = S_{i_0}) = P(X_{t+1} = S_j | X_t = S_{i_t})$.

Tato definice jinými slovy říká, že pravděpodobnost, že se po dalším kroku dostaneme do určitého stavu, závisí pouze na tom, v jakém stavu se nacházíme právě teď, nikoliv na tom, v jakých stavech jsme byli v předchozích krocích.

Například klikání na náhodné odkazy je jistě Markovův řetězec, neboť nezáleží, jak jsme se dostali na stránku, na které jsme, vzhledem k pravděpodobnosti s jakou se dostaneme na nějakou další stránku. Tyto pravděpodobnosti budeme nazývat přechodové pravděpodobnosti.

Definice 3.3. Čtvercovou matici $\boldsymbol{A}$ nazveme *stochastickou*, pokud je součet prvků v každém jejím sloupci roven jedné.

Poznamenejme, že pro libovolnou čtvercovou matici $\boldsymbol{A}$ platí $\rho(\boldsymbol{A}) \leq \|\boldsymbol{A}\|$, kde $\rho(\boldsymbol{A})$ značí spektrální poloměr a $\|\boldsymbol{A}\|$ značí libovolnou normu matice $\boldsymbol{A}$ (Meyer, 2001, kap. 7). Připomeňme, že 1-norma se dá spočítat jako maximum ze součtu absolutních hodnot prvků ve sloupci.

Pozorování 3.4. 1 je vlastním číslem stochastické matice.

Důkaz: Mějme stochastickou matici $\boldsymbol{B}$ řádu n. Označme $\boldsymbol{A} = \boldsymbol{B} - 1\boldsymbol{I}_n$. Víme, že 1 je vlastním číslem $\boldsymbol{B}$ právě tehdy, když $det(\boldsymbol{A}) = 0$, neboli když je matice A singulární.

Buďte $\boldsymbol{A}_{1*}, \boldsymbol{A}_{2*}, \ldots, \boldsymbol{A}_{n*}$ řádky matice $\boldsymbol{A}$. Víme, že matice $\boldsymbol{A}$ je singulární $\Leftrightarrow$ $\boldsymbol{A}$ má lineárně závislé řádky $\Leftrightarrow$ $\sum_{i=1}^{n} \alpha_i \boldsymbol{A}_{i*} = 0$ a nějaké α_i je nenulové $\Leftrightarrow$ Pro každé $j = 1, 2, \ldots, n$ platí $0 = \sum_{i=1}^{n} \alpha_i a_{ij}$. Tato suma ale pro $\alpha_i = 1$ značí součet všech prvků ve sloupci j a ten víme, že je roven nule, protože u matice $\boldsymbol{B}$ byl roven 1. $\square$

Důsledek 3.5. Spektrální poloměr nezáporné stochastické matice je roven 1.

Definice 3.6. Mějme stejnou množinu Ω jako v definici 3.2 a dále přechodové pravděpodobnosti $p_{ij}(t) = P(X_t = S_i | X_{t-1} = S_j)$. Matici $\boldsymbol{P}(t) = (p(t)_{ij})$ takovou, že $p(t)_{ij} = p_{ij}(t)$ nazveme *přechodovou maticí*.

Markovovy řetězce, které mají přechodovou matici neměnnou, tj. $\boldsymbol{P} = \boldsymbol{P}(t)$ pro všechna $t = 1, 2, \ldots$ se někdy označují jako homogenní řetězce. Právě tuhle vlastnost má i námi zkoumaný řetězec, a tak dále budeme přechodovou matici značit pouze $\boldsymbol{P}$.

Pokud bychom nyní sestrojili přechodovou matici Markovova řetězce daného klikáním na náhodné odkazy na stránce, dostali bychom matici $\boldsymbol{P}$, která by se od původní matice $\boldsymbol{H}$ lišila pouze v tom, že matice $\boldsymbol{H}$ má, na rozdíl od matice $\boldsymbol{P}$, nulové sloupce tam, kde z nějakých stránek nevedou žádné odkazy. Pokud by nevedl žádný odkaz ze stránky i, tak by i-tý sloupec matice $\boldsymbol{P}$ vypadal jako $\boldsymbol{e}_i$.

Každá přechodová matice Markovova řetězce je stochastická a na rozmyšlení necháme fakt, že každá stochastická matice určuje Markovův řetězec.

Definice 3.7. Vektor $\boldsymbol{p} \geq \boldsymbol{0}$ nazveme *pravděpodobnostní*, pokud platí $\|\boldsymbol{p}\|_1 = 1$.

Definice 3.8. Mějme Markovův řetězec s prostorem stavů $\{S_1, S_2, \ldots, S_n\}$ a s přechodovou maticí $\boldsymbol{P}$.
- Pravděpodobnostní vektor $\boldsymbol{p}$, pro který platí $\boldsymbol{P}\boldsymbol{p} = \boldsymbol{p}$ nazveme *stacionární pravděpodobnostní vektor*.
- Vektor $\boldsymbol{p}(k) = (p_1(k), p_2(k), \ldots, p_n(k))^T$, kde $p_i(k) = P(X_k = S_i)$ nazveme *k-tý pravděpodobnostní vektor* a pro $k = 0$ mluvíme o *počátečním pravděpodobnostním vektoru*.

Poznamenejme, že pokud existuje stacionární pravděpodobnostní vektor $\boldsymbol{p}$, tak je to vlastní vektor příslušný vlastnímu číslu 1. Podle pozorování 3.4 má stochastická matice vždy vlastní číslo 1, takže stacionární pravděpodobnostní vektor existuje vždy.

Dále také vidíme, že i-tá složka k-tého pravděpodobnostního vektoru řetězce, který vznikne klikáním na náhodné odkazy na stránkách je pravděpodobnost, že se přesně po k krocích nacházíme na i-té stránce. Důležitosti jednotlivých stránek bychom tedy mohli zjistit sečtením příslušných složek k-tých pravděpodobnostních vektorů pro $k = 1, 2, \ldots$ Stránka, která by měla nejvyšší hodnotu, by byla nejdůležitější atd. Důležitost stránky S_i bychom spočetli jako $\sum_{k=1}^{m} p_i(k)$, kde $m \in \mathbb{N}$

Tvrzení 3.9. Mějme Markovův řetězec s prostorem stavů $\{S_1, S_2, \ldots, S_n\}$ a s přechodovou maticí $\boldsymbol{P}$. Potom pro k-tý pravděpodobnostní vektor platí $\boldsymbol{p}(k) = \boldsymbol{P}^k \boldsymbol{p}(o)$.

Důkaz: Nejprve provedeme jednoduchý předvýpočet a následně tvrzení dokážeme indukcí podle k.

Zvolíme $t \in \mathbb{N}$. Pak pro libovolné $i = 1, 2, \ldots, n$ platí následující

$$p_i(t) = P(X_t = S_i) = \sum_{j=1}^{n} P(X_t = S_i, X_{t-1} = S_j) \ ,$$

tj. pravděpodobnost, že se nacházíme po t krocích na stránce S_i, je rovna součtu pravděpodobností, že se nacházíme po t krocích na stránce S_i a zároveň jsme po $t-1$ krocích byli na stránce S_j a sčítáme přes všechny stránky. Nyní použijeme definici podmíněné pravděpodobnosti, takže dostaneme

$$P(X_t = S_i, X_{t-1} = S_j) = P(X_t = S_i | X_{t-1} = S_j) P(X_{t-1} = S_j) \ .$$

Zbývá nám uvědomit si, že $P(X_t = S_i | X_{t-1} = S_j) = p_{ij}$ a $P(X_{t-1} = S_j) = p_j(t-1)$. Celkem tedy dostáváme $p_i(t) = \sum_{j=1}^{n} p_{ij} p_j(t-1)$. Z toho už přímo plyne, že $\boldsymbol{p}(t) = \boldsymbol{P}\boldsymbol{p}(t-1)$.

Nyní přistupme k indukci. Pro $k = 0$ platí zřejmě, protože $\boldsymbol{P}^0 = \boldsymbol{I}$ a pro $k = 1$ použijeme předvýpočet pro $t = 1$.

Nyní předpokládejme, že tvrzení platí pro $k-1$ a budeme chtít dokázat, že platí pro k. Rozepíšeme $\boldsymbol{P}^k \boldsymbol{p}(0) = \boldsymbol{P}\boldsymbol{P}^{k-1}\boldsymbol{p}(0)$, což se podle indukčního předpokladu rovná $\boldsymbol{P}\boldsymbol{p}(k-1)$. Nyní opět použijeme náš předvýpočet a dostáváme $\boldsymbol{P}^k \boldsymbol{p}(0) = \boldsymbol{p}(k)$. □

Tvrzení 3.10. Mějme Markovův řetězec s prostorem stavů $\{S_1, S_2, \ldots, S_n\}$ a s přechodovou maticí $\boldsymbol{P}$. Dále označme $\boldsymbol{P}^k = \left(p_{ij}^{(k)}\right)$. Pak $p_{ij}^{(k)} = P(X_k = S_i | X_0 = S_j)$.

Důkaz: Chceme dokázat, že pravděpodobnost, že se přesně po k krocích nacházíme na stránce S_i, pokud jsme vycházeli ze stránky S_j, odpovídá prvku na pozici i, j v matici $\boldsymbol{P}^k$. Zvolíme počáteční pravděpodobnostní vektor $\boldsymbol{p}(0)$ jako $\boldsymbol{e}_j$. To odpovídá tomu, že začínáme na stránce S_j.

Z předchozího tvrzení víme, že $\boldsymbol{p}(k) = \boldsymbol{P}^k \boldsymbol{p}(o)$, a můžeme tedy psát

$$\boldsymbol{p}(k) = \boldsymbol{P}^k \boldsymbol{e}_j = \left(p_{1j}^{(k)}, p_{2j}^{(k)}, \ldots, p_{nj}^{(k)}\right)^T \ .$$

Pro každé $i = 1, 2, \ldots, n$ tedy dostáváme $\boldsymbol{p}_i(k) = p_{ij}^{(k)}$ a to po rozepsání definice k-tého pravděpodobnostního vektoru znamená, že $P(X_k = S_i) = p_{ij}^{(k)}$, to jest pravděpodobnost, že se po k krocích nacházíme na stránce S_i je uvedena v matici $\boldsymbol{P}^k$ na pozici i, j a to jsme chtěli dokázat. □

3.3 Perronova věta

Vraťme se opět k matici $\boldsymbol{H}$, kterou jsme vytvořili dříve. Ta nebyla přechodovou maticí, neboť měla nějaké sloupce nulové, a tak by na ní nešel použít další postup. My jsme ale provedli dvě úpravy hypertextové struktury, které situaci změnily.

První z nich byl přechod k matici $\boldsymbol{S}$ tím, že jsme všechny nulové sloupce $\boldsymbol{H}$ nahradili sloupcem $\boldsymbol{e}/n$. Tím jsme docílili toho, že jsme z matice $\boldsymbol{H}$ udělali matici stochastickou, neboli

$$1 = \sum_{i=1}^{n} s_{i,j} = \boldsymbol{e}^T \boldsymbol{S} \boldsymbol{e}_j \ , \forall j = 1, 2, \ldots, n \ .$$

Matice $\boldsymbol{S}$ už je přechodovou maticí a od matice $\boldsymbol{P}$ se liší tím, že její i-tý sloupec (z i-té stránky nevede žádný odkaz) vypadá jako e/n. Získali jsme ale ještě jednu užitečnou vlastnost, která odpovídá zamezení ztrácení důležitosti.

Lemma 3.11. Mějme stochastickou matici $\boldsymbol{S}$ řádu n a nezáporný vektor $\boldsymbol{v}$. Potom platí $\|\boldsymbol{v}\|_1 = \|\boldsymbol{S}\boldsymbol{v}\|_1$.

Důkaz: Kýžený výsledek dostaneme jednoduchým rozepsáním normy:

$$\|\boldsymbol{S}\boldsymbol{v}\|_1 = \left\| \begin{pmatrix} s_{11}v_1 + s_{12}v_2 + \ldots + s_{1n}v_n \\ s_{21}v_1 + s_{22}v_2 + \ldots + s_{2n}v_n \\ \vdots \\ s_{n1}v_1 + s_{n2}v_2 + \ldots + s_{nn}v_n \end{pmatrix} \right\|_1 =$$

$$= |s_{11}v_1 + s_{12}v_2 + \ldots + s_{1n}v_n| + \ldots + |s_{n1}v_1 + s_{n2}v_2 + \ldots + s_{nn}v_n|$$

Díky tomu, že jsou matice $\boldsymbol{S}$ i vektor $\boldsymbol{v}$ nezáporné, můžeme psát

$$(s_{11}v_1 + s_{12}v_2 + \ldots + s_{1n}v_n) + \ldots + (s_{n1}v_1 + s_{n2}v_2 + \ldots + s_{nn}v_n) =$$

$$= (s_{11} + s_{12} + \ldots + s_{1n})v_1 + \ldots + (s_{n1} + s_{n2} + \ldots + s_{nn})v_n =$$

$$= v_1 + v_2 + \ldots + v_n = \sum_{i=1}^{n} |v_i| = \|\boldsymbol{v}\|_1 \, .$$

V posledních rovnostech jsme využili toho, že matice $\boldsymbol{S}$ je stochastická, takže pro každé i platí $s_{i1} + s_{i2} + \ldots + s_{in} = 1$. Pro přidání absolutních hodnot jsme potřebovali akorát nezáporný vektor $\boldsymbol{v}$. $\square$

Druhá úprava matice (nyní už $\boldsymbol{S}$) spočívala ve vynásobení určitým koeficientem α a následném přičtení matice, která měla na všech pozicích stejné kladné prvky. Tím jsme získali matici $\boldsymbol{G}$, která je pořád stochastická a navíc má každý prvek kladný ($\boldsymbol{G} > \boldsymbol{0}$). To nám zaručuje velice důležité vlastnosti matice $\boldsymbol{G}$, které využijeme dále.

