

Západočeská univerzita v Plzni
Fakulta aplikovaných věd
Katedra informatiky a výpočetní techniky

Bakalářská práce

Měření významnosti autorů v citační síti

Prohlášení

Prohlašuji, že jsem bakalářskou práci vypracoval samostatně a výhradně s použitím citovaných pramenů.

Tomáš Maršálek

Abstract

Obsah

1	Úvod	1
2	Sociální a citační sítě	3
2.1	Reprezentace sítí	3
2.2	Citační sítě	4
2.2.1	Citační síť publikací	4
2.2.2	Citační síť autorů	5
2.2.3	Vážené citační sítě	5
2.2.4	Orientované a neorientované sítě	6
3	Citační databáze	7
3.1	DBLP	7
3.2	CiteSeer	9
4	Analýza sítí	11
4.1	Souvislost a komponenty grafu	11
4.2	Klika v grafu	12
4.3	Významnost uzlů	12
4.3.1	Ocenění významných autorů	12
4.4	Míry centrality	13
4.4.1	Degree	14
4.4.2	Eigenvector centrality	14
4.4.3	Closeness centrality	16
4.4.4	Betweenness centrality	19
4.5	Hledání nejkratších cest	22
4.5.1	Single source shortest path	22
4.5.2	All-pair shortest paths	24
4.6	Ostatní míry významnosti autorů	25
4.6.1	H-index	25
5	Výsledky	27
5.1	Porovnání implementovaných metod	27
5.1.1	Spearmanův koeficient pořadové korelace	27
5.2	Žebříčky významných autorů	28
5.3	Porovnání metod s oceněními	30
5.4	Aproximace betweenness centrality	31
6	Diskuse	32
6.1	Podobnost výsledků jednotlivých metod	32

6.2	Vliv vah na výsledky	33
6.3	Vstupní a výstupní hrany	33
6.4	Shoda výsledků s oceněními	34
6.5	Aproximace betweenness	34
7	Závěr	35
A	Tabulky korelací	39
B	Porovnání pořadí implementovaných metod	43
B.1	DBLP	43
B.2	CiteSeer	45
C	Žebříčky významných autorů	47
C.1	DBLP	48
C.2	CiteSeer	62

1 Úvod

V době, kdy je celý svět propojen internetem a sociálními sítěmi, si podvědomě začínáme uvědomovat provázanost celého světa a všímáme si struktur kolem nás s charakterem sítě. Úspěchem je rozpoznat síťovou strukturu a popsat ji, ale můžeme jít s myšlenkou dál? Jaké informace zjistíme, pokud budeme blíže zkoumat uzly a jejich spojení? V této práci nás bude zajímat jedna konkrétní kvalitativní vlastnost uzlů - významnost.

Významnost není jednoznačně definována, ale intuitivně cítíme, pokud je například člověk významný mezi svými přáteli nebo můžeme zjistit, které město je klíčovým dopravním uzlem v železniční síti. Představme si, že neznáme jednotlivé lidi v síti přátel a žádné informace o nich. Jak zjistíme, kdo je nejvýznamnější pouze na základě vztahů mezi nimi? Naším cílem je zjistit nejen, které prvky obecné sítě jsou významné a které nikoliv, ale pokusíme se najít metody, jak seřadit prvky od nejvýznamnějšího po nejméně významný.

Velký díl k zodpovězení otázky relativní významnosti prvků definovali Freeman [17]; Bonacich [3], jehož práce je spojena s Hubbellovo [20] mírou sociometrického statusu; Coleman [7] se svou mírou síly a Burt [6] a jeho míra prestiže [18]. Významnost prvku bývá v sociální síti označována jako centralita a metody pro zjištění centrality jsou známé jako míry centrality (centrality measure). Původně byly vyvinuty v sociologickém kontextu pro analýzu sociálních sítí, ale jejich princip lze snadno zobecnit na obecný graf, proto můžeme využít těchto metod pro analýzu citačních nebo jiných komplexních sítí, které nemají čistě sociologický význam.

Tato práce je věnována citačním sítím, ale používáme metody z analýzy sociálních, dopravních, komunikačních a jiných sítí. V citační síti hledáme nejvýznamnější autory pouze podle toho, jak jsou provázáni s jinými autory podle referencí v publikacích, které vydali. Významnost autorů není úplnou neznámou, protože existuje množství ocenění, která byla udělena právě významným autorům a vědcům za jejich dílo. Samotné udělení ocenění mohlo tyto autory udělat významnými, přestože předtím nebyli. V jiném případě mohlo být důležité ocenění příčinou ještě větší významnosti autora.

Budou přiblíženy detaily o nejpoužívanějších mírách centrality (degree, eigenvector centrality, betweenness centrality, closeness centrality), bibliografické metodě h-index a jejich implementaci. Výsledkem budou porovnání jednotlivých metod aplikovaných na citační síť vytvořené z citačních databází DBLP a CiteSeer, které zároveň srovnáme s oceněními autorů v oblasti informatiky (ACM SIGMOD Edgar F. Codd Innovations Award, ACM Fellows, ACM A.M. Turing Award, ISI Highly Cited highlighted).

Cílem této práce je vytvořit knihovnu pro analýzu citační sítě s metodami pro měření významnosti autorů, použít ji pro nalezení žebříčků nejvýznamnějších autorů pro každou metodu. Tyto žebříčky poté mezi sebou porovnáme a jednotlivé metody

srovnáme s již udělenými oceněními. Očekáváme, že pokud měříme významnost autora nebo obecně prvku v síti, přestože neznáme její přesnou definici, bude se shodovat s těmito oceněními.

2 Sociální a citační sítě

Myšlenka sociální sítě existovala dlouho předtím, než je pod tímto termínem začali lidé rozpoznávat. Jedná se o komplexní struktury vztahů mezi členy sociálních uspořádání na všech úrovních - od osobních až po mezinárodní vztahy mezi organizacemi.

Nejčastěji se ale setkáme se sociální sítí jako strukturou tvořenou lidmi, kteří jsou svázáni nějakým sociálním vztahem, zejména v poslední době s rozmachem populárních webových sociálních sítí (MySpace, Facebook, G+, Lidé), jím bývá přátelství.

2.1 Reprezentace sítí

Abychom mohli pracovat s doposud abstraktním konceptem sítě, musíme být schopni ji reprezentovat jako datovou strukturu, na níž poté provedeme jakoukoliv analýzu. V odvětví matematiky teorie grafů je síť (graf) dvojice množin uzlů V (vrcholů) a spojení uzlů E (hran) $G = (V, E)$. Obecně můžeme uvažovat grafy s hranami s orientací či bez orientace. V obou případech se stále jedná o dvojici (V, E) , pouze pro orientovaný graf je množina hran množinou uspořádaných dvojic oproti množině neuspořádaných dvojic u neorientovaného grafu.

V definici grafu je množina hran E soubor dvojic, které označují koncové uzly hrany, neboli jejich spojení. Samotné spojení je jediná informace, kterou množina hran nese. Chceme-li zaznamenat nějakou další informaci, která je spojena se spojením dvou uzlů, namísto hrany jako dvojice koncových uzlů, nadefinujeme hranu jako n -tici, kde první dvě hodnoty jsou koncové uzly a zbylé hodnoty nesou libovolnou informaci. Ve většině případů si vystačíme s jednou dodatečnou informací a nazýváme ji váha hrany. Jiná možnost pro zavedení vah hran je váhová funkce $f : E \mapsto \mathbb{R}$, kde $f(e) = w$ je ohodnocení konkrétní hrany $e \in E$. V případě zavedení vah hovoříme o vážených sítích.

Při zavedení vah máme například možnost používat síť jako multigraf, tedy graf, u kterého je povoleno více spojení mezi dvěma stejnými uzly. Počet stejných hran pak pouze zaznamenáme celočíselnou hodnotou ve váze hrany.

Síť world wide web tvořená webovými stránkami je příkladem multigrafu, protože je povoleno z jedné stránky odkazovat na jinou na více místech. Při analýze takovýchto sítí využijeme právě vah hran a počet hypertextových odkazů mezi dvěma stránkami zaznamenáme vyšším ohodnocením hrany. V tomhle případě znamená vyšší váha silnější pouto mezi uzly.

