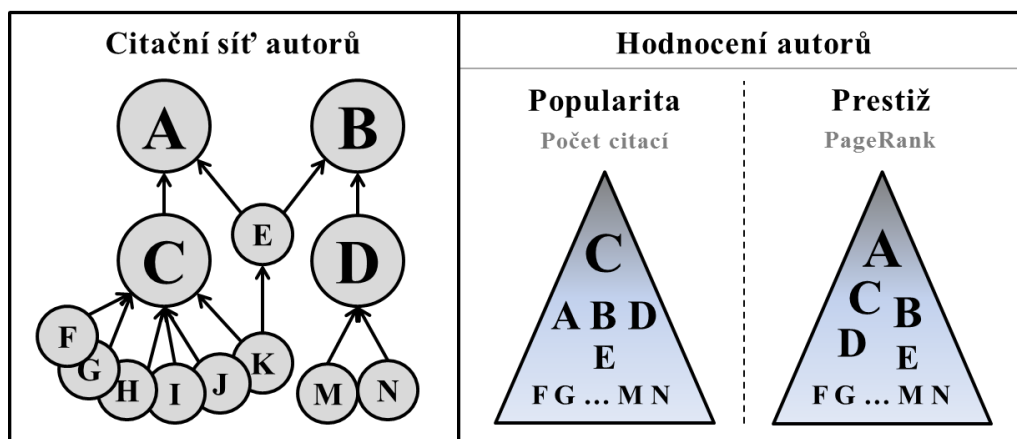


Hodnocení autorů obsažených v kolekci Web of Science

Michal Nykl¹

1 Hodnocení autorů vědeckých publikací

Pořadí autorů vědeckých publikací, které je vytvořeno dle jejich hodnot vědecké kvality, může být použito při vyhledávání nebo porovnávání expertů v určené oblasti, při výběrových řízeních, udílení odměn nebo ocenění apod. Dále zmíníme, jaké metody pro hodnocení autorů jsme použili v článku (Nykl et al., 2015). Hlavním algoritmem, který pro hodnocení autorů používáme, je iterační algoritmus PageRank (Brin a Page, 1998), který využívá významnosti citujících vrcholů v citačním grafu (web-stránky, publikace, autoři apod.) pro určení významnosti vrcholů citovaných. Jak zmiňuje Ding (2011), díky svým vlastnostem může být PageRank použit jako dobrý *ukazatel prestiže* autorů – prestižní autor je citován hodně citovanými autory. Často používaný počet citací ona označuje jako dobrý *ukazatel popularity* – populární autor je hodně citován normálními autory. Ideu odlišení popularity a prestiže znázorňuje obrázek 1, na kterém autora A můžeme označit za prestižního a autora C za populárního. Autor A je také významnější než autor B.



Obrázek 1: Rozdíl mezi popularitou (počet citací) a prestiží (PageRank).

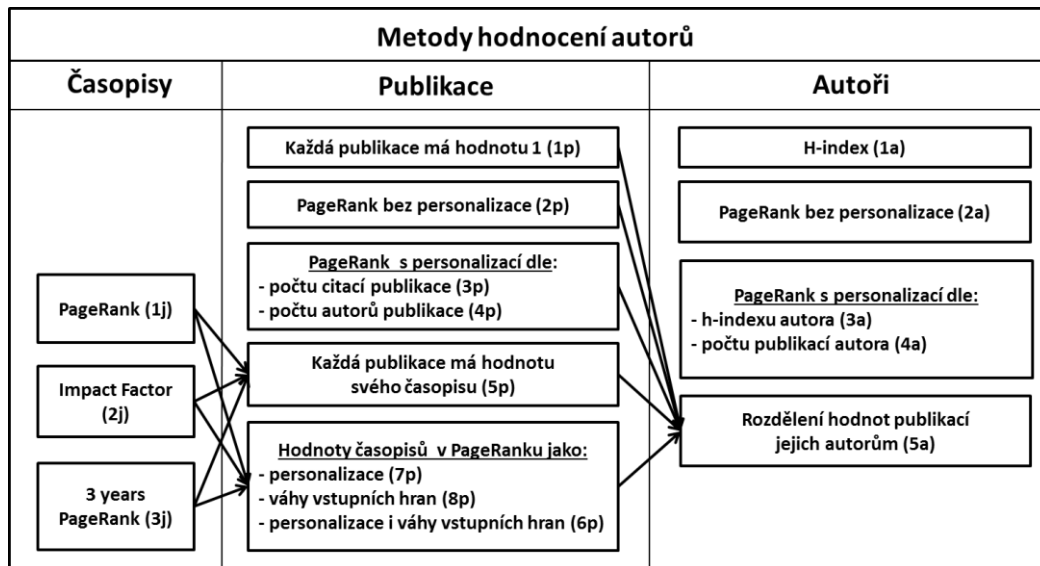
2 Použitá datová kolekce a referenční seznamy oceněných autorů