Definice 3.12. Čtvercová matice $\boldsymbol{A}$ řádu n se nazývá *ireducibilní*, pokud pro každou permutační matici $\boldsymbol{P}$ řádu n platí

$$\boldsymbol{P}\boldsymbol{A}\boldsymbol{P}^{-1} \neq \begin{pmatrix} \boldsymbol{X} & \boldsymbol{Y} \\ \boldsymbol{0} & \boldsymbol{Z} \end{pmatrix} \ ,$$

kde obě matice $\boldsymbol{X}$ a $\boldsymbol{Z}$ jsou čtvercové řádu alespoň 1.

Pro člověka je na první pohled obtížné poznat podle této definice, jestli zadaná matice je nebo není ireducibilní (Pokud není, mluvíme o reducibilní matici). Uvedeme proto ekvivalentní definici, která problém, zejména pro matice s menším řádem, značně zjednoduší.

Připomeňme ještě z teorie grafů, že matice vzdálenosti nějakého orientovaného grafu je taková matice, která má na pozici (i, j) váhu hrany, která vede z j-tého vrcholu do i-tého. Dále také, že takovýto graf je silně souvislý, pokud pro každé dva vrcholy platí, že z jednoho vede cesta do druhého. Následující tvrzení můžeme najít například na stránkách ČVUT (viz Pelantová, 2010).

Tvrzení 3.13. Buď $\boldsymbol{A}$ čtvercová matice řádu n. Pak $\boldsymbol{A}$ je ireducibilní právě tehdy, když orientovaný graf vzdálenosti $G(\boldsymbol{A})$ matice $\boldsymbol{A}$ je silně souvislý.

Důkaz: Obě implikace budeme dokazovat sporem.

Nejprve předpokládejme, že $\boldsymbol{A}$ je reducibilní. Existuje tedy permutace řádků a sloupců, po které je matice $\boldsymbol{A}$ ve tvaru

$$\begin{pmatrix} \boldsymbol{X} & \boldsymbol{Y} \\ \boldsymbol{0} & \boldsymbol{Z} \end{pmatrix}.$$

Označme k řád matice $\boldsymbol{X}$. Potom je řád matice $\boldsymbol{Z}$ roven $n-k$. Okamžitě vidíme, že z množiny vrcholů $S_1, S_2, \ldots, S_k$ nevede ani jedna hrana do množiny vrcholů $S_{k+1}, S_{k+2}, \ldots, S_n$. Určitě tedy neplatí, že pro každé dva vrcholy grafu $G(\boldsymbol{A})$ lze najít cestu z jednoho do druhého, což je spor s tím, že $G(\boldsymbol{A})$ je silně souvislý. Tím jsme dokázali implikaci zprava doleva.

Nyní předpokládejme, že $G(\boldsymbol{A})$ není silně souvislý. Existují tedy vrcholy S_i a S_j takové, že z prvního neexistuje v grafu $G(\boldsymbol{A})$ cesta do druhého.

Rozdělíme vrcholy do tří množin následovně.

$$\begin{aligned} M_1 &= \{\text{vrcholy, do kterých vede cesta z } S_i\} \\ M_2 &= \{\text{vrcholy, ze kterých vede cesta do } S_j\} \\ M_3 &= \{\text{vrcholy, které nejsou v } M_1 \text{ ani v } M_2\} \end{aligned}$$

Je zřejmé, že všechny tři množiny jsou po dvou disjunktní a také to, že M_1 a M_2 jsou neprázdné. Vhodným výběrem permutační matice můžeme matici $\boldsymbol{A}$ převést do tvaru, kde na prvních pozicích bude mít vrcholy z M_1, pak vrcholy z M_3 a nakonec vrcholy z M_2. Matice bude vypadat takto

$$\begin{array}{c} \\ M_1 \\ M_3 \\ M_2 \end{array} \begin{array}{c} \begin{array}{ccc} M_1 & M_3 & M_2 \end{array} \\ \begin{pmatrix} \boldsymbol{X}_{11} & \boldsymbol{X}_{12} & \boldsymbol{X}_{13} \\ \boldsymbol{X}_{21} & \boldsymbol{X}_{22} & \boldsymbol{X}_{23} \\ \boldsymbol{X}_{31} & \boldsymbol{X}_{32} & \boldsymbol{X}_{33} \end{pmatrix} \end{array}.$$

Nyní se pokusíme určit matice $\boldsymbol{X}_{31}$ a $\boldsymbol{X}_{21}$. To, že z S_i nevede cesta do S_j implikuje, že z množiny vrcholů M_1 nevede žádná hrana do množiny vrcholů M_2. To znamená, že $\boldsymbol{X}_{31} = \boldsymbol{0}$. Z definice M_3 také víme, že z M_1 nevede žádná hrana do M_3, a tudíž dostáváme $\boldsymbol{X}_{21} = \boldsymbol{0}$

Všimněme si, že $\boldsymbol{X}_{kl}$ jsou čtvercové matice, takže čtvercová je také matice

$$\begin{pmatrix} \boldsymbol{X}_{22} & \boldsymbol{X}_{23} \\ \boldsymbol{X}_{32} & \boldsymbol{X}_{33} \end{pmatrix}.$$

Celkem dostaneme takovouto reducibilní matici

$$\begin{array}{c} \\ M_1 \\ M_3 \\ M_2 \end{array}\begin{array}{c} \begin{array}{ccc} M_1 & M_3 & M_2 \end{array} \\ \begin{pmatrix} \boldsymbol{X}_{11} & \boldsymbol{X}_{12} & \boldsymbol{X}_{13} \\ \mathbf{0} & \boldsymbol{X}_{22} & \boldsymbol{X}_{23} \\ \mathbf{0} & \boldsymbol{X}_{32} & \boldsymbol{X}_{33} \end{pmatrix}. \end{array}$$

Došli jsme ke sporu s předpokladem tvrzení a tím máme dokázanou i druhou implikaci. □

Definice 3.14. Nezápornou čtvercovou matici $\boldsymbol{A}$ nazveme *primitivní*, pokud existuje $k \in \mathbb{N}$ takové, že $\boldsymbol{A}^k > \mathbf{0}$, tj. $\boldsymbol{A}^k$ je kladná matice.

Pozorování 3.15. Matice $\boldsymbol{G}$ definovaná v části 2.3 je ireducibilní a primitivní.

Nyní by mohla následovat Perronova - Frobeniova věta, která se zabývá ireducibilními maticemi. Nám bude ale stačit dokázat Perronovu větu, která nám dává podobné výsledky jako Perronova - Frobeniova, ale pouze pro kladné matice. Ještě před ní si ale všimneme několika věcí, které nám zpřehlední a zjednoduší její důkaz.

Lemma 3.16. Nechť $x_1, x_2, \dots, x_n \in \mathbb{R}$. Potom $\left|\sum_{j=1}^n x_j\right| = \sum_{j=1}^n |x_j|$ právě tehdy, když jsou buď všechna x_i nezáporná, nebo všechna x_i nekladná. To jest označíme-li $\boldsymbol{x} = (x_1, x_2, \dots, x_n)^T$, tak

$$\left|\sum_{j=1}^n x_j\right| = \sum_{j=1}^n |x_j| \iff \text{buď } \boldsymbol{x} \geq \mathbf{0}\text{, nebo } \boldsymbol{x} \leq \mathbf{0}.$$

Důkaz: Jedná se vlastně o případ rovnosti v trojúhelníkové nerovnosti, která nám říká, že

$$\left|\sum_{j=1}^n x_j\right| \geq \sum_{j=1}^n |x_j|.$$

Poznamenejme, že pokud by bylo nějaké $x_i = 0$, tak by to nezměnilo ani jednu ze sum. Můžeme tedy předpokládat $x_i \neq 0$ pro všechna $i = 1, 2, \dots, n$.

Nejprve dokážeme implikaci zprava doleva a uděláme to indukcí. Pro $n = 2$ to platí zjevně. Stačí si rozmyslet, jak je to se zápornými čísly.

Nyní předpokládejme, že tvrzení platí pro n reálných čísel a budeme chtít ukázat, že platí také pro $n+1$. Vezměme $k \in \mathbb{N}$ t.ž. $1 < k < n+1$. Potom platí

$$\sum_{i=1}^{n+1} |x_i| = \sum_{i=1}^{k} |x_i| + \sum_{i=k+1}^{n+1} |x_i| = \left|\sum_{i=1}^{k} x_i\right| + \left|\sum_{i=k+1}^{n+1} x_i\right| = \left|\sum_{i=1}^{n+1} x_i\right| ,$$

kde jsme v předposledním kroku použili indukční předpoklad na každou sumu zvlášť a v posledním kroku jsme použili indukční předpoklad ještě jednou.

Implikaci zleva doprava budeme dokazovat sporem. Předpokládejme tedy, že existují j a k takové, že $x_j < 0$ a zároveň $x_k > 0$. Bez újmy na obecnosti můžeme předpokládat $j = 1$ a $k = 2$. Pak určitě $|x_1| + |x_2| > |x_1 + x_2|$. Nyní můžeme počítat

$$\begin{aligned}\sum_{i=1}^{n} |x_i| = \sum_{i=3}^{n} |x_i| + |x_1| + |x_2| &> \sum_{i=3}^{n} |x_i| + |x_1 + x_2| \\ &\geq \left|\sum_{i=3}^{n} x_i\right| + |x_1 + x_2| \geq \left|\sum_{j=1}^{n} x_j\right| .\end{aligned}$$

V posledních dvou krocích jsme použili trojúhelníkovou nerovnost. Došli jsme ke sporu s předpoklady, a proto musí být buď $\boldsymbol{x} < \mathbf{0}$ nebo $\boldsymbol{x} > \mathbf{0}$. $\square$

Podobné tvrzení platí i v komplexních číslech. Nestačí ale jen aby měly reálné a imaginární části stejná znaménka, musí také poměry těchto částí být stejné.

Důsledek 3.17. Nechť $z_1, z_2, \ldots, z_n \in \mathbb{C}$. Potom $\left|\sum_{j=1}^{n} z_j\right| = \sum_{j=1}^{n} |z_j|$ právě tehdy, když existují reálná kladná α_j pro $j = 1, 2, \ldots, n$ taková, že $z_j = \alpha_j z_1$.

Důkaz: Důkaz můžeme provést za pomocí předchozího lemmatu, nebo ho můžeme nalézt ve cvičení 5.1.10. z knížky Matrix Analysis and Applied Linear Algebra (viz Meyer, 2000). $\square$

Lemma 3.18. Pro libovolné vlastní číslo λ kladné matice $\boldsymbol{A}$ řádu n platí, že $\frac{\lambda}{r}$ je vlastním číslem matice $\frac{\boldsymbol{A}}{r}$, kde $r \in \mathbb{R} \setminus \{0\}$.

Důkaz: Výsledek získáme rovnou z definice vlastního čísla. Označme $\boldsymbol{v}$ vlastní vektor matice $\boldsymbol{A}$ příslušný vlastnímu číslu λ. To jest $\boldsymbol{Av} = \lambda\boldsymbol{v}$. Pak ale také, pro nenulové r, platí $\frac{\boldsymbol{A}}{r}\boldsymbol{v} = \frac{\lambda}{r}\boldsymbol{v}$. $\square$

Pozorování 3.19. Buď $\boldsymbol{A}$ kladná reálná matice řádu n a $\boldsymbol{u}$, $\boldsymbol{v}$ reálné vektory typu $n \times 1$. Potom platí
1) $\mathbf{0} \neq \boldsymbol{v} \geq \mathbf{0} \Longrightarrow \boldsymbol{Av} > \mathbf{0}$,
2) $\mathbf{0} \leq \boldsymbol{u} < \boldsymbol{v} \Longrightarrow \boldsymbol{Av} < \boldsymbol{Av}$.

Důkaz: 1) Spočítáme i-tou složku vektoru $\boldsymbol{Av}$ jako $\sum_{j=1}^{n} a_{ij}v_j$. Z $\boldsymbol{v} \neq \mathbf{0}$ víme, že existuje k t.ž. $v_k > 0$. Dále také víme, že $a_{ij} > 0$ pro všechna i, j, takže dostáváme $0 < a_{ik}v_k \leq \sum_{j=1}^{n} a_{ij}v_j$ pro každé $i = 1, 2, \ldots, n$.

2) Opět se zaměříme na i-té složky porovnávaných vektorů. Předpoklad $\boldsymbol{u} < \boldsymbol{v}$ znamená, že $u_j < v_j$ pro všechna $j = 1, 2, \ldots, n$ a z toho hned vidíme výsledek

$$(\boldsymbol{Au})_i = \sum_{j=1}^{n} a_{ij}u_j < \sum_{j=1}^{n} a_{ij}v_j = (\boldsymbol{Av})_i \ .$$

$\square$

Lemma 3.20. Je-li λ vlastní číslo reálné matice $\boldsymbol{A}$ řádu n, pak je λ vlastní číslo také matice $\boldsymbol{A}^T$.