Jiným případem může být např. síť, kde sledujeme města a dopravní spojení mezi nimi. V tomto případě nás může zajímat vzdálenost nebo časová náročnost na dopravu mezi dvěma městy, které budou znamenat silnější pouto, pokud budou mít naopak menší váhu. Hledáme totiž nejkratší či nejrychlejší spojení.

Pro reprezentaci v paměti počítače se nejčastěji používají dva způsoby - matice

sousednosti a graf pomocí spojových seznamů. Hrany se uzlu v případě orientovaného grafu liší z pohledu jednoho uzlu. Pokud hrana vychází z tohoto uzlu, nazveme ji výstupní hrana, v opačném případě se bude jednat o vstupní hranu.

Matice sousednosti (adjacency matrix) je čtvercová matice A o velikosti počtu vrcholů grafu $|V|$, ve které prvek A_{uv} na řádku u a sloupci v určuje jestli existuje hrana od vrcholu u do vrcholu v . Pokud je hodnota A_{uv} 1, hrana existuje; pokud je hodnota 0, pak hrana neexistuje a pokud je hodnota w , pak hrana existuje s váhou w .

Jiným maticovým způsobem uchování grafu je incidenční matice B . Incidenční matice vyjadřuje vztah mezi vrcholy a hranami tak, že $B_{ue} = 1$, pokud vrchol u je spojený s hranou e , a 0 v opačném případě. V orientovaném grafu rozlišujeme mezi počátečním uzlem $B_{ue} = -1$ a koncovým uzlem $B_{ue} = 1$. Incidenční matice se pro výpočetní teorii grafů často nepoužívá z důvodu paměťové náročnosti, která je pro většinu grafů výrazně vyšší než u matice sousednosti ($\Theta(|V||E|)$ oproti $\Theta(|V|^2)$, kde množina hran dosahuje velikostí $O(|V|^2)$).

Nejčastěji používáme myšlenku sousednosti vrcholů, ale namísto reprezentace maticí, která je ve většině případů řídká a zbytečně obsahuje velké množství nul, použijeme reprezentaci řídké matice - řádek nahradíme seznamem vrcholů, které v matici sousednosti mají nenulovou hodnotu. Tento způsob je známý jako graf pomocí spojových seznamů (adjacency list representation of a graph).

2.2 Citační sítě

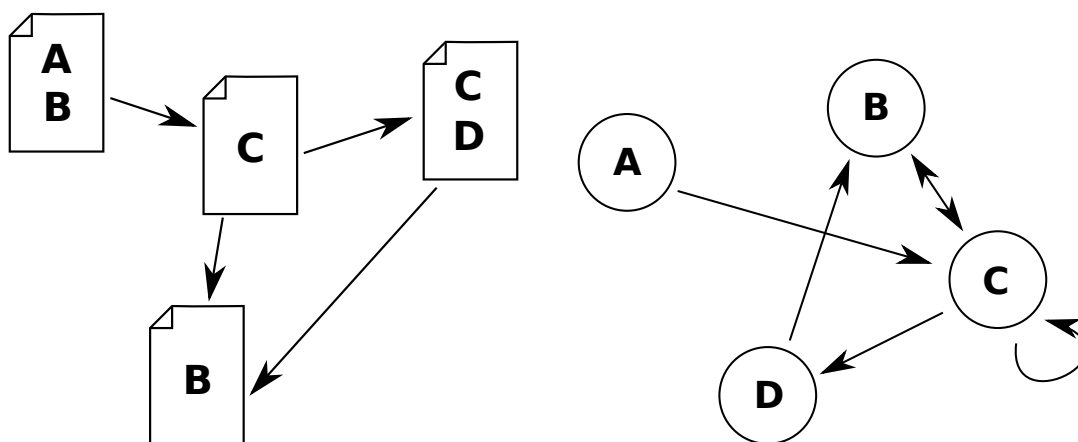
Citační sítě jsou podobné sociálním sítím, pouze místo uzlů, které představují osoby, se v citační síti jedná o publikace nebo autory těchto publikací. Pokud je uzlem publikace, pak hrany této sítě symbolizují citaci publikace jinou publikací. V druhém případě uvažujeme síť, kde uzly reprezentují autory knih, vědeckých článků, vědecké literatury a dalších publikací. Prvnímu typu říkáme síť publikací, druhému síť autorů. Citační sítě patří do kategorie bezškálových sítí. Citační síť autorů pro databázi DBLP i CiteSeer mají mocninné rozložení stupňů i vah hran (obr. 3.1, 3.3).

2.2.1 Citační síť publikací

Uvažujeme-li první případ, kde uzly reprezentují publikace a hrany přímo citace mezi těmito publikacemi, jedná se o síť publikací. Pokud publikace A odkazuje na publikaci B , budou existovat stejnojmenné uzly A a B a hrana mezi těmito uzly může mít dvě různé orientace podle svého uplatnění. Směr od citující publikace k citované (v našem příkladě od A do B) bude mít hrana, kterou označíme jako výstupní pro uzel A a vstupní pro uzel B . Výstupní hrana, laicky řečeno, označuje vztah „cituji“, kdežto vstupní hrana znamená „jsem citován“.

2.2.2 Citační síť autorů

Druhou citační sítí je citační síť autorů. Zde je uzel reprezentací autora a hrany spojují autory mezi sebou. Ve většině případech máme k dispozici data ve formátu, který přímo odpovídá síti publikací, tj. pro každou publikaci známe seznam jejích autorů a odkazů na další publikace. Síť autorů lze získat transformací sítě publikací tak, že každou hranu z původní sítě publikací přiřadíme každému z autorů citující publikace a duplikujeme ji pro každého z autorů citované publikace (obr. 2.1). Celkově vznikne nm nových hran, pokud odkazovaná publikace obsahuje n autorů a odkazující m autorů. Stejně jako v síti publikací, i zde uvažujeme opačné orientace hrany.



Obrázek 2.1: Transformace z citační sítě publikací na citační síť autorů

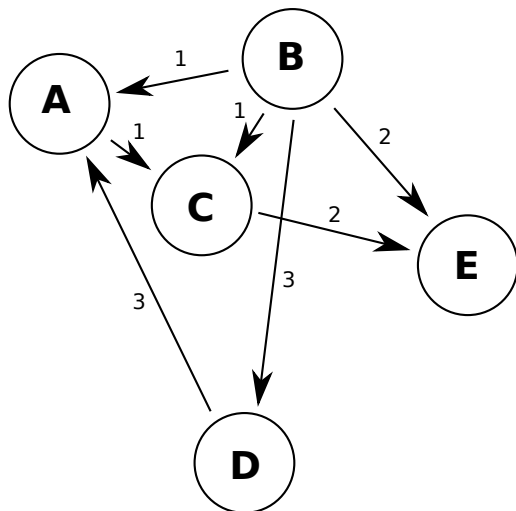
V síti autorů má pro naše účely smysl uvažovat ohodnocení hran. Existuje více způsobů, jak lze přiřadit ohodnocení (váhu) jednotlivým hranám, ale nejjednodušším způsobem je prosté přiřazení počtu publikací, jejichž autorem je daný autor A , které odkazují na publikace, jejichž autorem je autor B .

Druhou možností ohodnocení hran, který využívá implementovaná knihovna pro některé metody, je převrácená hodnota prvního způsobu ohodnocení. Důvodem je přímá souvislost mezi váhou hrany a vzdáleností mezi uzly. V prvním případě (obr. 2.2), kdy silnější pouto mezi autory vyjadřuje vyšší ohodnocení hrany, v druhém případě (obr. 2.3) je naopak nižší váha vyjádřením silnějšího vztahu, jelikož jsou si uzly blíže. Tento způsob je používán pro algoritmy, které pracují na myšlence nejkratších cest mezi uzly.

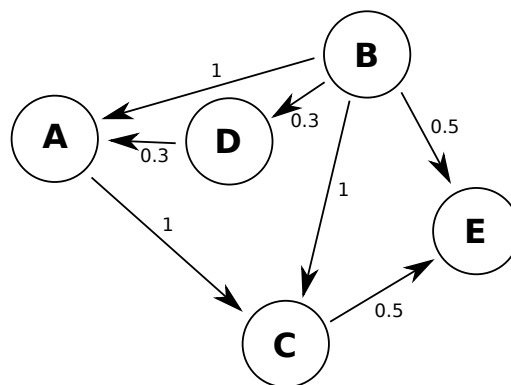
2.2.3 Vážené citační sítě

Pro citační sítě můžeme uvažovat ohodnocení hran obojího typu. Například mezi dvěma autory může být silnější vztah, pokud se citují ve více publikacích. Jestliže citační síť analyzujeme metodami, které jsou založené na myšlence hledání nejkratších

cest i v této síti, která nemá v podstatě žádný pojem vzdálenosti, použijeme druhý typ ohodnocení - menší váha, silnější pouto.