Pro testování našich metod jsme použili zakoupenou bibliografickou kolekci *Web of Science*. Ta obsahuje všechny časopisecké články z let 1996-2005, které byly v *Journal Citation Reports 2009* obsažené v kategoriích počítačových věd. Abychom porovnali kvalitu našich metod, vytvořili jsme referenční seznamy významných autorů a vypočtená pořadí autorů porovnávali na základě průměrné pozice, kterou tito autoři obsadili. Vypočtená pořadí obsahující autory z celé oblasti počítačových věd (tzv. globální hodnocení) jsme porovnali na základě osob *Association for Computing Machinery (ACM) Fellows*. Pořadí obsahující pouze autory ze zvolených oblastí *Umělá inteligence* a *Hardware* jsme porovnali na základě držitelů ocenění, která udělují *ACM Special Interest Groups* v odpovídajících oblastech.

¹ student doktorského studijního programu Inženýrská informatika, obor Informatika a výpočetní technika, specializace Modely a metody extrakce informací z Webu, e-mail: nyklm@kiv.zcu.cz

3 Naše metody pro hodnocení autorů

Naše metody pro hodnocení autorů shrnuje obrázek 2, na kterém obdélníky zastupují postupy či algoritmy hodnocení a šipky vyjadřují, jakým způsobem jsou vypočtené hodnoty využity při hodnocení autorů.



Obrázek 2: Naše metody pro hodnocení autorů obsažených ve Web of Science.

4 Shrnutí poznatků

Nejlepší naší metodou byla metoda (5a7p1j), která využívá PageRank časopisů jako personalizaci (statická část) PageRanku v citačním grafu publikací a hodnoty publikací následně rozděljuje jejich autorům. Tato metoda byla nejlepší jak pro globální hodnocení, tak pro hodnocení v kategoriích, proto můžeme konstatovat, že nedošlo k přetrénování metod. Hodnocení autorů na základě *prestiže* (PageRank) bylo lepší, než jejich hodnocení na základě *popularity* (počet citací), *vyspělosti* (h-index) nebo *produktivity* (počet publikací). Také bylo lepší vyhodnocovat citační graf publikací než graf autorů a hodnoty publikací rozdělovat autorům uniformně (každý autor získá z publikace stejnou hodnotu) než jiným způsobem. Personalizace publikací dle hodnot časopisů byla také lepší než jiné použité personalizace.

Výzkum, který jsme zde shrnuli, navazuje na článek (Nykl et al., 2014) a jeho výsledky budou publikovány v (Nykl et al., 2015). Oba články byly podpořeny grantem ZČU SGS-2013-029 „Pokročilé výpočetní a informační systémy“.

Literatura

- Brin, S., and Page, L. (1998). The anatomy of a large-scale hypertextual Web search engine. *Computer Networks and ISDN Systems*, Vol. 30. pp 107–117.
- Ding, Y. (2011). Applying weighted PageRank to author citation networks. *Journal of the American Society for Information Science and Technology*, Vol. 62. pp 236–245.
- Nykl, M., Ježek, K., Fiala, D., and Dostal, M. (2014). PageRank variants in the evaluation of citation networks. *Journal of Informetrics*, Vol. 8. doi:10.1016/j.joi.2014.06.005
- Nykl, M., Campr, M., and Ježek, K. (2015). Authors ranking based on personalized PageRank. *Journal of Informetrics*. Článek je ve schvalovacím řízení.