Důkaz: Připomeňme, že vlastní čísla $\boldsymbol{A}$ můžeme spočítat jako kořeny charakteristického polynomu matice $\boldsymbol{A}$, který budeme značit $\mathcal{X}_{\boldsymbol{A}}$ a je definován jako

$$\mathcal{X}_{\boldsymbol{A}}(x) = det(\boldsymbol{A} - x\boldsymbol{I}_n).$$

Všimneme si, že $(\boldsymbol{A} - x\boldsymbol{I}_n)^T = \boldsymbol{A}^T - (x\boldsymbol{I}_n)^T = \boldsymbol{A}^T - x\boldsymbol{I}_n$, protože $x\boldsymbol{I}_n$ je symetrická matice. Také víme, že $det(\boldsymbol{B}) = det(\boldsymbol{B}^T)$ pro libovolnou reálnou čtvercovou matici $\boldsymbol{B}$. Potom dostaneme

$$\mathcal{X}_{\boldsymbol{A}}(x) = det(\boldsymbol{A} - x\boldsymbol{I}_n) = det((\boldsymbol{A} - x\boldsymbol{I}_n)^T) = det(\boldsymbol{A}^T - x\boldsymbol{I}_n) = \mathcal{X}_{\boldsymbol{A}^T}(x).$$

Máme tedy rovnost charakteristických polynomů $\boldsymbol{A}$ a $\boldsymbol{A}^T$. Z toho dostáváme, že se všechna vlastní čísla $\boldsymbol{A}$ a $\boldsymbol{A}^T$ musí shodovat. □

Perronova věta 3.21. Nechť $\boldsymbol{A}$ je reálná kladná matice řádu n. Pak platí následující tvrzení.
1) Existuje reálné vlastní číslo λ_1 matice $\boldsymbol{A}$ takové, že $\lambda_1 > 0$.
2) Existuje vlastní vektor $\boldsymbol{\pi}$ matice $\boldsymbol{A}$ příslušný vlastnímu číslu λ_1, pro který platí $\boldsymbol{\pi} > \boldsymbol{0}$ a $\|\boldsymbol{\pi}\|_1 = 1$.
3) Pro každé vlastní číslo $\lambda \neq \lambda_1$ matice $\boldsymbol{A}$ platí $|\lambda| < \lambda_1$.
4) Algebraická násobnost vlastního čísla λ_1 je 1.

Důkaz: Na úvod poznamenejme, že matice $\boldsymbol{A}$ má alespoň jedno vlastní číslo různé od nuly. Kdyby tomu tak nebylo, tak pro vhodné pevné k platí $\boldsymbol{A}^k = \boldsymbol{T}(\boldsymbol{J}_A)^k\boldsymbol{T}^{-1} = \boldsymbol{0}$, kde $(\boldsymbol{J}_A)^k$ je Jordanův kanonický tvar matice $\boldsymbol{A}$. To ale nemůže platit, protože $\boldsymbol{A}$ je kladná matice.

Dále budeme uvažovat $\rho(\boldsymbol{A}) = 1$, kde $\rho(\boldsymbol{A})$ je spektrální poloměr matice $\boldsymbol{A}$. Pokud by totiž $\rho(\boldsymbol{A}) \neq 1$, tak budeme počítat s maticí $\boldsymbol{A}' = \frac{\boldsymbol{A}}{\rho(\boldsymbol{A})}$, ta je rovněž kladná a reálná a následně použijeme lemma 3.18.

1), 2) Buď λ vlastní číslo matice $\boldsymbol{A}$ takové, že $|\lambda| = 1$ a $0 \neq \boldsymbol{v}$ vlastní vektor příslušný λ. Budeme chtít ukázat, že potom je 1 vlastní číslo matice $\boldsymbol{A}$ s příslušným vlastním vektorem $|\boldsymbol{v}|$.

S využitím trojúhelníkové nerovnosti vidíme, že platí $|\boldsymbol{A}\boldsymbol{v}| \leq |\boldsymbol{A}|\,|\boldsymbol{v}|$. Pokud si ještě všimneme, že platí $|\boldsymbol{A}| = \boldsymbol{A}$, protože $\boldsymbol{A}$ je kladná matice, dostáváme

$$1\,|\boldsymbol{v}| = |\lambda|\,|\boldsymbol{v}| = |\lambda\boldsymbol{v}| = |\boldsymbol{A}\boldsymbol{v}| \leq |\boldsymbol{A}|\,|\boldsymbol{v}| = \boldsymbol{A}\,|\boldsymbol{v}|\ .$$

Máme $1\,|\boldsymbol{v}| \leq \boldsymbol{A}\,|\boldsymbol{v}|$, naším cílem bude ukázat rovnost a uděláme to sporem.

Předpokládejme tedy $|\boldsymbol{v}| < \boldsymbol{A}\,|\boldsymbol{v}|$. Přenásobením obou stran nerovnice kladnou maticí $\boldsymbol{A}$ dostaneme, že také $\boldsymbol{A}\,|\boldsymbol{v}| < \boldsymbol{A}\,(\boldsymbol{A}\,|\boldsymbol{v}|)$ dle pozorování 3.19. Máme zde ostrou nerovnost, a proto existuje malé $\epsilon > 0$ takové, že bude pořád platit $\boldsymbol{A}\,|\boldsymbol{v}| < \frac{\boldsymbol{A}}{1+\epsilon}\,(\boldsymbol{A}\,|\boldsymbol{v}|)$. Matice $\frac{\boldsymbol{A}}{1+\epsilon}$ je kladná, a proto po vynásobení obou stran nerovnice touto maticí dostaneme opět $\frac{\boldsymbol{A}}{1+\epsilon}\,(\boldsymbol{A}\,|\boldsymbol{v}|) < \frac{\boldsymbol{A}}{1+\epsilon}\left(\frac{\boldsymbol{A}}{1+\epsilon}\,(\boldsymbol{A}\,|\boldsymbol{v}|)\right)$. Takto můžeme nerovnici maticí $\frac{\boldsymbol{A}}{1+\epsilon}$ přenásobovat pořád dál a dostaneme nakonec řetězec

$$\boldsymbol{A}\,|\boldsymbol{v}| < \frac{\boldsymbol{A}}{1+\epsilon}\boldsymbol{A}\,|\boldsymbol{v}| < \left(\frac{\boldsymbol{A}}{1+\epsilon}\right)^2\boldsymbol{A}\,|\boldsymbol{v}| < \left(\frac{\boldsymbol{A}}{1+\epsilon}\right)^3\boldsymbol{A}\,|\boldsymbol{v}| < \ldots$$

Užitím lemma 3.18 vidíme, že $\rho(\frac{\boldsymbol{A}}{1+\epsilon}) = \frac{1}{1+\epsilon} < 1$ a z mocnění Jordanových buněk pak získáme

$$\lim_{k\to\infty}\left(\frac{\boldsymbol{A}}{1+\epsilon}\right)^k = \lim_{k\to\infty}\boldsymbol{T}(\boldsymbol{J}_{\frac{\boldsymbol{A}}{1+\epsilon}})^k\boldsymbol{T}^{-1} = \boldsymbol{0}\ .$$

Z toho plyne, že $\boldsymbol{A}\,|\boldsymbol{v}| < \boldsymbol{0}$, ale z definice $\boldsymbol{A}\,|\boldsymbol{v}|$ víme, že to není možné.

Došli jsme tedy ke sporu a máme $|\boldsymbol{v}| = \boldsymbol{A}\,|\boldsymbol{v}|$. To znamená, že 1 je vlastní číslo matice $\boldsymbol{A}$ s příslušným kladným vlastním vektorem $|\boldsymbol{v}| = \boldsymbol{\pi}$. K tomu, aby součet komponent vektoru $\boldsymbol{\pi}$ byl roven jedné, stačí tento vektor přenásobit vhodnou konstantou.

3) Vezměme si nějaké vlastní číslo λ matice $\boldsymbol{A}$ pro které platí $|\lambda| = 1$ a jemu příslušný vlastní vektor $\boldsymbol{v} \neq \boldsymbol{0}$. Z důkazu 1) a 2) už víme, že $|\boldsymbol{v}| = \boldsymbol{A}\,|\boldsymbol{v}|$. Nejprve ukážeme, že $\boldsymbol{A}$ má reálný vlastní vektor příslušný λ a že složky tohoto vektoru mají stejné znaménko a nakonec, že nutně $\lambda = 1$.

Nyní budeme počítat dvěma způsoby, čemu se rovná k-tá složka vektoru $\boldsymbol{v}$.

$$|v_k| = (\boldsymbol{A}\,|\boldsymbol{v}|)_k = \sum_{j=1}^{n} a_{kj}\,|v_j| = \sum_{j=1}^{n} |a_{kj}v_j|$$

$$|v_k| = |\lambda|\,|v_k| = |\lambda v_k| = |(\lambda \boldsymbol{v})_k| = |(\boldsymbol{A}\boldsymbol{v})_k| = \left|\sum_{j=1}^{n} a_{kj}v_j\right|.$$

Dostali jsme tedy

$$\sum_{j=1}^{n} |a_{kj}v_j| = \left|\sum_{j=1}^{n} a_{kj}v_j\right|$$

a z důsledku 3.17 víme, že to nastane právě tehdy, existují-li reálná $\alpha_j > 0$ t.ž. $a_{kj}v_j = \alpha_j a_{k1}v_1$. Z toho hned vidíme, že reálné i imaginární části prvků $v_1, v_2, \ldots, v_n$ mají stejná znaménka a také, že

$$\frac{\boldsymbol{v}}{v_1} \in \mathbb{R}^n \text{ a označíme } \boldsymbol{v}'' = \frac{\boldsymbol{v}}{v_1}.$$

Díky tomu dostáváme $\lambda\boldsymbol{v}'' = \boldsymbol{A}\boldsymbol{v}''$, tedy $\boldsymbol{v}''$ je vlastním vektorem příslušným vlastnímu číslu λ, a také $\boldsymbol{v}'' \leq \boldsymbol{0}$, nebo $\boldsymbol{v}'' \geq \boldsymbol{0}$.

Protože $\boldsymbol{v}$ má alespoň jednu složku nenulovou a $\boldsymbol{A} > \boldsymbol{0}$, platí $\boldsymbol{A}\,|\boldsymbol{v}| > \boldsymbol{0}$ (pozorování 3.19). Z $|\boldsymbol{v}| = \boldsymbol{A}\,|\boldsymbol{v}|$, také vidíme, že platí $|\boldsymbol{v}| > \boldsymbol{0}$. Z toho dostáváme také $|\boldsymbol{v}''| > \boldsymbol{0}$ a dohromady tedy buď $\boldsymbol{v}'' > \boldsymbol{0}$, nebo $\boldsymbol{v}'' < \boldsymbol{0}$.

Nyní definujme vektor $\boldsymbol{v}' = |\boldsymbol{v}''|$. Takovýto vektor je určitě kladný a zároveň je to pořád vlastní vektor matice $\boldsymbol{A}$ příslušný λ, protože je to buď přímo vektor $\boldsymbol{v}''$, nebo vektor $-\boldsymbol{v}''$. Hledaný výsledek už dostaneme jednoduchým výpočtem.

$$\begin{aligned}
\lambda\boldsymbol{v}' &= \boldsymbol{A}\boldsymbol{v}' && (\boldsymbol{v}' \text{ je vlastní vektor příslušný } \lambda)\\
&= |\boldsymbol{A}\boldsymbol{v}'| && (\boldsymbol{A} > \boldsymbol{0} \text{ a } \boldsymbol{v}' > \boldsymbol{0})\\
&= |\lambda\boldsymbol{v}'| && (\boldsymbol{v}' \text{ je vlastní vektor příslušný } \lambda)\\
&= |\lambda|\,\boldsymbol{v}' && (|\lambda v_i'| = |\lambda|\,|v_i'| = |\lambda|\,v_i', \text{ protože } v_i' > 0)\\
&= 1\boldsymbol{v}'
\end{aligned}$$

Zjistili jsme, že $\lambda = 1$, tedy, že 1 je jediným vlastním číslem na spektrálním kruhu matice $\boldsymbol{A}$. To znamená, že pro libovolné vlastní číslo λ' matice $\boldsymbol{A}$ platí $|\lambda'| < 1$.

4) Již jsme dokázali, že pro libovolný nenulový vlastní vektor $\boldsymbol{v}$ matice $\boldsymbol{A}$ příslušný vlastnímu číslu 1 platí, že všechny jeho složky mají stejné znaménko.

Bez újmy na obecnosti můžeme dále předpokládat $\boldsymbol{v} > \mathbf{0}$, neboť jinak bychom ho přenásobili -1.

Nejprve dokážeme, že $geo(1) = 1$ (geometrická násobnost). Vezmeme dva vlastní vektory $\boldsymbol{u} > \mathbf{0}$ a $\boldsymbol{v} > \mathbf{0}$ matice $\boldsymbol{A}$ příslušné 1. Protože platí

$$\boldsymbol{A}\left(v_1\boldsymbol{u} - u_1\boldsymbol{v}\right) = v_1\boldsymbol{A}\boldsymbol{u} - u_1\boldsymbol{A}\boldsymbol{v} = v_1 1\boldsymbol{u} - u_1 1\boldsymbol{v} = 1\left(v_1\boldsymbol{u} - u_1\boldsymbol{v}\right) \ ,$$

je také vektor $(v_1\boldsymbol{u} - u_1\boldsymbol{v})$ vlastním vektorem matice $\boldsymbol{A}$. Tento vektor má první složku rovnu nule, ale na začátku této části jsme již zmínili, že každá jeho složka musí mít stejné znaménko. Z toho plyne, že $(v_1\boldsymbol{u} - u_1\boldsymbol{v}) = 0$, neboli $\boldsymbol{v} = \left(\frac{v_1}{u_1}\right)\boldsymbol{u}$. Každé dva vlastní vektory matice $\boldsymbol{A}$ příslušné 1 jsou lineárně závislé. Dostáváme tedy $geo(1) = 1$.

Nyní přichází na řadu algebraická násobnost. Pro spor předpokládejme, že $alg(1) > 1$. Vezměme vlastní vektor $\boldsymbol{v}$ matice $\boldsymbol{A}$ příslušný 1 a vektor $\boldsymbol{u}$ takový, že $(\boldsymbol{A} - 1\boldsymbol{I})\,\boldsymbol{u} = \boldsymbol{v}$, neboli $\boldsymbol{A}\boldsymbol{u} = \boldsymbol{u} + \boldsymbol{v}$ (tvoří Jordanův řetízek).