Obrázek 2.2: Vážená síť autorů, kde váha odpovídá vzdálenosti mezi uzly



Obrázek 2.3: Stejná síť autorů s převrácenou hodnotou váhy ($\frac{1}{w}$)

2.2.4 Orientované a neorientované sítě

V případě sociálních sítí nejčastěji uvažujeme sítě bez orientace, protože nejčastěji modelovaný vztah „přítel-přítel“ je ekvivalentní z pohledu obou koncových uzlů. Pro citační síť jsou na místě orientované hrany, protože vztahy autor odkazujícího na jiného autora nebo publikace citující jinou publikaci mají očividně jinou interpretaci z pohledu koncových uzlů. Buďto se jedná o citovaného nebo citujícího autora či publikaci.

3 Citační databáze

Bibliografická citační databáze poskytují možnost vyhledávání bibliografických citací. Velké množství z dnešních citačních databází se zaměřuje na jeden obor. [23]. Jiné jsou multioborové s možností volby prohledávaného oboru (Scopus, Web of Science).

3.1 DBLP

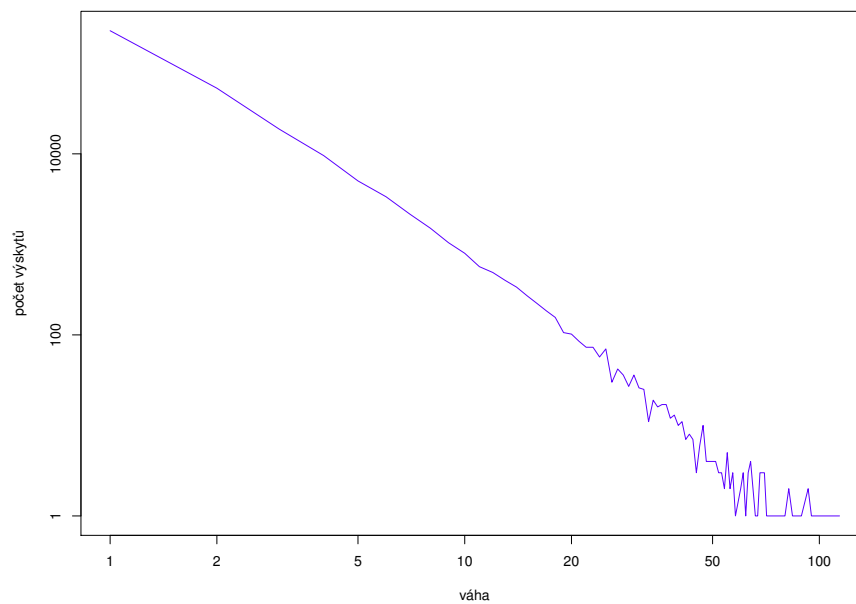
DBLP [35] je webová bibliografická databáze hostovaná na univerzitě Trier. Od 80. let indexovala literaturu z oblasti databázových systémů a logického programování, ale postupně se její zaměření zobecnilo a nyní je bibliografickou databází obecně pro obor informatiky. V roce 2012 obsahovala více než 2,1 milionu článků. Metody implementované v této práci jsou aplikovány na verzi z roku 2004. Tab. 3.1 poskytuje statistiky o vytvořených citačních sítích z této databáze. Obr. 3.1 znázorňuje vlastnosti bezškálové sítě pro váženou citační síť autorů - čím vyšší váha, tím menší pravděpodobnost výskytu v síti. Obr. 3.2 znázorňuje mocinné rozdělení stupňů uzlů pro stejnou síť.

Tabulka 3.1: Statistiky pro databázi DBLP 2004

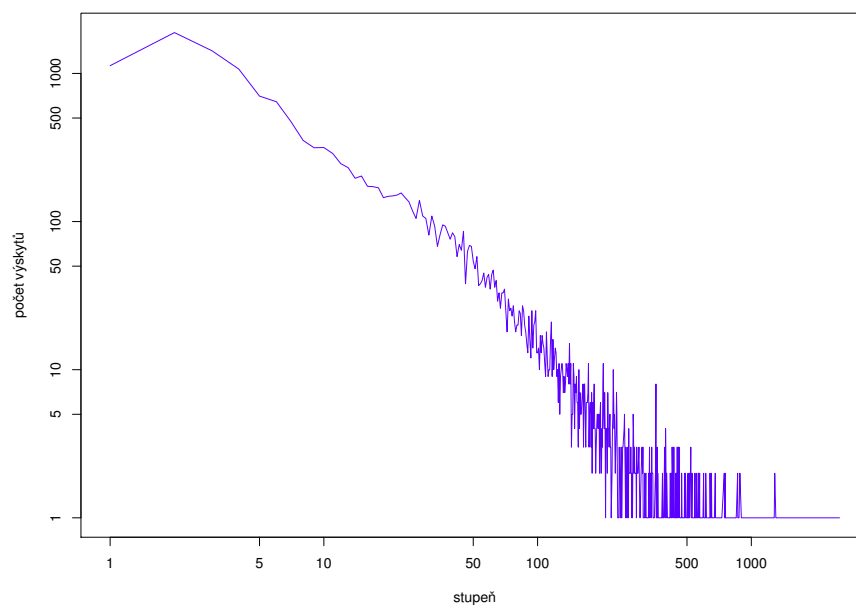
	hodnota
Počet publikací	470 554
Počet hran v síti publikací	109 130
Počet autorů	315 485
Počet hran v síti autorů	331 245
Počet samocitací	3 095
Průměrný počet spoluautorů	2.278
Největší silně spojitá komponenta	1.637%
Průměrná nejkratší cesta ¹	2.894
Průměrná nejkratší cesta ²	1.194
Poloměr grafu ¹	8.0
Poloměr grafu ²	6.637

¹Platí pro neváženou síť autorů

²Platí pro váženou síť autorů



Obrázek 3.1: Log-log graf mocninného rozdělení vah hran citační sítě autorů DBLP



Obrázek 3.2: Log-log graf mocninného rozdělení stupňů uzlů citační sítě autorů DBLP

3.2 CiteSeer

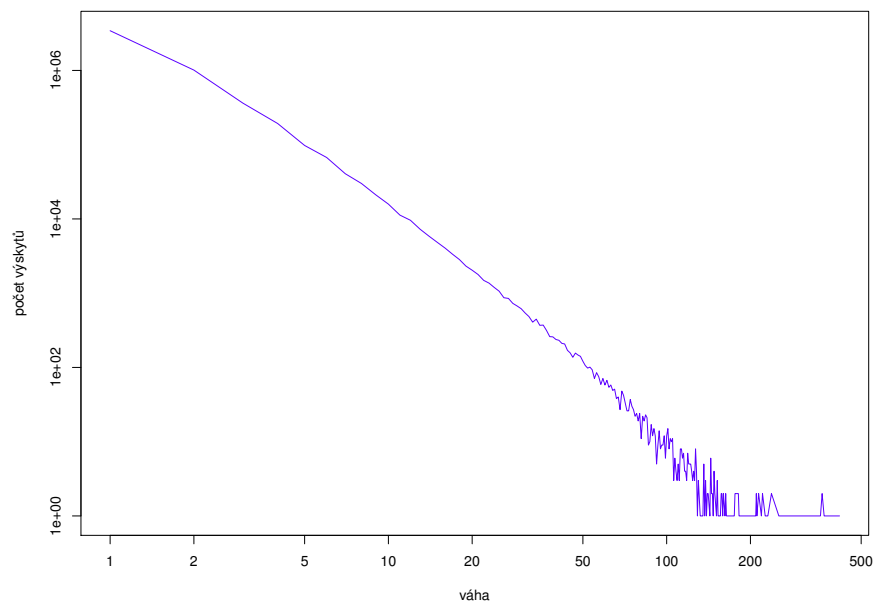
CiteSeer [36] (nyní CiteSeer^X) je považován za první automatizovaný systém shromažďování publikací a autonomní indexace citací v nich obsažených. Publikace jsou zejména z oboru informatiky a informační vědy. V dnešní době obsahuje přes dva miliony dokumentů s téměř dvěma miliony autorů a čtyřiceti miliony citací. Zde používáme verzi z roku 2005. Stejně jako u databáze DBLP jsou v tab. 3.2 uvedeny statistiky pro vytvořené citační sítě a v obr. 3.3 a 3.4 ukázka vlastností bezškálové sítě pro citační síť autorů.