Podle lemmatu 3.20 víme, že 1 je největší vlastní číslo také matice $\boldsymbol{A}^T > \mathbf{0}$, takže můžeme vzít její vlastní vektor $\boldsymbol{w} > \mathbf{0}$ příslušný 1. Nyní si na pomoc vezmeme skalární součin a téměř okamžitě získáme výsledek.

$$\begin{aligned}(\boldsymbol{w}\cdot\boldsymbol{u}) &= (1\boldsymbol{w}\cdot\boldsymbol{u}) = \left(\boldsymbol{A}^T\boldsymbol{w}\cdot\boldsymbol{u}\right) = \left(\boldsymbol{A}^T\boldsymbol{w}\right)^T\boldsymbol{u}\\ &= \boldsymbol{w}^T\boldsymbol{A}\boldsymbol{u} = (\boldsymbol{w}\cdot\boldsymbol{A}\boldsymbol{u}) = (\boldsymbol{w}\cdot(\boldsymbol{u}+\boldsymbol{v}))\\ &= (\boldsymbol{w}\cdot\boldsymbol{u}) + (\boldsymbol{w}\cdot\boldsymbol{v})\end{aligned}$$

Z toho plyne, že $(\boldsymbol{w}\cdot\boldsymbol{v}) = 0$ což je spor, neboť $\boldsymbol{w} > \mathbf{0}$ a $\boldsymbol{v} > \mathbf{0}$.

Dokázali jsme tedy, že algebraická násobnost vlastního čísla 1 je 1, a tím jsme zakončili i celý důkaz Perronovy věty. □

Google matice $\boldsymbol{G}$ má podle důsledku 3.5 všechna vlastní čísla menší nebo rovné 1. Z předchozí věty vidíme, že má vlastní číslo 1 a všechna ostatní vlastní čísla jsou v absolutní hodnotě ostře menší než 1. Na tomto faktu je založen samotný výpočet vektoru $\boldsymbol{\pi}$, kterému budeme říkat Perronův vektor, a naším dalším cílem bude ukázat, že právě on určuje důležitosti stránek pro Google matici.

V této části jsme vycházeli z knížky Matrix Analysis and Applied Linear Algebra (Meyer, 2001, kap. 8), kde se dá nalézt spoustu užitečných faktů z lineární algebry.

Na závěr této části ještě zformulujeme Perronovu - Frobeniovu větu. Využijeme ji v další kapitole, kde nebudeme pracovat pouze s kladnými maticemi. Její důkaz můžeme najít například v knize Dynamical Systems (viz Sternberg, 2010, kap. 9).

Perronova - Frobeniova věta 3.22. Nechť $\boldsymbol{A}$ je reálná nezáporná ireducibilní matice řádu n. Pak platí následující tvrzení.
1) Existuje reálné vlastní číslo λ_1 matice $\boldsymbol{A}$ takové, že $\lambda_1 > 0$.
2) Existuje vlastní vektor $\boldsymbol{\pi}$ matice $\boldsymbol{A}$ příslušný vlastnímu číslu λ_1, pro který platí $\boldsymbol{\pi} > \mathbf{0}$ a $\|\boldsymbol{\pi}\|_1 = 1$.
3) Pro každé vlastní číslo $\lambda \neq \lambda_1$ matice $\boldsymbol{A}$ platí $|\lambda| \leq \lambda_1$.
4) Algebraická násobnost vlastního čísla λ_1 je 1.
5) Pokud je navíc $\boldsymbol{A}$ primitivní, tak pro každé vlastní číslo $\lambda \neq \lambda_1$ matice $\boldsymbol{A}$ platí $|\lambda| < \lambda_1$.

3.4 Výpočet vektoru popularity

Lemma 3.23. Nechť $\boldsymbol{A} > \boldsymbol{0}$ je čtvercová matice řádu n taková, že součet prvků v každém řádku je 1 (řádkově stochastická). Pak vlastní vektor $\boldsymbol{v}$ matice $\boldsymbol{A}$ příslušný vlastnímu číslu 1 vypadá jako $\boldsymbol{e}c$, kde $c > 0$.

Důkaz: Matice $\boldsymbol{A}^T$ je stochastická, a tak má vlastní číslo 1. Podle důsledku 3.5 je to největší (v absolutní hodnotě) vlastní číslo této matice a podle lemmatu 3.20 platí to samé pro matici $\boldsymbol{A}$. Máme tedy $\boldsymbol{Av} = \boldsymbol{v}$, pro nějaký vektor $\boldsymbol{v}$. Nyní použijeme Perronovu větu a dostaneme, že $\boldsymbol{v} > \boldsymbol{0}$. Podrobnější rozepsání násobení nám dává

$$\boldsymbol{v} = \boldsymbol{Av} = \begin{pmatrix} a_{11}v_1 + a_{12}v_2 + a_{13}v_3 + \ldots + a_{1n}v_n \\ a_{21}v_1 + a_{22}v_2 + a_{23}v_3 + \ldots + a_{2n}v_n \\ a_{31}v_1 + a_{32}v_2 + a_{33}v_3 + \ldots + a_{3n}v_n \\ \vdots \\ a_{n1}v_1 + a_{n2}v_2 + a_{n3}v_3 + \ldots + a_{nn}v_n \end{pmatrix}.$$

Pokud by bylo $v_1 = v_2 = \ldots = v_n = c$, pro $c > 0$, tak by rovnost platila, neboť $\boldsymbol{A}$ je řádkově stochastická. Z Perronovy věty ale víme, že matice $\boldsymbol{A}$ má až na násobky jediný vlastní vektor příslušný vlastnímu číslu 1. Toto je tedy jediné možné řešení pro vektor $\boldsymbol{v}$. □

Věta 3.24. Buď $\boldsymbol{P}$ přechodová matice pro Markovův řetězec taková, že $\boldsymbol{P} > \boldsymbol{0}$. Buď navíc $\boldsymbol{\pi}$ její Perronův vektor. Potom platí
1) $\lim_{k\to\infty} \boldsymbol{P}^k = \big(\boldsymbol{\pi}|\boldsymbol{\pi}|\ldots|\boldsymbol{\pi}\big)$,
2) $\lim_{k\to\infty} \boldsymbol{p}(k) = \boldsymbol{\pi}$, pro libovolný počáteční pravděpodobnostní vektor $\boldsymbol{p}(0)$.

Důkaz: ad 1) Z Perronovy věty a z toho, že matice $\boldsymbol{P}$ je stochastická víme, že 1 jako vlastní číslo matice $\boldsymbol{P}$ má algebraickou násobnost 1. Z teorie o Jordanově kanonickém tvaru víme, že existuje matice $\boldsymbol{T}$ řádu n taková, že

$$\boldsymbol{P} = \boldsymbol{T} \begin{pmatrix} 1 & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{J} \end{pmatrix} \boldsymbol{T}^{-1},$$

kde $\boldsymbol{J}$ je v blokově diagonálním tvaru tvořená Jordanovými buňkami. Navíc se v prvním sloupci matice $\boldsymbol{T}$ nachází nějaký vlastní vektor příslušný vlastnímu číslu 1, což je $\boldsymbol{\pi}$. To můžeme vysvětlit tak, že s předchozí rovnicí je ekvivalentní zápis

$$\boldsymbol{PT} = \boldsymbol{T} \begin{pmatrix} 1 & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{J} \end{pmatrix}$$

a když porovnáme první sloupce matic na obou stranách, dostáváme $\boldsymbol{PT}_{\star 1} = \boldsymbol{T}_{\star 1}1$, neboli $\boldsymbol{PT}_{\star 1} = 1\boldsymbol{T}_{\star 1}$, kde $\boldsymbol{T}_{\star 1}$ značí první sloupec matice $\boldsymbol{T}$.

Když si nyní vzpomeneme, jak se mocní Jordanova buňka a vůbec celá matice v Jordanově kanonickém tvaru, uvidíme, že $\lim_{k\to\infty} \boldsymbol{J}^k = \boldsymbol{0}$, protože pro každé vlastní číslo $\lambda \neq 1$ matice $\boldsymbol{P}$ platí $|\lambda| < 1$. Potom tedy dostáváme

$$\lim_{k\to\infty} \boldsymbol{P}^k = \lim_{k\to\infty} \boldsymbol{T} \begin{pmatrix} 1^k & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{J}^k \end{pmatrix} \boldsymbol{T}^{-1} = \boldsymbol{T} \begin{pmatrix} 1 & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{pmatrix} \boldsymbol{T}^{-1}.$$

Označíme-li prvky matice $\boldsymbol{T}$ t_{ij} a prvky matice $\boldsymbol{T}^{-1}$ t'_{ij}, tak po roznásobení dostáváme

$$\lim_{k\to\infty} \boldsymbol{P}^k = \begin{pmatrix} t_{11}t'_{11} & t_{11}t'_{12} & t_{11}t'_{13} & \dots & t_{11}t'_{1n} \\ t_{21}t'_{11} & t_{21}t'_{12} & t_{21}t'_{13} & \dots & t_{21}t'_{1n} \\ t_{31}t'_{11} & t_{31}t'_{12} & t_{31}t'_{13} & \dots & t_{31}t'_{1n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ t_{n1}t'_{11} & t_{n1}t'_{12} & t_{n1}t'_{13} & \dots & t_{n1}t'_{1n} \end{pmatrix}.$$

Nyní se blíže podívejme vektor $\begin{pmatrix} t_{11} & t_{21} & t_{31} & \dots & t_{n1} \end{pmatrix}^T$, což je první sloupec matice $\boldsymbol{T}$, tedy $\boldsymbol{\pi}$, a vektor $\begin{pmatrix} t'_{11} & t'_{12} & t'_{13} & \dots & t'_{1n} \end{pmatrix}^T$, což je první řádek matice $\boldsymbol{T}^{-1}$. První řádek matice $\boldsymbol{T}^{-1}$ je vlastním vektorem matice $\boldsymbol{P}^T$ příslušný vlastnímu číslu 1. To proto, že

$$\boldsymbol{P}^T = (\boldsymbol{T}^{-1})^T \begin{pmatrix} 1 & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{J}^T \end{pmatrix} \boldsymbol{T}^T.$$

Z lemma 3.23 víme, že tento vektor vypadá jako $\boldsymbol{e}c$. Zbývá nám tedy určit konstantu c.

Z rovnosti $\boldsymbol{T}^{-1}\boldsymbol{T} = \boldsymbol{I}_n$ dostáváme $t_{11}t'_{11} + t_{21}t'_{12} + t_{31}t'_{13} + \ldots + t_{n1}t'_{1n} = 1$. Víme, že $t_{i1} > 0$, pro všechna $i = 1, 2, \ldots, n$ a také, že $t'_{1i} = c > 0$, pro všechna $i = 1, 2, \ldots, n$. Dostáváme tedy $1 = (t_{11} + t_{21} + t_{31} + \ldots + t_{n1})c = 1c = c$.

Pokud nyní označíme $\boldsymbol{\pi} = (\pi_1, \pi_2, \ldots, \pi_n)$, tak ve výsledku dostáváme

$$\lim_{k\to\infty} \boldsymbol{P}^k = \begin{pmatrix} \pi_1 & \pi_1 & \dots & \pi_1 \\ \pi_2 & \pi_2 & \dots & \pi_2 \\ \vdots & \vdots & \ddots & \vdots \\ \pi_n & \pi_n & \dots & \pi_n \end{pmatrix} = \big(\boldsymbol{\pi}|\boldsymbol{\pi}|\ldots|\boldsymbol{\pi}\big).$$

ad 2) Plyne z 1) a tvrzení 3.9. Nejprve si všimněme, že $\lim_{k\to\infty} \boldsymbol{P}^k = \big(\boldsymbol{\pi}|\boldsymbol{\pi}|\ldots|\boldsymbol{\pi}\big) = \boldsymbol{\pi e}$. Protože pro každý pravděpodobnostní vektor platí $\boldsymbol{ep}(0) = 1$, můžeme psát

$$\lim_{k\to\infty} \boldsymbol{p}(k) = \lim_{k\to\infty} \boldsymbol{P}^k \boldsymbol{p}(o) = \boldsymbol{\pi e p}(0) = \boldsymbol{\pi}\ .$$

□

Tvrzení, která jsme uvedli, platí pro Google matici $\boldsymbol{G}$, kterou zkoumáme. Můžeme tedy konečně uvést výsledek, ke kterému jsme směřovali.

Už dříve jsme nastínili, že důležitost stránek získáme součtem pravděpodobnostních vektorů. Když nyní víme, že tyto vektory konvergují k Perronovu vektoru, bude k němu také konvergovat součet $\sum_{k=1}^{m} \frac{\boldsymbol{p}(k)}{m}$. Vydělení číslem m nám nezmění poměry mezi jednotlivými složkami, a tedy nám vydělení číslem m nezmění ani pořadí stránek. Ve výsledku můžeme říci, že seřazení stránek podle důležitosti můžeme provést pomocí vektoru $\boldsymbol{\pi}$.

Jak v současné době Google fakticky počítá Perronův vektor, není známo. My si zde pouze v příkladu naznačíme jednu z metod, které jsou pro počítání vlastního vektoru tak velké matice vůbec možné. Anglicky se jí říká Power method, a jak naznačuje název, trik je v umocňování dané matice.

Příklad: Mějme Google matici $\boldsymbol{G}$ a pro zjednodušení předpokládejme, že je diagonalizovatelná. Víme, že má vlastní číslo 1 a ostatní jsou v absolutní hodnotě menší. Označme je a seřaďme je podle velikosti takto $1 > |\lambda_2| \geq |\lambda_3| \geq \ldots \geq |\lambda_n|$. Těmto vlastním číslům odpovídají vektory $\boldsymbol{\pi}, \boldsymbol{v}_2, \boldsymbol{v}_3, \ldots, \boldsymbol{v}_n$, které tvoří bázi $\mathbb{C}^n$.