Tabulka 3.2: Statistiky pro databázi CiteSeer 2005

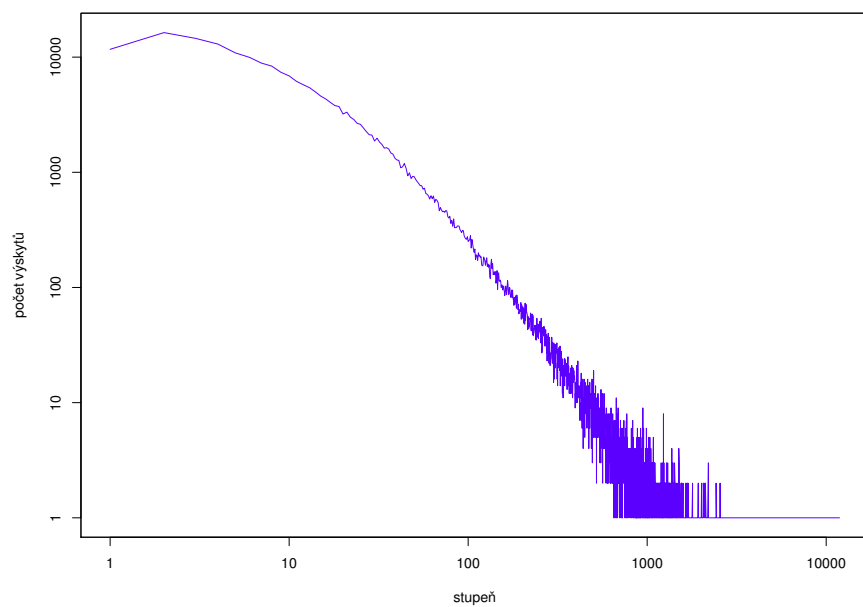
	hodnota
Počet publikací	659 814
Počet hran v síti publikací	1 624 758
Počet autorů	406 465
Počet hran v síti autorů	5 398 239
Počet samocitací	60 639
Průměrný počet spoluautorů	2.525
Největší silně spojená komponenta	29.763%
Průměrná nejkratší cesta ¹	4.134
Průměrná nejkratší cesta ²	1.592
Poloměr grafu ¹	15.0
Poloměr grafu ²	11.489

¹Platí pro neváženou síť autorů

²Platí pro váženou síť autorů



Obrázek 3.3: Log-log graf mocninného rozdělení vah hran citační sítě autorů CiteSeer



Obrázek 3.4: Log-log graf mocninného rozdělení stupňů uzlů citační sítě autorů CiteSeer

4 Analýza sítí

Více než 40 let byly považovány všechny komplexní sítě za naprosto náhodné. Paul Erdős a Alfréd Rényi v roce 1959 navrhli modelování komunikačních sítí a sítí, které se vyskytují v přírodních vědách, spojením uzlů náhodnými hranami. Tato jednoduchá metoda vytvoření náhodného grafu způsobí rozložení stupňů vrcholů (počet spojení vrcholu s ostatními) podle Poissonova rozdělení pravděpodobnosti s charakteristickou křivkou připomínající zvon - většina uzlů má zhruba stejný stupeň. V roce 1998 bylo na univerzitě v Notre Dame (Barabási a kolegové) provedeno mapování sítě World Wide Web s očekáváním, že výsledkem bude náhodná síť. Přestože byl zmapován pouze zlomek celé sítě, výsledkem bylo přes všechna očekávání, zcela jiné rozdělení stupňů - mocninné. Přes 80% uzlů mělo méně než čtyři spojení, ale méně než 0.01% uzlů mělo více než tisíc spojení. Sítě, které se řídí mocninným rozdělením stupňů, nazvali Barabási a jeho kolegové bezškálovými sítěmi (scale-free network). Rozpoznání tohoto jevu vedlo k lepšímu porozumění šíření virů a epidemií nebo proč některé sítě fungují takřka beze změny i přes poruchu většiny jejich uzlů. Sociální a citační sítě se řadí do kategorie bezškálových sítí. Například autor vědecké literatury, jehož dílo je v dané oblasti známé, má velkou šanci, že bude citován dalšími autory, především těmi novými. Stejně tak osoba s velkým počtem přátel má velkou šanci, že bude představen novým lidem a rozšíří si tak svůj okruh přátel ještě více. Tomuto jevu v bezškálových sítích se říká „bohatší se stává bohatším“.

Rozvinutou disciplínou v oblasti sítí je analýza sociálních sítí, která se stala klíčovou technikou v moderní sociologii a stala se významnou v různých vědeckých oblastech (antropologie, biologie, komunikace, ekonomie, informační věda, geografie, historie, politologie, ...).

Analýza sociálních sítí zkoumá povahu vztahů (homofilie, multiplexita, vzájemnost, blízkost vztahů, ...), rozdělení vlastností v síti (míry centrality, hustota sítě, ...) nebo segmentaci (souvislost grafu, komponenty grafu, kliky, koeficient shlukování, ...).

4.1 Souvislost a komponenty grafu

Neorientovaný graf je souvislý, pokud pro každé jeho dva vrcholy u a v existuje alespoň jedna cesta z u do v . U neorientovaného grafu hovoříme o slabé souvislosti. Pro orientovaný uvažujeme silnou souvislost, protože přestože existuje cesta z u do v , není zaručeno, že existuje cesta z v do u . Slabě souvislý orientovaný graf znamená, že neorientovaný graf, který by vznikl nahrazením orientovaných hran neorientovanými (symetrizace grafu), by byl souvislý.

Komponenta maximálně souvislý podgraf. Jinak řečeno komponenta je podgraf takový, že všechny jeho vrcholy jsou spojeny nějakou cestou. Komponentou je i samostatný vrchol.

Všechny slabě souvislé komponenty grafu najdeme pomocí jednoduchých algo-

ritmů „prohledávání do šířky“ nebo „do hloubky“. Spuštění prohledávání najde celou komponentu, ve které se výchozí vrchol nachází. Pokud zaznamenáváme, které vrcholy byly nalezeny, a spustíme prohledávání ze všech nenalezených vrcholů, najdeme všechny komponenty.

Silně spojené komponenty nenajdeme pouhým prohledáním do šířky nebo do hloubky, ale použijeme sofistikovanější algoritmy (Kosarajův, Tarjanův, ...), které principově vycházejí z prohledávání do hloubky.

4.2 Klika v grafu

Klika (clique) grafu je úplný podgraf. To znamená, že všechny vrcholy kliky jsou spojeny přímo hranou.

V sociologii slovo klika popisuje skupinu dvou až dvanácti lidí, kteří jsou na sebe vázáni více než na jiné lidi v tomtéž prostředí [31]. Klika je silněji spojená skupina lidí než sociální kruh.

Algoritmus pro nalezení největší kliky v grafu je přímočaré otestování n vrcholů podgrafu pro všech 2^L podgrafů grafu, kde L je horní limit velikosti podgrafu. Pokud je všech $\frac{n(n-1)}{2}$ párů vrcholů daného podgrafu spojených hranou, pak se jedná o kliku. Problém má exponenciální složitost, proto je horní hranice velikosti podgrafu ve výpočtech omezena na 20.

4.3 Významnost uzlů

Významnost autorů je jedním předmětem zájmu analýzy sociálních sítí. Kdybychom se měli rozhodnout, kterého člena sítě zvolit jako vůdce nebo přes které členy nejrychleji rozšíříme zprávu, koho bychom měli vybrat?

4.3.1 Ocenění významných autorů

Významní autoři vědecké literatury bývají za své dílo oceněni významnou cenou nebo zařazeni do seznamů významných členů.

Autory, kteří byli oceněni těmito cenami, můžeme považovat za významné a častokrát citované už jen proto, že přítomností jejich jména v seznamu oceněných prestižní cenou se dostanou do podvědomí mnoha jiných, zejména začínajících autorů.

ACM A.M. Turing Award

ACM A.M. Turing Award je ocenění ročně udělované skupinou ACM (Association for Computing Machinery) jedincům vybraným pro kontribuce technického ducha do výpočetního světa [13].