Nyní zvolíme libovolný vektor $\boldsymbol{p}$ a vyjádříme ho jako $\boldsymbol{p} = a_1\boldsymbol{\pi} + a_2\boldsymbol{v}_2 + \ldots + a_n\boldsymbol{v}_n$. Konečně přichází na řadu slibované umocňování maticí $\boldsymbol{G}$.

$$\begin{aligned}\boldsymbol{G}\boldsymbol{p} &= a_1\boldsymbol{G}\boldsymbol{\pi} + a_2\boldsymbol{G}\boldsymbol{v}_2 + \ldots + a_n\boldsymbol{G}\boldsymbol{v}_n \\ &= a_1\boldsymbol{\pi} + a_2\lambda_2\boldsymbol{v}_2 + \ldots + a_n\lambda_n\boldsymbol{v}_n\end{aligned}$$

Pokud výsledek znovu přenásobíme maticí $\boldsymbol{G}$, dostaneme

$$\begin{aligned}\boldsymbol{G}^2\boldsymbol{p} &= a_1\boldsymbol{G}\boldsymbol{\pi} + a_2\lambda_2\boldsymbol{G}\boldsymbol{v}_2 + \ldots + a_n\lambda_n\boldsymbol{G}\boldsymbol{v}_n \\ &= a_1\boldsymbol{\pi} + a_2\lambda_2^2\boldsymbol{v}_2 + \ldots + a_n\lambda_n^2\boldsymbol{v}_n\ .\end{aligned}$$

Vidíme, jak mocnění bude pokračovat dál, a proto napíšeme obecný vzoreček

$$\boldsymbol{G}^k\boldsymbol{p} = a_1\boldsymbol{\pi} + a_2\lambda_2^k\boldsymbol{v}_2 + \ldots + a_n\lambda_n^k\boldsymbol{v}_n\ .$$

Uvědomme si, jak velká jsou čísla λ_i, je to $1 > |\lambda_i|$ pro všechna $i = 2, 3, \ldots, n$. Určitě tedy platí $\lim_{k\to\infty} \lambda_i^k = 0$. Z toho také hned dostáváme $\lim_{k\to\infty} \boldsymbol{G}^k\boldsymbol{p} = a_1\boldsymbol{\pi}$

Opakovaným mocněním jsme se dostali až k Perronovu vektoru. Otázkou zůstává, kolik iterací musíme udělat, abychom dostali "dost přesný" výsledek. Odpověď záleží na velikosti vlastního čísla λ_2, protože ostatní vlastní čísla, kromě 1, konvergují k nule stejně rychle nebo ještě rychleji než λ_2.

Ukázalo se, že hodnotu druhého největšího vlastního čísla můžeme ovlivnit různou volbou koeficientu α při sestavování matice $\boldsymbol{G}$ z matice $\boldsymbol{S}$. Pro skutečnou Google matici se uvádí $\alpha = 0,85$ a pro takovouto hodnotu je potřeba pouze asi 50 iterací.

Příklad: V předešlém příkladu jsme si ukázali ideu výpočtu Perronova vektoru. Nyní si opět vezmeme Google matici $\boldsymbol{G}$, ale nebudeme na ni klást žádné další nároky. Tato matice s největší pravděpodobností diagonalizovatelná nebude, a proto nebudeme mít k dispozici bázi složenou z vlastních vektorů. Naštěstí zde ale zafunguje báze, která vznikne spojením Jordanových řetízků příslušných vlastním číslům matice $\boldsymbol{G}$.

Podobně jako v předešlém příkladu seřadíme vlastní čísla matice, to znamená $1 > |\lambda_2| \geq |\lambda_3| \geq \ldots \geq |\lambda_m|$, kde $m \leq n$. Dále označíme vektory zmíněné báze prostoru $\mathbb{C}^n$ jako $\boldsymbol{\pi}, \boldsymbol{v}_2, \boldsymbol{v}_3, \ldots, \boldsymbol{v}_n$ tak, že pokud je vektor $\boldsymbol{v}_i$ součástí Jordanova řetízku příslušného vlastnímu číslu λ_j a zároveň to není jeho počátek, tak platí $\boldsymbol{G}\boldsymbol{v}_i = \boldsymbol{v}_{i-1} + \lambda_j\boldsymbol{v}_i$.

Nyní opět zvolíme libovolný počáteční vektor $\boldsymbol{p}$, vyjádříme ho pomocí vektorů dané báze jako $\boldsymbol{p} = a_1\boldsymbol{\pi} + a_2\boldsymbol{v}_2 + \ldots + a_n\boldsymbol{v}_n$ a budeme chtít vědět, co se stane, když ho budeme stále přenásobovat maticí $\boldsymbol{G}$.

Pokud bude $\boldsymbol{v}_i$ vlastní vektor příslušný vlastnímu číslu λ_j, pak dostaneme jako v předchozím příkladu $\boldsymbol{G}^k\boldsymbol{v}_i = \lambda_j^k\boldsymbol{v}_i$. Pokud ovšem bude $\boldsymbol{v}_i$ součástí Jordanova řetízku příslušného vlastnímu číslu λ_j a nebude to vlastní číslo, tak dostaneme následující.

$$
\begin{aligned}
\boldsymbol{G}\boldsymbol{v}_i &= \boldsymbol{v}_{i-1} + \lambda_j \boldsymbol{v}_i \\
\boldsymbol{G}^2\boldsymbol{v}_i &= \boldsymbol{v}_{i-2} + 2\lambda_j \boldsymbol{v}_{i-1} + \lambda_j^2 \boldsymbol{v}_i \\
&\vdots \\
\boldsymbol{G}^k\boldsymbol{v}_i &= x_l^{(k)} \lambda_j^{k-l} \boldsymbol{v}_{i-l} + x_{l-1}^{(k)} \lambda_j^{k-(l-1)} \boldsymbol{v}_{i-(l-1)} + \ldots + x_1^{(k)} \lambda_j^{k-1} \boldsymbol{v}_{i-1} + x_0^{(k)} \lambda_j^k \boldsymbol{v}_i
\end{aligned}
$$

Vektor $\boldsymbol{v}_{i-l}$ už je vlastním vektorem příslušný vlastnímu číslu λ_j. Jednoduchým rozepsáním bychom také zjistili, že $x_j^{(k)} = x_j^{(k-1)} + x_{j-1}^{(k-1)}$, a proto konvergence $\boldsymbol{G}^k \boldsymbol{v}_i$ bude záležet na hodnotě λ_j.

Uvědomme si, že $1 > |\lambda_2| \geq |\lambda_3| \geq \ldots \geq |\lambda_m|$, takže $\lim_{k\to\infty} \lambda_j^k = 0$. Ihned už dostáváme výsledek jako v minulém příkladu

$$\lim_{k\to\infty} \boldsymbol{G}^k \boldsymbol{p} = \boldsymbol{\pi}.$$

4 Aplikace na Synot ligu

4.1 První ohodnocení

Při čtení toho, jakým způsobem Google ohodnocuje stránky podle důležitosti, musí člověka napadnout, jestli by takovýto způsob šel využít i jinde. Odpověď je zřejmá, ano šel. Co nás pravděpodobně napadne jako první, je použití ohodnocování ve sportu, a sice tam, kde spolu soupeří dva celky, nebo dva jednotlivci. My si podrobně ukážeme aplikaci PageRanku na první českou fotbalovou ligu (Synot ligu).

Naším prvním úkolem bude sestrojit nějaký graf, který bude ilustrovat vzájemné počínání jednotlivých týmů. Ihned nás napadne jak na to, totiž pokud tým A prohraje s týmem B, tak z vrcholu A povede hrana do vrcholu B. Když spolu týmy remizují, tak povedeme hranu oběma směry. Následně sestrojíme matici $\boldsymbol{H}$ stejně jako tomu bylo v případě internetu. Označíme-li l počet hran, které odcházejí z vrcholu A, pak v matici $\boldsymbol{H}$, ve sloupci odpovídajícímu vrcholu A bude hodnota $1/l$ na pozici, která odpovídá odchozí hraně a 0 na pozici těch vrcholů, do kterých z A hrana nevede.

Uvědomme si, že každé dva týmy proti sobě v jedné sezoně nastoupí dvakrát - v podzimní části a v jarní části. My se nejprve zaměříme pouze na výsledky z proběhlé podzimní části sezony 2014/2015 a uvidíme, jestli se naše ohodnocení bude shodovat s výsledky tabulky po podzimní části. Všechny výsledky, které zde budeme uvádět, můžeme najít na www.synotliga.cz (viz LFA a kol., 2015).

Domácí	Bohemians	Brno	Budějovice	Dukla	Hradec K.	Jablonec	Jihlava	Liberec	Mladá B.	Ostrava	Plzeň	Příbram	Slavia	Slovácko	Sparta	Teplice
Bohemians		0:0	3:0		1:1	1:2		2:4					2:0		1:2	
Brno			2:2		3:0			1:0			2:3	1:0		0:1	1:3	1:1
Budějovice								1:1	2:1	1:0	0:4	0:3	1:1	2:2		
Dukla	0:1	1:1	3:1				4:1			0:0		0:0	2:2		1:0	
Hradec K.			2:3	2:4		0:2	0:3	0:0	0:0					2:0		0:0
Jablonec		2:0	0:0	6:0			2:0		1:0					1:1		4:2
Jihlava	1:2	2:0	0:0						0:1		2:1	0:1	1:0			
Liberec				0:0		0:1	2:2		0:0	6:0	1:1	0:0		2:2		
Mladá B.	1:0	5:1		0:1							0:3	4:1	2:1	2:0		3:0
Ostrava	1:0	1:0			2:1	1:2	0:0		1:0							1:1
Plzeň	2:1			2:1	4:0	3:1				2:0					2:0	1:0
Příbram	2:3				1:0	1:4				3:1	2:2				0:1	1:1
Slavia		1:3			1:1	1:1		4:1		3:1	1:0	3:2			0:2	
Slovácko	4:1			5:1			2:0			1:2	0:1	4:1	1:2		0:2	
Sparta			4:0		3:1	2:0	3:0	1:0	1:0	0:1						
Teplice	4:1		4:0	1:0			2:3	2:2					2:1	4:0	0:0	

Tabulka 1: křížová tabulka Synot ligy 2014/2015 po podzimní části

Z uvedené tabulky budeme vycházet a jen pro pořádek uveďme, že domácí zápasy například Plzně jsou uvedeny v řádku označeném Plzeň a venkovní zápasy jsou ve sloupci označeném Plzeň. V grafu vedou z vrcholu Plzeň celkem 4 hrany a to do vrcholů Jihlava, Liberec, Příbram, Slavia. Tyto týmy s Plzní totiž remizovali, nebo nad ní dokonce vyhráli. Z ostatních vrcholů vede naopak hrana směrem do vrcholu Plzeň a stejně tak z vrcholů Liberec a Příbram kvůli zmiňované remíze.

Nyní sestrojíme matici $\boldsymbol{H}$ způsobem, jaký jsme popsali výše. Bude vypadat následovně.

$$
\begin{array}{c}
\begin{matrix}
\text{Bohemians} & \text{Brno} & \text{Budějovice} & \text{Dukla} & \text{Hradec K.} & \text{Jablonec} & \text{Jihlava} & \text{Liberec} & \text{Mladá B.} & \text{Ostrava} & \text{Plzeň} & \text{Příbram} & \text{Slavia} & \text{Slovácko} & \text{Sparta} & \text{Teplice}
\end{matrix} \\
\begin{pmatrix}
0 & \frac{1}{11} & \frac{1}{12} & \frac{1}{10} & \frac{1}{14} & 0 & \frac{1}{10} & 0 & 0 & 0 & 0 & \frac{1}{11} & \frac{1}{10} & 0 & 0 & 0 \\
\frac{1}{10} & 0 & \frac{1}{12} & \frac{1}{10} & \frac{1}{14} & 0 & 0 & \frac{1}{13} & 0 & 0 & 0 & \frac{1}{11} & \frac{1}{10} & 0 & 0 & \frac{1}{10} \\
0 & \frac{1}{11} & 0 & 0 & \frac{1}{14} & \frac{1}{5} & \frac{1}{10} & \frac{1}{13} & \frac{1}{8} & \frac{1}{9} & 0 & 0 & \frac{1}{10} & \frac{1}{10} & 0 & 0 \\
0 & \frac{1}{11} & \frac{1}{12} & 0 & \frac{1}{14} & 0 & \frac{1}{10} & \frac{1}{13} & \frac{1}{8} & \frac{1}{9} & 0 & \frac{1}{11} & \frac{1}{10} & 0 & \frac{1}{4} & 0 \\
\frac{1}{10} & 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{13} & \frac{1}{8} & 0 & 0 & 0 & \frac{1}{10} & \frac{1}{10} & 0 & \frac{1}{10} \\
\frac{1}{10} & \frac{1}{11} & \frac{1}{12} & \frac{1}{10} & \frac{1}{14} & 0 & \frac{1}{10} & \frac{1}{13} & \frac{1}{8} & \frac{1}{9} & 0 & \frac{1}{11} & \frac{1}{10} & \frac{1}{10} & 0 & \frac{1}{10} \\
0 & \frac{1}{11} & \frac{1}{12} & 0 & \frac{1}{14} & 0 & 0 & \frac{1}{13} & 0 & \frac{1}{9} & \frac{1}{4} & 0 & \frac{1}{10} & 0 & 0 & \frac{1}{10} \\
\frac{1}{10} & 0 & \frac{1}{12} & \frac{1}{10} & \frac{1}{14} & 0 & \frac{1}{10} & 0 & \frac{1}{8} & \frac{1}{9} & \frac{1}{4} & \frac{1}{11} & 0 & \frac{1}{10} & 0 & \frac{1}{10} \\
\frac{1}{10} & \frac{1}{11} & 0 & 0 & \frac{1}{14} & 0 & \frac{1}{10} & \frac{1}{13} & 0 & 0 & 0 & \frac{1}{11} & \frac{1}{10} & \frac{1}{10} & 0 & \frac{1}{10} \\
\frac{1}{10} & \frac{1}{11} & 0 & \frac{1}{10} & \frac{1}{14} & 0 & \frac{1}{10} & 0 & \frac{1}{8} & 0 & 0 & 0 & 0 & \frac{1}{10} & \frac{1}{4} & \frac{1}{10} \\
\frac{1}{10} & \frac{1}{11} & \frac{1}{12} & \frac{1}{10} & \frac{1}{14} & \frac{1}{5} & 0 & \frac{1}{13} & \frac{1}{8} & \frac{1}{9} & 0 & \frac{1}{11} & 0 & \frac{1}{10} & \frac{1}{4} & \frac{1}{10} \\
0 & 0 & \frac{1}{12} & \frac{1}{10} & \frac{1}{14} & 0 & \frac{1}{10} & \frac{1}{13} & 0 & \frac{1}{9} & \frac{1}{4} & 0 & 0 & 0 & 0 & \frac{1}{10} \\
0 & 0 & \frac{1}{12} & \frac{1}{10} & \frac{1}{14} & \frac{1}{5} & 0 & \frac{1}{13} & 0 & \frac{1}{9} & \frac{1}{4} & \frac{1}{11} & 0 & \frac{1}{10} & 0 & 0 \\
\frac{1}{10} & \frac{1}{11} & \frac{1}{12} & \frac{1}{10} & 0 & \frac{1}{5} & \frac{1}{10} & \frac{1}{13} & 0 & 0 & 0 & \frac{1}{11} & 0 & 0 & 0 & 0 \\
\frac{1}{10} & \frac{1}{11} & \frac{1}{12} & 0 & \frac{1}{14} & \frac{1}{5} & \frac{1}{10} & \frac{1}{13} & \frac{1}{8} & 0 & 0 & \frac{1}{11} & \frac{1}{10} & \frac{1}{10} & 0 & \frac{1}{10} \\
\frac{1}{10} & \frac{1}{11} & \frac{1}{12} & \frac{1}{10} & \frac{1}{14} & 0 & 0 & \frac{1}{13} & 0 & \frac{1}{9} & 0 & \frac{1}{11} & \frac{1}{10} & \frac{1}{10} & \frac{1}{4} & 0
\end{pmatrix}
\end{array}
$$

Vidíme, že žádný sloupec této matice není nulový, neprovedeme tedy žádnou modifikaci, a proto $\boldsymbol{S} = \boldsymbol{H}$. Nyní by byl ještě na řadě přechod k matici $\boldsymbol{G}$, ale to nebude nutné, protože $\boldsymbol{S}$ je ireducibilní a z Perronovy - Frobeniovy věty dostaneme existenci Perronova vektoru. To, že je tato matice ireducibilní, je jednoduše vidět z odpovídajícího grafu za pomoci tvrzení 3.13. Pro odlišení budeme tuto matici označovat $\boldsymbol{F}$, jako fotbalová matice.

V tabulce 2 uvádíme vypočítaný Perronův vektor $\boldsymbol{\pi}$, pořadí, které nám tento vektor dává, pořadí dle ligové tabulky a rozdíl mezi oběma pořadími.

Vidíme, že nějaké týmy své pořadí nezměnily vůbec, nebo jen o jednu nebo dvě pozice. To by pro nás bylo dostačující. Bohužel jsou ale i takové týmy, které se posunuly o řadu pozic (Liberec o 12, Mladá Boleslav o 9) a to nás rozhodně neuspokojí.

Je nám jasné, že Liberec nemůže být druhý nejlepší tým, když vyhrál pouze dvakrát. Jeho dobré umístění je pravděpodobně způsobeno velkým počtem remíz. Podle našeho modelu má totiž remíza podobnou váhu jako výhra. Když tedy Liberec remizoval s Plzní, posunulo ho to v žebříčku vysoko.

Na druhé straně Mladá Boleslav v pořadí velice klesla navzdory tomu, že má na kontě celkem 7 vítězství. Kdybychom počítali i remízy, je to dohromady 9, ale Liberec měl dohromady 11, je tedy jasně vidět, že s tímto problémem musíme něco udělat.

tým	vektor $\boldsymbol{\pi}$	pořadí dle $\boldsymbol{\pi}$	ligové pořadí	rozdíl
Bohemians	0,0380212	15	11	-4
Brno	0,0447810	14	13	-1
Budějovice	0,0591406	11	15	4
Dukla	0,0685389	7	7	0
Hradec K.	0,0359422	16	16	0
Jablonec	0,0731057	5	3	-2
Jihlava	0,0622501	9	10	1
Liberec	0,0785199	2	14	12
Mladá B.	0,0485277	13	4	-9
Ostrava	0,0605678	10	6	-4
Plzeň	0,0924503	1	1	0
Příbram	0,0636706	8	12	4
Slavia	0,0758742	3	8	5
Slovácko	0,0523298	12	9	-3
Sparta	0,0741432	4	2	-2
Teplice	0,0721369	6	5	-1

Tabulka 2: počáteční ohodnocení dle matice $\boldsymbol{F}$

4.2 Úpravy fotbalové matice

Podobně jako Page a Brin modifikovali matici pro webové stránky, tak také my budeme modifikovat výše zmíněnou matici $\boldsymbol{F}$, abychom dostali věrohodnější výsledky. Jeden z problémů byl, že jsme remízu nějakých týmů považovali za výhru každého z nich. Zkusíme tedy remízám dát menší váhu. Ihned se nabízí následující řešení.

Úpravu budeme předvádět opět na Plzni. Ta má na kontě 2 remízy a 2 prohry. Za každou prohru uděláme Plzni 2 čárky a za každou remízu 1 čárku. Dohromady tedy bude mít 6 čárek. Na pozici odpovídající prohře s daným týmem tedy nyní bude hodnota 2/6 a na pozici odpovídající remíze bude 1/6. Tím jsme docílili toho, že remíza nad Plzní má oproti výhře poloviční váhu. V ligovém bodování má sice váhu třetinovou, ale podle mého názoru si zaslouží trochu větší.

Takto upravíme celou matici $\boldsymbol{F}$ a dostaneme matici $\boldsymbol{F}_2$, kterou zde nebudeme explicitně uvádět, neboť si jí jistě dokážeme představit. Co ale můžeme udělat, je, že spočítáme Perronův vektor $\boldsymbol{\pi}_2$ této matice. Po úpravě je totiž stále ireducibilní ze stejného důvodu jako matice $\boldsymbol{F}$. Opět dostaneme pořadí týmů podle vektoru $\boldsymbol{\pi}_2$, a tak můžeme znovu uvést tabulku, nyní pro $\boldsymbol{F}_2$ (tabulka 3).

Vidíme, že podle vektoru $\boldsymbol{\pi}_2$ už Liberec neposkočil o 12 příček, ale jen o 5. Jediný tým, který nás ještě neuspokojuje je Mladá Boleslav. Zčásti je to tím, že Mladá Boleslav vyhrává nad slabšími soupeři, ale se silnějšími, jako jsou Plzeň nebo Sparta, vždy prohraje. Pokles v žebříčku je tedy oprávněný. S tím souvisí i druhá věc, kterou popíšeme na příkladu.

Dejme tomu, že proti sobě má nastoupit Plzeň a Slavie. Každý tým vsadí na svojí výhru 500 korun. Zápas vyhraje Slavie a po právu jí tedy náleží celá

tým	vektor $\boldsymbol{\pi}_2$	pořadí dle $\boldsymbol{\pi}_2$	ligové pořadí	rozdíl
Bohemians	0,0427721	14	11	-3
Brno	0,0396257	15	13	-2
Budějovice	0,0475641	13	15	2
Dukla	0,0673444	6	7	1
Hradec K.	0,0252078	16	16	0
Jablonec	0,0777956	4	3	-1
Jihlava	0,0731361	5	10	5
Liberec	0,0584954	9	14	5
Mladá B.	0,0536700	11	4	-7
Ostrava	0,0635501	7	6	-1
Plzeň	0,1106770	1	1	0
Příbram	0,0553685	10	12	2
Slavia	0,0836341	3	8	5
Slovácko	0,0500879	12	9	-3
Sparta	0,0889490	2	2	0
Teplice	0,0621216	8	5	-3

Tabulka 3: ohodnocení dle matice $\boldsymbol{F}_2$

výhra 1000 korun. Podívejme se, co to znamená v matici $\boldsymbol{F}_2$. Z pohledu Plzně je vše v pořádku. Ve sloupci Plzně je na odpovídající pozici hodnota 2/6 - to odpovídá dvěma pětistovkám. Když se ale koukneme na sloupec Slavie, zjistíme, že Slavie svou výhru vlastně nedostala - tisícovka měla připadnout jí, ale na pozici odpovídající Slavii je nula.

Zkusíme ještě udělat jednu modifikaci matice. Každý tým hraje 15 zápasů a do každého zápasu vloží dva týmy svůj díl (pomyslných 500 korun). Za každý zápas se tedy rozdělí 2 díly a za všechny zápasy jednoho týmu se rozdělí 30 dílů. Vezměme si opět na pomoc Plzeň. Na pozici, která odpovídá Slavii, bude hodnota 2/30 (Plzeň prohrála). Na pozici, která odpovídá Liberci, bude 1/30 (remíza) a Plzeň si přičte také 1/30. Na pozicích týmů, nad kterými Plzeň zvítězila, budou nuly, ale za každý si Plzeň přičte 2/30. Tuto matici označíme $\boldsymbol{F}_3$ a raději si jí přímo napíšeme.

$$
\begin{array}{c}
\begin{array}{cccccccccccccccc}
\text{Bohemians} & \text{Brno} & \text{Budějovice} & \text{Dukla} & \text{Hradec K.} & \text{Jablonec} & \text{Jihlava} & \text{Liberec} & \text{Mladá B.} & \text{Ostrava} & \text{Plzeň} & \text{Příbram} & \text{Slavia} & \text{Slovácko} & \text{Sparta} & \text{Teplice}
\end{array} \\
\begin{pmatrix}
\frac{12}{30} & \frac{1}{30} & \frac{1}{15} & \frac{1}{15} & \frac{1}{30} & 0 & \frac{1}{15} & 0 & 0 & 0 & 0 & \frac{1}{15} & \frac{1}{15} & 0 & 0 & 0 \\
\frac{1}{30} & \frac{12}{30} & \frac{1}{30} & \frac{1}{30} & \frac{1}{15} & 0 & 0 & \frac{1}{15} & 0 & 0 & 0 & \frac{1}{15} & \frac{1}{15} & 0 & 0 & \frac{1}{30} \\
0 & \frac{1}{30} & \frac{12}{30} & 0 & \frac{1}{15} & \frac{1}{30} & \frac{1}{30} & \frac{1}{30} & \frac{1}{15} & \frac{1}{15} & 0 & 0 & \frac{1}{30} & \frac{1}{30} & 0 & 0 \\
0 & \frac{1}{30} & \frac{1}{15} & \frac{15}{30} & \frac{1}{15} & 0 & \frac{1}{15} & \frac{1}{30} & \frac{1}{15} & \frac{1}{30} & 0 & \frac{1}{30} & \frac{1}{30} & 0 & \frac{1}{15} & 0 \\
\frac{1}{30} & 0 & 0 & 0 & \frac{7}{30} & 0 & 0 & \frac{1}{30} & \frac{1}{30} & 0 & 0 & 0 & \frac{1}{30} & \frac{1}{15} & 0 & \frac{1}{30} \\
\frac{1}{15} & \frac{1}{15} & \frac{1}{30} & \frac{1}{15} & \frac{1}{15} & \frac{23}{30} & \frac{1}{15} & \frac{1}{15} & \frac{1}{15} & \frac{1}{15} & 0 & \frac{1}{15} & \frac{1}{30} & \frac{1}{30} & 0 & \frac{1}{15} \\
0 & \frac{1}{15} & \frac{1}{30} & 0 & \frac{1}{15} & 0 & \frac{13}{30} & \frac{1}{30} & 0 & \frac{1}{30} & \frac{1}{15} & 0 & \frac{1}{15} & 0 & 0 & \frac{1}{15} \\
\frac{1}{15} & 0 & \frac{1}{30} & \frac{1}{30} & \frac{1}{30} & 0 & \frac{1}{30} & \frac{13}{30} & \frac{1}{30} & \frac{1}{15} & \frac{1}{30} & \frac{1}{30} & 0 & \frac{1}{30} & 0 & \frac{1}{30} \\
\frac{1}{15} & \frac{1}{15} & 0 & 0 & \frac{1}{30} & 0 & \frac{1}{15} & \frac{1}{30} & \frac{16}{30} & 0 & 0 & \frac{1}{15} & \frac{1}{15} & \frac{1}{15} & 0 & \frac{1}{15} \\
\frac{1}{15} & \frac{1}{15} & 0 & \frac{1}{30} & \frac{1}{15} & 0 & \frac{1}{30} & 0 & \frac{1}{15} & \frac{15}{30} & 0 & 0 & 0 & \frac{1}{15} & \frac{1}{15} & \frac{1}{30} \\
\frac{1}{15} & \frac{1}{15} & \frac{1}{15} & \frac{1}{15} & \frac{1}{15} & \frac{1}{15} & 0 & \frac{1}{30} & \frac{1}{15} & \frac{1}{15} & \frac{24}{30} & \frac{1}{30} & 0 & \frac{1}{15} & \frac{1}{15} & \frac{1}{15} \\
0 & 0 & \frac{1}{15} & \frac{1}{30} & \frac{1}{15} & 0 & \frac{1}{15} & \frac{1}{30} & 0 & \frac{1}{15} & \frac{1}{30} & \frac{12}{30} & 0 & 0 & 0 & \frac{1}{30} \\
0 & 0 & \frac{1}{30} & \frac{1}{30} & \frac{1}{30} & \frac{1}{30} & 0 & \frac{1}{15} & 0 & \frac{1}{15} & \frac{1}{15} & \frac{1}{15} & \frac{14}{30} & \frac{1}{15} & 0 & 0 \\
\frac{1}{15} & \frac{1}{15} & \frac{1}{30} & \frac{1}{15} & 0 & \frac{1}{30} & \frac{1}{15} & \frac{1}{30} & 0 & 0 & 0 & \frac{1}{15} & 0 & \frac{13}{30} & 0 & 0 \\
\frac{1}{15} & \frac{1}{15} & \frac{1}{15} & 0 & \frac{1}{15} & \frac{1}{15} & \frac{1}{15} & \frac{1}{15} & \frac{1}{15} & 0 & 0 & \frac{1}{15} & \frac{1}{15} & \frac{1}{15} & \frac{23}{30} & \frac{1}{30} \\
\frac{1}{15} & \frac{1}{30} & \frac{1}{15} & \frac{1}{15} & \frac{1}{30} & 0 & 0 & \frac{1}{30} & 0 & \frac{1}{30} & 0 & \frac{1}{30} & \frac{1}{15} & \frac{1}{15} & \frac{1}{30} & \frac{16}{30}
\end{pmatrix}
\end{array}
$$