Turingova cena je brána jako nejvyšší vyznamenání v informatice a je lidově nazývána Nobelovou cenou pro informatiku [10, p. 317].

ACM SIGMOD Edgar F. Codd Innovations Award

ACM SIGMOD Edgar F. Codd Innovations Award je ohodnocení životního díla skupinou ACM SIGMOD (Special Interest Group on Management of Data) za inovativní a vysoce ceněné kontribuce k rozvoji, porozumění a použití databázových systémů a databází [34].

ACM Fellows

„The ACM Fellows Program“ byl založen v roce 1993, aby našel a ocenil vynikající členy ACM za jejich dílo v informatice a informační vědě a pro jejich významné kontribuce pro účel ACM. Členové ACM Fellows slouží jako význační kolegové, ke kterým ACM a jejich členové vzhlíží jako k autoritám v době rozvoje informačních technologií [12].

ISI Highly Cited highlighted

ISI Highly Cited je databáze často citovaných autorů v článcích posledního desetiletí, které byly publikovány institutem ISI (Institute for Scientific Information). Ten v dnešní době spadá pod agenturu Thomson Reuters, na jejíchž webových stránkách nalezneme seznam autorů ISI Highly Cited highlighted z let 2000 až 2008 napříč 21 vědeckými obory [29].

4.4 Míry centrality

V sítích dopravní infrastruktury nás zajímá, po které cestě se nejrychleji a nejvýhodněji dostat z bodu A do bodu B . V sociálních a citačních sítích nemůžeme intuitivně hovořit o nějakých cestách mezi uzly, protože ani přesně nevíme, jak takovou cestu interpretovat. Nejkratší cesta mezi přáteli v sociální síti může znamenat přes které přátele se mezi nimi nejpravděpodobněji šíří informace. V sítích spolupráce vědeckých autorů se například setkáme s tzv. Erdősovým číslem, které vyjadřuje nejkratší vzdálenost mezi osobou a matematikem Paulem Erdősem v rámci spolupráce na vědeckých článcích v oboru matematiky.

Použijeme-li metody z dopravních sítí pro analýzu sociálních a citačních sítí, které v jádře spočívají v hledání nejkratších cest, setkáme se se dvěma nejznámějšími mírami centrality, a to closeness centrality a betweenness centrality.

Necht' cesta z bodu $u \in V$ do bodu $v \in V$ je střídající se posloupnost vrcholů a hran takových, že spojují předcházející a následující vrchol v této posloupnosti. Délka cesty je pak součet vah hran této cesty nebo pouze počet hran v případě neváženého grafu. Vzdálenost vrcholů $d_G(u, v)$ je délka nejkratší z cest, která spojuje vrcholy u a v .

Jiné míry jsou založeny na počtu spojení jednoho uzlu s ostatními uzly a nejkratší cesty neuvažují (degree, eigenvector).

4.4.1 Degree

Stupeň je počet hran spojených s uzlem. Pro orientovaný graf můžeme uvažovat vstupní (indegree) a výstupní stupeň (outdegree) vrcholu nebo obecný stupeň (degree), tedy součet těchto dvou. Vstupní stupeň se často označuje jako deg^- a výstupní jako deg^+ . Necht' C_D označuje míru centrality degree a C_{Din} , C_{Dout} centralitry indegree a outdegree, respektive. Pak můžeme vyjádřit hodnoty centrality pomocí matice sousednosti \mathbf{A} .

$$C_{Din}(u) = deg^-(u) = \sum_{v \in V} \mathbf{A}_{uv} \quad (4.1)$$

$$C_{Dout}(u) = deg^+(u) = \sum_{v \in V} \mathbf{A}_{vu} \quad (4.2)$$

$$C_D(v) = C_{Din} + C_{Dout} \quad (4.3)$$

Pokud uvažujeme pouze vstupní stupeň, vypočtená hodnota určuje významnost uzlu, kdežto výstupní stupeň ukazuje jakousi společenskost či otevřenost uzlu.

Degree centrality je výpočetně velmi jednoduchý způsob, jak změřit významnost prvku v síti. Tato metoda je však příliš jednoduchá, protože do výpočtu hodnoty centrality nezahrnuje uzly, které jsou od daného uzlu vzdálenější než jeden krok. Tento fakt je známý problém a důvod pro zavedení dalších a složitějších metod pro výpočet významnosti.

4.4.2 Eigenvector centrality

Eigenvector centrality (také známá jako Gould's index of accessibility of a Network [39] nebo Bonacich's centrality [19]) je míra vlivu vrcholu v grafu, která doslova znamená „Důležitý uzel má důležité sousedy“ [40]. Hodnotu vlivu získáme z vlastního vektoru \mathbf{x} matice sousednosti grafu:

$$\mathbf{A}\mathbf{x} = \lambda\mathbf{x}$$

\mathbf{A} je matice sousednosti, \mathbf{x} je vlastní vektor matice \mathbf{A} a řešením této rovnice. Rovnice má více řešení. Ke každému řešení náleží vlastní číslo λ . Pro měření významnosti nás však zajímá pouze to řešení, které má pouze nezáporné hodnoty. Podle Perron-Frobeniovy věty pro každou nezápornou primitivní matici existuje právě jedno takové řešení, které zároveň patří k největšímu vlastnímu číslu λ [22].

Rovnici můžeme rozepsat z maticového tvaru do jednotlivých složek:

$$x_u = \frac{1}{\lambda} \sum_{v \in V} \mathbf{A}_{uv} x_v$$

Kde x_u je prvek vlastního vektoru \mathbf{x} náležící vrcholu u a \mathbf{A}_{uv} je prvek matice sousednosti \mathbf{A} , který leží na řádce u a sloupci v .

$$x_{ui+1} = \frac{1}{\lambda} \sum_{v \in V} \mathbf{A}_{uv} x_{vi}$$

V tomto rekurentním tvaru je vidět předpis pro iterační výpočet eigenvector centrality. Algoritmus se nazývá mocinná metoda, která se používá pro řešení problému vlastních čísel v numerické matematice. Výsledkem mocinné metody je dominantní vlastní číslo a odpovídající vlastní vektor. Pro eigenvector centrality nás zajímá právě tohle řešení a žádné jiné.

Z druhé rovnice si navíc povšimneme, že se jedná o přímé rozšíření degree centrality (4.1). Výsledek předchozí iterace použijeme jako vstup do následující a iterujeme tak dlouho, dokud nedosáhneme požadované přesnosti.

PageRank

V roce 1998 vyvinuli Sergey Brin a Larry Page algoritmus PageRank (nesoucí jméno druhého autora) jako součást výzkumu na novém druhu webového vyhledávače [5, 24, 27]. PageRank přiřazuje relativní hodnocení webovým stránkám podle hypertextových odkazů z jiných webových stránek, které na ně směřují, a podle jejich PageRankové významnosti. Sama definice je rekurzivní a po nahlédnutí na vzorec zjistíme, že se jedná o rozšířenou variantu algoritmu pro eigenvector centrality.

$$x_{ui+1} = \frac{1-d}{|V|} + d \sum_{v \in V} \mathbf{A}_{uv} \frac{x_{vi}}{\deg^+(v)}$$

\mathbf{A} je opět matice sousednosti, V je množina vrcholů a $\deg^+(v)$ je výstupní stupeň vrcholu v . V původní myšlence, kdy PageRank modeluje chování náhodného surfaře, damping factor d je pravděpodobnost, že daný surfař přestane opakovaně klikat na odkazy, které najde na webové stránce, na kterou se dostal z předchozího odkazu, a otevře zcela novou stránku, ze které povede podobný sled surfování přes odkazy. Damping factor bývá ze zkušenosti nastavován na 85%.

Hodnota PageRanku je z matematického hlediska pravděpodobnost, že surfař, který náhodně kliká na odkazy, se dostane na konkrétní stránku. Součet hodnot PageRanku všech uzlů v síti je tedy 1, protože PageRank je rozdělení pravděpodobnosti.