Z konstrukce této matice je jasné, že zůstane ireducibilní. Najdeme tedy její Perronův vektor $\boldsymbol{\pi}_3$, z něho získáme nové pořadí týmů a porovnáme ho s ligovým pořadím.

tým	vektor $\boldsymbol{\pi}_3$	pořadí dle $\boldsymbol{\pi}_3$	ligové pořadí	rozdíl
Bohemians	0,0274478	14	11	-3
Brno	0,0254287	15	13	-2
Budějovice	0,0305229	13	15	2
Dukla	0,0518597	5	7	2
Hradec K.	0,0126598	16	16	0
Jablonec	0,1283740	3	3	0
Jihlava	0,0496938	7	10	3
Liberec	0,0397459	10	14	4
Mladá B.	0,0442815	9	4	-5
Ostrava	0,0489378	8	6	-2
Plzeň	0,2130720	1	1	0
Příbram	0,0355312	11	12	1
Slavia	0,0603786	4	8	4
Slovácko	0,0340332	12	9	-3
Sparta	0,1467780	2	2	0
Teplice	0,0512547	6	5	-1

Tabulka 4: ohodnocení dle matice $\boldsymbol{F}_3$

Nyní se pokusíme vysvětlit posuny některých týmů oproti ligovému pořadí v tabulce 4. Nejprve se podívejme na Mladou Boleslav. Ta si pohoršila o pět příček, ale jak už jsme řekli, je to způsobeno tím, že sice vyhrává se slabými soupeři, ale není schopná porazit kvalitnější mužstva. Posunu Liberce jsme se již také věnovali. Řekli jsme, že remízám budeme dávat větší váhu, než mají v ligovém bodování a Liberec to jen dokazuje.

Co se Slavie týče, její posun o 4 příčky směrem nahoru je způsoben tím, že vyhrála nad Plzní a tím získala pomyslné bodíky navíc. Plzeň porazila také Jihlava, ale její posun není tak patrný. To si můžeme vysvětlit tím, že například Příbram se v pořadí dostala před ní, protože s Plzní remizovala a navíc nad samotnou Jihlavou vyhrála.

Ukázali jsme si, že posuny v pořadí jsou opodstatněné. Nyní záleží pouze na nás, jaké hodnocení nám více vyhovuje. Ohodnocování týmu pomocí algoritmu PageRank má oproti klasickému bodování výhodu v tom, že vítězství nad silným soupeřem má větší váhu než nad slabším.

Liga se ale nevyhrává pouze podzimní částí. Pokud bychom znali všechny výsledky i z jarní části, můžeme stejným způsobem určit matici $\boldsymbol{F}_3^{'}$ pro jarní část. Následně bychom obě matice sloučili takto $\frac{1}{2}\boldsymbol{F}_3 + \frac{1}{2}\boldsymbol{F}_3^{'}$. Pro matici, kterou bychom získali, bychom vypočetli Perronův vektor a dostali bychom pořadí týmů.

V tento okamžik ale známe pouze výsledky do 20. kola, to znamená pouhých 5 zápasů z jarní části. Křížovou tabulku zde uvádět nebudeme, ale uvedeme rovnou matici $\boldsymbol{F}_3^{'}$, která z této tabulky vznikne stejným způsobem, jako vznikla $\boldsymbol{F}_3$.

$$\begin{array}{c}
\begin{array}{cccccccccccccccc}
\text{Bohemians} & \text{Brno} & \text{Budějovice} & \text{Dukla} & \text{Hradec K.} & \text{Jablonec} & \text{Jihlava} & \text{Liberec} & \text{Mladá B.} & \text{Ostrava} & \text{Plzeň} & \text{Příbram} & \text{Slavia} & \text{Slovácko} & \text{Sparta} & \text{Teplice}
\end{array}\\
\begin{pmatrix}
\frac{23}{30} & 0 & 0 & 0 & 0 & 0 & \frac{1}{30} & 0 & 0 & 0 & 0 & \frac{1}{15} & 0 & 0 & 0 & 0\\
0 & \frac{25}{30} & \frac{1}{15} & 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{30} & 0 & 0 & 0 & 0 & 0 & \frac{1}{15}\\
0 & 0 & \frac{22}{30} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{15}\\
0 & \frac{1}{15} & 0 & \frac{24}{30} & 0 & 0 & 0 & \frac{1}{15} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
\frac{1}{15} & \frac{1}{15} & 0 & 0 & \frac{27}{30} & 0 & 0 & 0 & \frac{1}{30} & 0 & 0 & 0 & 0 & \frac{1}{15} & 0 & 0\\
\frac{1}{15} & 0 & 0 & 0 & 0 & \frac{26}{30} & \frac{1}{30} & 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{15} & \frac{1}{30} & 0\\
\frac{1}{30} & 0 & \frac{1}{15} & 0 & 0 & \frac{1}{30} & \frac{24}{30} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{21}{30} & 0 & \frac{1}{30} & 0 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & \frac{1}{15} & \frac{1}{15} & \frac{1}{30} & 0 & 0 & 0 & \frac{27}{30} & 0 & 0 & 0 & \frac{1}{15} & 0 & 0 & 0\\
0 & \frac{1}{30} & 0 & 0 & 0 & 0 & 0 & \frac{1}{30} & 0 & \frac{24}{30} & 0 & 0 & \frac{1}{30} & 0 & \frac{1}{30} & 0\\
0 & 0 & \frac{1}{15} & \frac{1}{15} & \frac{1}{15} & \frac{1}{15} & 0 & 0 & 0 & 0 & 1 & 0 & \frac{1}{15} & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{15} & \frac{1}{15} & 0 & 0 & \frac{26}{30} & 0 & \frac{1}{15} & 0 & 0\\
0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{15} & 0 & \frac{1}{30} & 0 & 0 & \frac{24}{30} & 0 & 0 & \frac{1}{30}\\
\frac{1}{15} & 0 & 0 & 0 & 0 & 0 & \frac{1}{15} & 0 & 0 & 0 & 0 & 0 & 0 & \frac{24}{30} & 0 & 0\\
0 & 0 & 0 & \frac{1}{15} & 0 & \frac{1}{30} & \frac{1}{15} & 0 & 0 & \frac{1}{30} & 0 & \frac{1}{15} & 0 & 0 & \frac{28}{30} & 0\\
0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{15} & 0 & \frac{1}{15} & 0 & 0 & \frac{1}{30} & 0 & 0 & \frac{25}{30}
\end{pmatrix}
\end{array}$$

Na první pohled vidíme, že tato matice je reducibilní. Můžeme to nahlédnout i tak, že v našem grafu této matice nevede z vrcholu Plzeň ani jedna hrana do jiných vrcholů. Pokud bychom nyní spočítali vlastní vektor příslušný vlastnímu číslu 1, dostali bychom vektor $\boldsymbol{e}_i$, kde i odpovídá pozici Plzně.

Na řadu by mohla přijít úprava s přičtením matice $\frac{1}{16}\boldsymbol{e}\boldsymbol{e}^T$, my ale zkusíme jinou úpravu. Ve sloupci odpovídajícímu Plzni nahradíme 1 ($=\frac{30}{30}$) hodnotou $\frac{29}{30}$ a zbylou $\frac{1}{30}$ rozdělíme rovnoměrně mezi ostatní týmy. Na pozici, kde předtím byla 0, bude tedy nyní hodnota $\frac{1}{450}$. Plzni takto sice nepatrně ublížíme, ale odpovídá to zhruba jedné remíze, což Plzeň bezpečně udrží na první pozici.

Vzniklou matici označíme $\boldsymbol{F}_4^{'}$ a můžeme i přeznačit matici z podzimní části jako $\boldsymbol{F}_4$, která se ale od matice $\boldsymbol{F}_3$ nebude vůbec lišit.

Nyní vyvstává ještě jedna otázka, a sice jakým způsobem spojit výše zmíněné matice. Můžeme to udělat tak, jak už jsme navrhli, to znamená $\frac{1}{2}\boldsymbol{F}_4 + \frac{1}{2}\boldsymbol{F}_4^{'}$. Určíme její Perronův vektor $\boldsymbol{\pi}_4$ a pomocí něj určíme pořadí týmů.

Zmíněný způsob dává stejnou váhu podzimní i jarní části, a proto zkusíme váhu rozdělit podle odehraných kol. Zatím bylo odehráno 15 kol v podzimní části a 5 kol v jarní části. Zmíněné matice tedy můžeme sečíst následovně $\frac{15}{20}\boldsymbol{F}_4 + \frac{5}{20}\boldsymbol{F}_4^{'}$ a její Perronův vektor označíme $\boldsymbol{\pi}_5$.

Poslední tabulka (tabulka 5), kterou zde uvedeme, bude obsahovat vektory $\boldsymbol{\pi}_4$ a $\boldsymbol{\pi}_5$, pořadí dané těmito vektory, pořadí dle ligy po 20. kole a rozdíl pořadí dle $\boldsymbol{\pi}$ oproti ligovému pořadí, jako jsme uváděli dříve.

tým	vektor $\boldsymbol{\pi}_4$	vektor $\boldsymbol{\pi}_5$	pořadí $\boldsymbol{\pi}_4$	pořadí $\boldsymbol{\pi}_5$	ligové pořadí	rozdíl $\boldsymbol{\pi}_4$	rozdíl $\boldsymbol{\pi}_5$
Bohemians	0,0247658	0,0262892	15	15	13	-2	-2
Brno	0,0281786	0,0265399	13	14	11	-2	-3
Budějovice	0,0253164	0,0283478	14	13	15	1	2
Dukla	0,0446738	0,0491986	9	7	6	-3	-1
Hradec K.	0,0194767	0,0150512	16	16	14	-2	-2
Jablonec	0,1101040	0,1203930	3	3	3	0	0
Jihlava	0,0471145	0,0490945	8	8	12	4	4
Liberec	0,0291844	0,0352380	12	11	16	4	5
Mladá B.	0,0523351	0,0476097	5	9	4	-1	-5
Ostrava	0,0501004	0,0495318	6	6	9	3	3
Plzeň	0,2357140	0,2223010	1	1	1	0	0
Příbram	0,0400485	0,0373876	10	10	8	-2	-2
Slavia	0,0571287	0,0593343	4	4	10	6	6
Slovácko	0,0301583	0,0324313	11	12	7	-4	-5
Sparta	0,1572920	0,1510060	2	2	2	0	0
Teplice	0,0484094	0,0502460	7	5	5	-2	0

Tabulka 5: ohodnocení dle matic $(\frac{1}{2}\boldsymbol{F}_4 + \frac{1}{2}\boldsymbol{F}_4^{'})$ a $(\frac{15}{20}\boldsymbol{F}_4 + \frac{5}{20}\boldsymbol{F}_4^{'})$

Ukázali jsme si alternativní hodnocení fotbalové ligy, které má od ligového hodnocení odlišné některé parametry jako například, že remíze dává oproti výhře poloviční váhu. V ligovém hodnocení je to třetinová váha. Posuny týmů jsme si vysvětlili a podle mého názoru jsou opodstatněné. Toto ale není jediný způsob ohodnocení, který můžeme vytvořit.

4.3 Možnost jiného ohodnocení

V této sekci si ještě ukážeme, jak vytvořit fotbalovou matici, aby více odpovídala skutečným ligovým výsledkům. V tabulce 5 jsme si všimli, že se Slavie posunula oproti ligovému pořadí dokonce o šest příček. Zkusíme tedy vytvořit matici $\boldsymbol{F}_6$ jiným způsobem než dříve.

Zatím jsme se vždy snažili držet fotbalovou matici stochastickou a to nás nepatrně omezovalo. Z Perronovy - Frobeniovy věty víme, že pokud matici udržíme nezápornou a ireducibilní, tak budeme mít pořád k dispozici Perronův vektor. Podle tohoto vektoru opět zkusíme seřadit týmy a uvidíme, co nám vyjde.

Matici $\boldsymbol{F}_6$ tedy sestrojíme následujícím způsobem. Do matice budeme zapisovat přímo body, které jednotlivé týmy získaly. Ve sloupci odpovídajímu například Plzni zapíšeme 0 na pozici odpovídající týmu, nad kterým Plzeň zvítězila, 1 na pozici odpovídající týmu, se kterým remizovala a 3 na pozici odpovídající týmu, se kterým prohrála. Pokud spolu týmy sehrály více zápasů, tak jednotlivé výsledky jednoduše sečteme.