Jedním problémem algoritmu PageRank jsou „visící uzly“ (dangling nodes), tj. uzly bez výstupních hran. Protože musíme v každé iteraci algoritmu zachovat vlastnost rozdělení pravděpodobnosti, tj. suma všech pravděpodobností je 1, je třeba zajistit, aby se přenášená hodnota mezi iteracemi neztrácela právě v uzlech bez výstupních hran. Problém lze řešit tak, že se tyto uzly z výpočtu vynechají, nebo přidáním zpětných odkazů z těchto uzlů zpět do sítě. V každé iteraci předem vypočítáme součet

hodnot PageRanku, které by se ztratily v uzlech bez výstupních hran (D). Tahle hodnota je v rámci iterace konstantní a pouze ji rovnoměrně rozdělíme mezi uzly sítě (s váhou $1/|V|$).

$$x_{ui+1} = \frac{1-d}{|V|} + d \sum_{v \in V} A_{uv} \frac{x_{vi}}{\deg^+(v)} + \frac{1}{|V|} \sum_{w \in D} x_{wi}$$

Algoritmus 1 PageRank

```

1:  $x_0[v] \leftarrow \frac{1}{|V|}, v \in V$  ▷ Uzly začínají se stejnou pravděpodobností
2: for  $i \leftarrow 0, K$  do ▷ Iterujeme  $K$ -krát, dokud není dosažena požadovaná přesnost
3:    $s \leftarrow 0$  ▷  $s$  je suma PageRanků pro dangling nodes
4:   for  $v \in V$  do
5:     if  $\deg^+(v) = 0$  then
6:        $s \leftarrow s + x_i[v]$ 
7:     end if
8:   end for
9:    $x_{i+1}[v] \leftarrow \frac{s}{|V|}, v \in V$  ▷ Každý uzel začíná s přebytkem z dangling nodes
10:  for  $v \in V$  do
11:    for  $k \in \text{inLinks}[v]$  do
12:       $x_{i+1}[v] \leftarrow x_{i+1} + \frac{x_i[k]}{\deg^+(k)}$ 
13:    end for
14:  end for
15:   $x_{i+1}[v] \leftarrow \frac{1-d}{|V|} + dx_{i+1}[v], v \in V$  ▷ Přidání damping factoru
16: end for

```

Přestože je PageRank původně určený pro webovou síť, lze ho použít na jakoukoliv orientovanou váženou i neváženou síť, tedy i na sociální a citační síť, o kterých je zde řeč. Pro neorientovaný graf je hodnota PageRanku pro jednotlivé uzly velmi blízká stupňům grafu, ale ne totožná [28].

4.4.3 Closeness centrality

Closeness neboli „blízkost“ je definována jako převrácená hodnota míry farness, tedy „dalekost“ [30]. Dalekost je součet všech vzdáleností od uzlu do všech ostatních, tzn. $f(u) = \sum_{v \in V} d_G(u, v)$ a $c(u) = \sum_{v \in V} \frac{1}{d_G(u, v)}$. Podle jiné definice je closeness převrácená hodnota průměrné nejkratší cesty. V podstatě se od předchozí příliš neliší, protože průměrná nejkratší cesta je rovna $\frac{1}{|V|-1} \sum_{v \in V} d_G(u, v)$ a closeness podle této definice:

$$c(u) = \frac{|V| - 1}{\sum_{v \in V} d_G(u, v)}$$

Pro obě definice platí, že čím vyšší hodnota $c(u)$, tím je uzel u významnější podle této míry. Zde se budeme soustředit na druhou definici, protože je častou volbou autorů zabývajících se touto problematikou a existuje pro ni aproximační algoritmus, který si zde uvedeme.

Closeness, stejně jako ostatní míry centrality, modeluje rozptýlení informace napříč sítí. Výše uvedené klasické definici je vytýkáno, že pro přenos informace uvažuje pouze nejkratší cesty, které nejsou vždy jedinou komunikační cestou v síti. Alternativu navrhli Noh a Rieger (2004), kde namísto nejkratších cest používají náhodné procházky (random walk closeness centrality). Náhodná procházka (random walk) je cesta, kde v každém uzlu je náhodně vybrán směr dalšího postupu.

Příkladem může být oběh mincí mezi lidmi. Tento jev nemá s nejkratšími cestami mnoho společného, proto je vhodnější ho modelovat náhodnými procházkami. Oproti tomu například poštovní zásilky očividně cestují po nejkratších cestách. Pokud uvažujeme citační sítě, nemáme jasnou představu o významu náhodných procházek nebo nejkratších cest jako v případě mince nebo dopisu. I přesto očekáváme vysokou podobnost této metody s ostatními.

Nevýhodou closeness centrality je nutnost uvažovat souvislý graf, tedy takový, který obsahuje pouze jednu komponentu. Pokud by měl více komponent, pak by vždy existovala cesta s nekonečnou vzdáleností. Hodnota farness by pak byla automaticky nekonečná a closeness, tedy převrácená hodnota, by byla nulová.

Existuje několik upravených definic, které se mají vypořádat s problémem konektivity a druhotně jsou numericky stabilnější. Jedna z nich zaměňuje převrácenou hodnotu součtu vzdáleností za součet převrácených hodnot vzdáleností $c(u) = \sum_{v \in V} \frac{1}{d_G(u, v)}$ (Opsahl) a druhá $c(u) = \sum_{v \in V} 2^{-d_G(u, v)}$ (Dangalchev). Přesto se nejvíce používá původní definice closeness a výpočet se omezí na největší komponentu.

Algoritmus

Closeness pro všechny vrcholy můžeme přesně vypočítat v čase $O(|E| + |V|^2 \log |V|)$, kde V a E jsou množiny vrcholů a hran sítě [15].

Algoritmus vychází z definice, tedy vyřeší problém všech párů nejkratších cest, čímž rovnou získá hodnoty farness $f(u) = \sum_{v \in V} d_G(u, v)$ a zjištění closeness je poté triviální podle jedné z výše uvedených definic. Výše uvedená složitost platí při použití Dijkstrova algoritmu pro všechny páry cest.

Algoritmus 2 Closeness

```

1: for  $s \in V$  do
2:    $f \leftarrow 0$  ▷ Farness
3:    $d[v] \leftarrow \infty, v \in V$  ▷ Zpočátku jsou uzly nedosažitelné
4:    $d[u] \leftarrow 0$ 
5:    $Q \leftarrow \{s\}$  ▷ Prioritní fronta  $Q$  začíná se zdrojovým vrcholem
6:   while  $Q \neq \emptyset$  do ▷ Dokud není fronta prázdná
7:      $u \leftarrow \text{extract-min}(Q)$  ▷ Vytáhneme uzal s min. vzdáleností  $d$ 
8:      $f \leftarrow f + d[u]$  ▷ Zvýšíme farness pro nový uzal
9:     for sousedící vrchol  $v \in \text{Adj}[u]$  do ▷ Přidáme nově nalezené vrcholy
10:      if  $d[v] > d[u] + w(u, v)$  then
11:         $d[v] \leftarrow d[u] + w(u, v)$ 
12:         $Q \leftarrow Q \cup \{v\}$ 
13:      end if
14:    end for
15:  end while
16:   $c[s] \leftarrow \frac{n-1}{f}$  ▷ Closeness je převrácená hodnota farness
17: end for

```

Výpočet closeness lze snadno paralelizovat, jelikož výpočet stromu nejkratších cest je nezávislá úloha pro každý z $|V|$ vrcholů.

Pro rozsáhlé sítě s miliony uzlů (sociální sítě k dnešnímu datu) je tato metoda příliš náročná. Eppstein a Wang vyvinuli aproximační algoritmus s náročností $O(\frac{\log |V|}{\epsilon}(|V| \log |V| + |E|))$ s chybou $\epsilon \delta$ pro převrácenou hodnotu closeness (s pravděpodobností alespoň $1 - \frac{1}{|V|}$), kde $\epsilon > 0$ a δ je poloměr sítě (nejdelší z nejkratších cest). Na základě tohoto aproximačního algoritmu byl vytvořen jiný aproximační algoritmus pro nalezení k nejvýznamnějších uzlů hodnocených podle closeness centrality TOPRANK [26].

Aproximace

Algoritmus TOPRANK najde prvních k nejvýznamnějších uzlů s vysokou přesností a pro každý z nich přesnou hodnotu closeness. Algoritmus pracuje s myšlenkou, že zjistíme přibližné pořadí uzlů tak, že pro jeden strom nejkratších cest nebudeme počítat se všemi koncovými uzly, ale jen s dostatečně velkým vzorkem této množiny. Přesné hodnoty closeness dosáhneme použitím exaktního algoritmu, který použijeme jen na nejvýznamnější uzly získané z prvního aproximovaného kroku. Klíčovou otázkou je, kolik nejvýznamnějších uzlů musíme uvažovat, aby se jednalo o dostatečně přesný výsledek. Autoři algoritmu uvádějí tento algoritmus s heuristikou, která najde přibližně místo, ve kterém je vhodné výpočet ukončit a považovat za dostatečně

přesný. Sami uvádějí, že tento algoritmus je pouze první krok k návrhu efektivnějšího způsobu jak najít prvních k nejvýznamnějších uzlů [26].