V našem případě získáme jistě ireducibilní matici, neboť v podzimní části soutěže už hrál každý tým s každým, a proto bude graf této matice silně souvislý. Spočteme tedy vektor $\boldsymbol{\pi}_6$ a podobně jako v předchozí sekci jej porovnáme s ligovým pořadím.

tým	vektor $\boldsymbol{\pi}_6$	pořadí dle $\boldsymbol{\pi}_6$	ligové pořadí	rozdíl
Bohemians	0,0460106	13	13	0
Brno	0,0470046	12	11	-1
Budějovice	0,0422385	14	15	1
Dukla	0,0591685	8	6	-2
Hradec K.	0,0382386	16	14	-2
Jablonec	0,0900784	3	3	0
Jihlava	0,0565120	9	12	3
Liberec	0,0383252	15	16	1
Mladá B.	0,0705605	4	4	0
Ostrava	0,0609948	5	9	4
Plzeň	0,1213230	1	1	0
Příbram	0,0554022	10	8	-2
Slavia	0,0592607	7	10	3
Slovácko	0,0519274	11	7	-4
Sparta	0,1026420	2	2	0
Teplice	0,0603125	6	5	-1

Tabulka 6: ohodnocení dle matice $\boldsymbol{F}_6$

Z tabulky číslo 6 vidíme, že při takovémto ohodnocení stránek je maximální posun týmu o 4 příčky oproti ligovému pořadí. Toto hodnocení tedy opravdu více odpovídá skutečnosti.

Na konec kapitoly ještě můžeme zmínit fakt, že takto by šly porovnat týmy z celé Evropy. Důležité spojení nám zajistí například Liga mistrů nebo Evropská liga. Podobně jako v jarní části Synot ligy, kde spolu všechny týmy nehrály a stejně jsme je dokázali ohodnotit, bychom mohli porovnat evropské týmy.

5 Dodatečná pozorování

Dříve už jsme si všimli, že pokud se v matici $\boldsymbol{H}$ objevil nějaký nulový sloupec, to odpovídá tomu, že z nějaké stránky nevede žádný odkaz, tak vektor popularity konvergoval k nulovému vektoru. Takovému jevu jsme říkali ztrácení důležitosti. Naopak, pokud byla matice $\boldsymbol{H}$ stochastická, tj. z každé stránky vedl alespoň jeden odkaz, tak se žádná důležitost neztrácela (viz lemma 3.11). V této sekci se budeme zabývat obdobnými jevy a jejich dopadem na matice $\boldsymbol{H}$, $\boldsymbol{S}$ a také $\boldsymbol{G}$.

Na vrcholech (stránkách) zavedeme relaci ekvivalence následovně. Vrchol S_1 je v relaci s vrcholem S_2, pokud z S_1 vede orientovaná cesta do S_2. Není těžké rozmyslet si, že jde skutečně o ekvivalenci. Tím se nám graf rozpadne na třídy ekvivalence, mezi kterými vedou hrany už jen jedním směrem. Na těchto třídách můžeme opět zavést relaci podobnou té výše popsané, a dostaneme tak částečné uspořádání (stránky, do kterých nevede žádný odkaz, ani z nich žádný nevychází, nebudeme brát v úvahu). Maximální prvky budou ty třídy, do kterých nevedou z jiných tříd žádné odkazy, a naopak minimální prvky budou ty třídy, z kterých už nevedou žádné odkazy. Pokud bude v takovéto třídě více vrcholů, budeme mluvit o koncové třídě a pokud tam bude pouze jeden vrchol, budeme mluvit o odtoku.

Podobná situace je například na prohlídce zámku. Začínáme u vchodu (maximální prvek) a jakmile projdeme do první místnosti, tak už se zpátky vrátit nemůžeme. Zde se můžeme dle libosti kochat obrazy, ale jakmile se dostaneme do druhé místnosti, tak se opět už vrátit nemůžeme a tak dále. Nakonec se ocitneme u východu, což je minimální prvek. Dle prohlídkových okruhů máme i na výběr cestu, kudy půjdeme a také se začíná a končí jinde (více maximálních a minimálních prvků).

Nejprve se podívejme na to, jaký vliv mají odtoky a koncové třídy. To se týká matic $\boldsymbol{H}$ a $\boldsymbol{S}$.

Vzpomeňme si na náhodné klikání na odkazy na stránce. Pokud se kliknutím na odkaz dostaneme na stránku v jiné třídě ekvivalence, tak už nemáme šanci se do předchozí třídy znovu dostat. Kdyby ano, byly by tyto stránky ve stejné třídě. Takto klikáme dál, až se nakonec jednou ocitneme v koncové třídě, nebo odtoku.

Pokud se ocitneme v koncové třídě a budeme zde klikat na odkazy, do jiné třídy už se nedostaneme. Z toho důvodu budou mít ve výsledku stránky z koncových tříd všechnu popularitu a stránky, které nejsou v koncových třídách, nebudou mít žádnou.

Pokud se ocitneme v odtoku, tak už nemáme kam dál jít a důležitost by se zde hromadila podobně jako v koncové třídě. Matice $\boldsymbol{H}$ je ovšem konstruována tak, že se tato důležitost dokonce ztrácí (odtéká). Celkem bychom tedy z vektoru popularity ztratili tolik důležitosti, kolik by jí měla stránka, která je odtokem.

Ještě uvedeme dvě pozorování o tom, jak odtoky a koncové třídy ovlivňují vlastní čísla matice $\boldsymbol{H}$.

Pozorování 5.1. Buď k počet odtoků v síti. Pak geometrická násobnost vlastního čísla 0 matice $\boldsymbol{H}$ je nejméně k.

Buď l počet stránek, na které žádná stránka neodkazuje. Pak geometrická násobnost vlastního čísla 0 matice $\boldsymbol{H}$ je nejméně l.

Důkaz: Pokud $k \geq 1$, tak je matice $\boldsymbol{H}$ singulární, a tedy 0 je vlastní číslo. Nyní spočtěme dimenzi nulového prostoru $\boldsymbol{H}$. Víme, že $\boldsymbol{H}$ obsahuje k nulových sloupců a z toho hned dostáváme $dim(Ker(\boldsymbol{H})) \geq k$. Dokázali jsme tedy, že geometrická násobnost vlastního čísla 0 je větší nebo rovna počtu odtoků.

Pro stránky, na které žádná stránka neodkazuje, máme analogický argument. Všimneme si, že pokud taková stránka existuje (tj. $l \geq 1$), tak je $\boldsymbol{H}$ singulární, protože obsahuje nulový řádek. Nulových řádků tato matice obsahuje přesně l. Pak použijeme tvrzení, že $dim(Ker(\boldsymbol{H})) = dim(Ker(\boldsymbol{H}^T))$ a jsme hotovi. $\square$

Pozorování 5.2. Nechť $\boldsymbol{H}$ je matice určená grafem stránek $s_1, s_2, \ldots, s_n$ a nechť l je počet koncových tříd tohoto grafu. Pak 1 je vlastní číslo matice $\boldsymbol{H}$ a geometrická násobnost vlastního čísla 1 je alespoň l.

Důkaz: O stochastických maticích víme, že mají vlastní číslo 1. Matice $\boldsymbol{H}$ ale díky odtokům stochastická být nemusí, a tak budeme muset ukázat, že navzdory přítomnosti odtoků má vlastní číslo 1.

Označme $s_{i_1}, s_{i_2}, \ldots, s_{i_k}$ stránky jedné z koncových tříd grafu. Budeme chtít dokázat, že $det(\boldsymbol{H} - 1\boldsymbol{I})$ je roven nule.

Víme, že to je právě tehdy, když je matice $(\boldsymbol{H} - 1\boldsymbol{I})$ singulární a to je právě tehdy, když má lineárně závislé sloupce. Vezmeme si sloupce $i_1, i_2, \ldots, i_k$, které odpovídají stránkám $s_{i_1}, s_{i_2}, \ldots, s_{i_k}$ a ukážeme, že už tyto jsou lineárně závislé.

$$\begin{array}{cc}
 & \begin{matrix} \cdots & s_{i_1} & \cdots & s_{i_2} & \cdots & s_{i_3} & \cdots & \cdots & s_{i_k} & \cdots \end{matrix} \\
\begin{matrix} \\ \vdots \\ \\ s_{i_1} \\ \\ \vdots \\ \\ s_{i_2} \\ \\ \vdots \\ \\ s_{i_3} \\ \\ \vdots \\ \\ s_{i_k} \\ \\ \vdots \\ \end{matrix} &
\begin{pmatrix}
0 & 0 & 0 & 0 \\
\vdots & \vdots & \vdots & \vdots \\
0 & 0 & 0 & 0 \\
-1 & x & x & x \\
0 & 0 & 0 & 0 \\
\vdots & \vdots & \vdots & \vdots \\
0 & 0 & 0 & 0 \\
x & -1 & x & x \\
0 & 0 & 0 & 0 \\
\vdots & \vdots & \vdots & \vdots \\
0 & 0 & 0 & 0 \\
x & x & -1 & x \\
0 & 0 & 0 & 0 \\
\vdots & \vdots & \vdots & \vdots \\
0 & 0 & 0 & 0 \\
x & x & -1 & x \\
0 & 0 & 0 & 0 \\
\vdots & \vdots & \vdots & \vdots \\
0 & 0 & 0 & 0
\end{pmatrix}
\end{array}$$

Všimněme si, že i_j-tý sloupec má nenulové prvky maximálně na pozicích $i_1, i_2, \ldots, i_k$, pro $j = 1, 2, \ldots, k$, protože s_{i_j} je v koncové třídě. Na ostatních pozicích má tedy nuly, a tak lineární závislost sloupců $i_1, i_2, \ldots, i_k$ na těchto $n - k$ pozicích nezáleží.

Označme nyní t_j sloupec, který vznikne ze sloupce i_j vynecháním oněch $n - k$ nulových pozic, pro $j = 1, 2, \ldots, k$. Pak sloupce $i_1, i_2, \ldots, i_k$ jsou lineárně závislé právě, když jsou lineárně závislé sloupce $t_1, t_2, \ldots, t_k$. To nastane právě tehdy, když je matice $(t_1|t_2|\ldots|t_k)$ singulární. Jelikož víme, že součet každého sloupce této matice je roven nule, tak tato matice singulární je.

Dokázali jsme, že 1 je vlastní číslo matice $\boldsymbol{H}$ a zbývá nám ukázat, že geometrická násobnost 1 je alespoň l. K tomu nám stačí definice geometrické násobnosti jako dimenze nulového prostoru, neboli jádra, matice. Pro každou koncovou třídu máme množinu lineárně závislých vektorů, koncových tříd je l, a tedy $dim(Ker(\boldsymbol{H} - 1\boldsymbol{I})) \geq l$. $\square$

Příklad: Nyní si ukážeme význam dalších vlastních vektorů (tedy ne jenom Perronova) v mocnící metodě při získávání právě Perronova vektoru. Budeme pracovat s maticí $\boldsymbol{G}$, která je uvedená na konci kapitoly 2, a také vycházet z příkladu, který je na konci kapitoly 3.

U matice řádu 10 je jednoduché spočítat vlastní čísla a vlastní vektory, a proto zde můžeme rovnou uvést Perronův vektor matice $\boldsymbol{G}$, který kvůli přehlednosti zaokrouhlíme na tři desetinná místa.

$$\boldsymbol{\pi} = \begin{pmatrix} 0,033 & 0,075 & 0,058 & 0,050 & 0,108 & 0,106 & 0,160 & 0,192 & 0,183 & 0,033 \end{pmatrix}$$

Počáteční vektor $\boldsymbol{p}(0)$ zvolíme tak, jak jsme navrhovali v kapitole 2, tedy

$$\boldsymbol{p}(0) = \begin{pmatrix} 0,1 & 0,1 & 0,1 & 0,1 & 0,1 & 0,1 & 0,1 & 0,1 & 0,1 & 0,1 \end{pmatrix}.$$

Připomeňme nyní, že počáteční vektor vyjádříme vzhledem k bázi složené z vlastních vektorů, tedy

$$\boldsymbol{p}(0) = a_1\boldsymbol{\pi} + a_2\boldsymbol{v}_2 + \ldots + a_{10}\boldsymbol{v}_{10} \; .$$

Podíváme-li se na první složku tohoto vektoru (odpovídá důležitosti první stránky), vidíme, že oproti vektoru $\boldsymbol{\pi}$ musíme první stránce přidat $0,067$ důležitosti. Na druhou stranu u deváté stránky musíme důležitost ubrat, konkrétně $0,083$.

Z toho plyne, že vektory $\boldsymbol{v}_2, \boldsymbol{v}_3, \ldots \boldsymbol{v}_{10}$ musí vyrovnávat rozdíly mezi libovolně zvoleným počátečním vektorem a Perronovým vektorem.

Podobných tvrzeníček by šla zformulovat a dokázat jistě celá řada. My se ale kvůli rozsahu práce spokojíme s výše uvedeným s tím, že pokud by čtenáře zajímala další fakta, může se inspirovat v uvedené literatuře.

Seznam použité literatury

BARTO, L. a TŮMA, J. (2014). Zápisky z přednášek lineární algebry. URL http://www.karlin.mff.cuni.cz/~tuma/LinAlg14-15/skripta_la4.pdf. [Online; přístupný 21.04.2015].

LANGEVILLE, A. N. a MEYER, C. D. (2006). *Google's PageRank and Beyond: The Science of Search Engine Rankings*. Princeton Universitiy Press, New Jersey. ISBN 0-691-12202-4.

LFA, 2SCORE S.R.O. a ESPORTS.CZ S.R.O. (2015). SYNOT liga. URL http://www.synotliga.cz/. [Online; přístupný 18.03.2015].

MEYER, C. D. (2000). Matrix Analysis and Applied Linear Algebra: SolutionsManual. URL http://www.matrixanalysis.com/SolutionsManual.pdf. [Online; přístupný 05.05.2015].

MEYER, C. D. (2001). *Matrix Analysis and Applied Linear Algebra*. SIAM: Society for Industrial and Applied Mathematics, Philadelphia. ISBN 0898714540.

PELANTOVÁ, E. (2010). TEMA KapitolaCtvercoveMatice_Final. URL http://people.fjfi.cvut.cz/pelanedi/TEMA/KapitolaCtvercoveMatice_Final.pdf. [Online; přístupný 22.04.2015].

STERNBERG, S. (2010). *Dynamical Systems*. Dover Books on Mathematics. Dover Publications, New York. ISBN 0486477053.