4.4.4 Betweenness centrality

Betweenness je druhá metoda, která modeluje šíření informace sítí pomocí nejkratších cest. Princip betweenness spočívá ve zvýhodnění uzlů, přes kterou teče nejvíce informace. Pokud uzel A komunikuje s uzlem C , můžeme tvrdit, že uzel B , který leží mezi nimi, bude mít roli prostředníka. Být tímto prostředníkem mezi více uzly intuitivně napovídá, že takový uzel bude centrální. „Čím více lidí na mně závisí k vytvoření spojení s jinými lidmi, tím mám větší moc“ [19]. Betweenness měří, na kolika nejkratších cestách se uzel nachází. Více se ale setkáme s definicí, kde do sumy zahrneme poměr cest, na kterých se uzel nachází, k celkovému počtu cest mezi dvěma uzly [16, 1, 4].

$$C(v) = \sum_{s \in V} \sum_{t \in V \setminus s} \frac{\sigma_{st}(v)}{\sigma_{st}}$$

$C_b(v)$ značí hodnotu betweenness centrality pro uzel v , V množinu všech uzlů grafu, σ_{st} je počet nejkratších cest mezi uzly s a t a $\sigma_{st}(v)$ je počet nejkratších cest, které navíc procházejí uzlem v .

Normalizovaný betweenness je hodnota v intervalu od 0 do 1, kterou získáme tak, že betweenness vydělíme celkovým počtem možných cest, tj. $(|V| - 1)(|V| - 2)$ pro orientované grafy a $\frac{(|V|-1)(|V|-2)}{2}$ pro neorientované grafy. Normalizované hodnoty metod centralit jsou nezávislé na velikosti grafu [38].

$$C_b(v) = \frac{1}{(|V| - 1)(|V| - 2)} \sum_{s \in V} \sum_{t \in V \setminus s} \frac{\sigma_{st}(v)}{\sigma_{st}}$$

Vznik betweenness je připisován sociologovi Lintonu Freemanovi [16] a byl původně definován pro neorientované grafy.

Brandesův algoritmus

Ve své práci [4] Ulrik Brandes zmiňuje (do té doby nejrychlejší) algoritmus pro výpočet betweenness centrality s časovou náročností $\theta(|V|^3)$ a $\theta(|V|^2)$ paměťovými nároky. Tento způsob přistupuje k problému nejkratších cest způsobem all-pair shortest paths. Brandesův způsob využívá algoritmu pro nalezení nejkratších cest z jednoho bodu,

kde výsledný algoritmus pracuje s pamět'ovou náročností $O(|V| + |E|)$ a běží v čase $O(|V||E|)$ pro nevážený graf nebo $O(|V||E| + |V|^2 \log |V|)$ pro vážený graf.

Brandes ve své práci o algoritmu uvádí pseudokód pro nevážený graf, který je následně snadné pozměnit pro vážený graf drobnými úpravami a zaměněním obyčejné fronty za prioritní frontu; kompletní důkaz správnosti algoritmu a porovnání standardního algoritmu s tímto [4].

Algoritmus 3 Brandes

```

1:  $b[v] \leftarrow 0, v \in V$  ▷ Inicializujeme hodnoty betweenness na 0
2: for  $s \in V$  do ▷ V exaktním výpočtu počítáme s celou množinou V
3:    $S \leftarrow$  prázdný zásobník ▷ Zásobník pro „dokončené“ vrcholy
4:    $P[w] \leftarrow$  prázdný seznam,  $w \in V$  ▷ Seznam předchůdců
5:    $\sigma[t] \leftarrow 0, t \in V$ 
6:    $\sigma[s] \leftarrow 1$ 
7:    $d[t] \leftarrow -1, t \in V$  ▷ Vzdálenosti jsou inicializovány na -1, tj. symbolicky nekonečno
8:    $d[s] \leftarrow 0$ 
9:    $Q \leftarrow \{s\}$  ▷ Fronta nebo prioritní fronta začíná se startovním vrcholem
10:  while  $Q \neq \emptyset$  do ▷ Opakujeme, dokud není fronta prázdná
11:     $v \leftarrow dequeue(Q)$ 
12:     $push(S, v)$ 
13:    for sousedící vrchol  $w \in Adj[v]$  do
14:      if  $d[w] < 0$  then
15:         $enqueue(Q, w)$ 
16:         $d[w] \leftarrow d[v] + 1$ 
17:      end if
18:      if  $d[w] = d[v] + 1$  then ▷ Nejkratší cesta do  $w$  přes  $v$ ?
19:         $\sigma[w] \leftarrow \sigma[w] + \sigma[v]$ 
20:         $push(P[w], v)$ 
21:      end if
22:    end for
23:  end while
24:   $\delta[v] \leftarrow 0, v \in V$  ▷  $\delta$  je závislost uzlu  $s$  na ostatních
25:  while  $S \neq \emptyset$  do ▷  $S$  vrátí vrcholy v pořadí s nezvyšující se vzdáleností od  $s$ 
26:     $w \leftarrow pop(S)$ 
27:    for  $v \in P[w]$  do
28:       $\delta[v] \leftarrow \delta[v] + \frac{\sigma[v]}{\sigma[w]}(1 + \delta[w])$ 
29:    end for
30:    if  $w \neq s$  then ▷ Betweenness je součet dependencí  $\delta[w]$ 
31:       $b[w] \leftarrow b[w] + \delta[w]$ 
32:    end if
33:  end while
34: end for

```

Algoritmus lze paralelizovat stejně jako algoritmus pro closeness centrality, pokud zajistíme synchronizaci vláken při přístupu k hodnotám betweenness. Narozdíl od algoritmu pro closeness, kde výpočet pro jeden uzel ovlivňuje hodnotu closeness

pouze pro tento uzel, u betweenness výpočet vycházející z jednoho uzlu ovlivňuje hodnotu betweenness potencionálně i pro všechny ostatní uzly.

Aproximace

I přes použití rychlejšího Brandesova algoritmu je výpočet betweenness centrality příliš náročný výpočet pro síť reálného světa (např. biologické, dopravní nebo webové sítě) a pokud nám jde více o relativní pořadí uzlů podle hodnoty betweenness než o hodnotu samotnou, lze oželit přesný výpočet přibližným, který příliš nezmění výsledné umístění v žebříčku nejvýznamnějších uzlů.

Bader, Kintali, Madduri, Mihail ukazují aproximační algoritmus pro betweenness s odhadem chyby. Myšlenkou je jednoduchá lineární extrapolace Brandesova algoritmu, pokud do výpočtu zahrneme pouze náhodný vzorek namísto celé množiny vrcholů. Necht' k je velikost vzorku množiny vrcholů, se kterým počítáme, pak extrapolovaná hodnota betweenness je $\frac{|V|S}{k}$, kde S je vypočtená přibližná hodnota.

4.5 Hledání nejkratších cest

Hledání nejkratších cest v grafu je historicky starý problém, jehož matematický výzkum přišel relativně pozdě v porovnání s jinými problémy kombinatorické optimalizace (nejmenší kostra grafu, přiřazovací a dopravní problém). Pravděpodobně byl výzkum opožděn, protože se jedná o intuitivní a relativně jednoduchý problém, ale jakmile se dostal do středu zájmu, bylo nezávisle na sobě nalezeno několik metod řešení různými lidmi (Shimbel [33], Ford [14], Dantzig [9], Bellman [2], Moore [25], Dijkstra [11]) [32].

Z hlediska metod řešení můžeme uvažovat několik kategorií algoritmů - nalezení všech párů nejkratších cest (all-pairs shortest paths problem), nalezení nejkratší cesty mezi počátečním a koncovým vrcholem (source-target) nebo nalezení stromu nejkratších cest, máme-li zadán počáteční vrchol (single source shortest path problem).

4.5.1 Single source shortest path

Pokud hledáme pouze jednu cestu mezi dvěma vrcholy (source-target), nemusíme počítat celý strom nejkratších cest, ale můžeme zastavit výpočet při dosažení požadovaného vrcholu.

BFS

Prohledávání do šířky z jednoho bodu (breadth first search) je algoritmus, který najde nejkratší cesty z jednoho bodu do všech ostatních v případě neváženého grafu v čase $O(E)$. Pro vážený graf by zjistil nejkratší cesty, kde metrika vzálenosti by byla počet skoků mezi uzly.

Algoritmus 4 Prohledávání do šířky

```

1: function BFS( $G, s$ )
2:    $d[v] \leftarrow \infty, v \in V$                                 ▷ Uzly jsou zpočátku nedosažitelné
3:    $d[s] \leftarrow 0$ 
4:    $p[v] \leftarrow \text{NIL}, v \in V$                                 ▷ A nemají implicitně žádné předchůdce
5:    $Q \leftarrow \{s\}$                                            ▷ Fronta začíná s počátečním vrcholem
6:   while  $Q \neq \emptyset$  do                                    ▷ Dokud není prázdná
7:      $u \leftarrow \text{dequeue}(Q)$ 
8:     for sousedící vrchol  $v \in \text{Adj}[u]$  do
9:       if  $d[v] = -1$  then                                       ▷ Pro všechny nově objevené vrcholy
10:         $d[v] \leftarrow d[u] + 1$                                 ▷ Nově vypočtená vzdálenost
11:         $Q \leftarrow Q \cup \{v\}$                                 ▷ Přidáme nově objevené uzly do fronty
12:         $p[v] \leftarrow u$                                        ▷ Předchůdce uzlu  $v$  je  $u$ 
13:      end if
14:    end for
15:  end while
16:  return  $p$ 
17: end function

```

Bellman-Fordův algoritmus

Bellman-Fordův algoritmus je aplikací dynamického programování na nejkratší cesty z jednoho vrcholu do všech ostatních. Jeho využití najdeme zejména tam, kde se mohou objevit záporné váhy hran. Časová náročnost je $O(|V||E|)$.

Dijkstrův algoritmus

Dijkstrův algoritmus řeší „single source shortest path problem“ pomocí greedy (žravé) strategie. Časová náročnost běžné implementace $O((|E| + |V|) \log |V|)$ je lepší než u Bellman-Fordova algoritmu, ale Dijkstrova greedy strategie funguje, pouze pokud v grafu neexistují záporně ohodnocené hrany. Časová náročnost závisí především na implementaci klíčové datové struktury prioritní fronty, která v každém kroku vybírá uzel s nejmenší vzdáleností od zdrojového vrcholu (greedy strategie). V případě použití obyčejného spojového seznamu jako prioritní fronty získáme kvadratickou náročnost $O(|V|^2)$. Nejčastěji se setkáme s prioritní frontou implementovanou pomocí binární haldy, která zajišťuje čas běhu právě $O((|E| + |V|) \log |V|)$. Pomocí Fibonacciho haldy získáme doposud nejrychlejší Dijkstrův algoritmus s časem běhu $O(|E| + |V| \log |V|)$, ale pro běžné účely (grafy menší než miliony vrcholů) je nejvýhodnější binární halda [8, 15]. Pro nevážený graf je Dijkstrův algoritmus ekvivalentní prohledávání do šířky.

Dijkstrův algoritmus je klíčovým pro všechny implementované metody založené na nejkratších cestách, proto uvedeme pseudokód [8].

Algoritmus 5 Dijkstrův algoritmus

```

1: function DIJKSTRA( $G, w, s$ )           ▷  $s$  je zdrojový vrchol a  $w$  je váhová funkce
2:    $d[v] \leftarrow \infty, v \in V$            ▷ Vrcholy jsou zpočátku nedosažitelné
3:    $d[s] \leftarrow 0$ 
4:    $p[v] \leftarrow \text{NIL}, v \in V$            ▷ Předchůdci vrcholů zpočátku neexistují
5:    $Q \leftarrow V$                            ▷ Prioritní fronta  $Q$  obsahuje všechny vrcholy
6:   while  $Q \neq \emptyset$  do                 ▷ Postupujeme, dokud není prázdná
7:      $u \leftarrow \text{extract-min}(Q)$        ▷ Vytáhneme z fronty vrchol s nejmenší hodnotou
        $d[u]$ 
8:     for sousedící vrchol  $v \in \text{Adj}[u]$  do
9:       if  $d[v] > d[u] + w(u, v)$  then
10:         $d[v] \leftarrow d[u] + w(u, v)$            ▷ Relaxace hrany  $\{u, v\}$ 
11:         $p[v] \leftarrow u$ 
12:       end if
13:     end for
14:   end while
15:   return  $p$                                ▷ Strom předchůdců
16: end function

```

Výsledkem je strom předchůdců, který reprezentuje strom nejkratších cest. Princip Dijkstrova algoritmu se objevuje i u algoritmů pro closeness, betweenness a BFS. V tomto případě do prioritní fronty prvotně zahrneme všechny vrcholy. Jinou možností je přidávat je postupně, aby vybírání z prioritní fronty ($\text{extract-min}(Q)$) bylo rychlejší. Tento způsob je použit v algoritmu pro closeness.

4.5.2 All-pair shortest paths

Do této kategorie spadají maticové metody, tj. graf je zadán jako matice sousednosti nebo matice sousednost s váhami hran.

Shimbelova metoda

Shimbelova metoda (1955) používá upravené maticové násobení k získání $|V|$ -té mocniny matice sousednosti. Celková časová náročnost je $O(|V|^4)$, protože provedeme $|V|$ „násobení“ čtvercové matice o složitosti $O(|V|^3)$. Shimbelovo upravené násobení nahrazuje sčítání a násobení za minimum a sčítání:

$$x + y \equiv \min(x, y)$$

$$xy \equiv x + y$$

Floyd-Warshallův algoritmus

Floyd-Warshallův algoritmus snižuje časovou náročnost na $O(|V|^3)$ použitím dynamického programování. Graf je opět zadán jako vážená matice sousednosti. Rekurentní vzorec dynamického programování pro tento algoritmus je:

$$d_0(u, v) = \mathbf{A}_{uv}$$

$$d_{k+1}(u, v) = \min(d_k(u, v), d_k(u, k) + d_k(k, v))$$

Jednoduše zkusíme, zda je kratší cesta mezi vrcholy u a v , kterou již známe, nebo jiná cesta za použití nějakého vrcholu k , který leží mezi nimi. Výpočet provádíme pro všechny páry vrcholů pro každý vrchol k ($|V|^2|V|$).

Johnsonův algoritmus

Johnsonův algoritmus nepatří mezi maticové metody, protože využívá metod single source shortest path pro všechny vrcholy. V principu jednoduše použijeme Dijkstrův algoritmus pro každý vrchol zvlášť, ale dovoluujeme váhy hran i záporné. V případě záporných hran je nutné provést transformaci vah pomocí Bellman-Fordova algoritmu, která v grafu nepozmění nejkratší cesty [8].

4.6 Ostatní míry významnosti autorů

Míry centrality jsou určeny pro relativní seřazení uzlů pro obecný graf. Jelikož zde pracujeme speciálně s citačními sítěmi, zajímá nás, jestli existují i jiné metody ohodnocení autorů, které ani nevyžadují konstrukci citačních sítí, ale vyžadují pouze údaje v databázi.

Setkáme se s metodami AWCR, AW-index, Eigenfactor, Egghův g-index, E-index, I10-index, R-impact, Wu Index a zejména H-index [21], který zde byl implementován jako jediná metoda, která nespadá pod míry centrality.

4.6.1 H-index

H-index je metoda pro měření produktivity autora a významu jeho publikované vědecké práce. Metoda byla navržena fyzikem Jorge E. Hirschem pro zjišťování relativní významnosti vědců publikujících v oboru teoretické fyziky. Původní Hirschova definice zní:

Vědec má index h pokud h z jeho N_p publikací má každá alespoň h citací a žádná ze zbylých $(N_p - h)$ publikací nemá více než h citací.

Tato metoda je specifická pro citační sítě a nelze ji aplikovat na sociální nebo obecnou komplexní síť.

Výpočet probíhá tak, že seřadíme autorovy publikace P sestupně podle počtu citací $c : P \mapsto \mathbb{N}$ a poté od začátku tohoto seřazeného seznamu hledáme tu publikaci, jejíž pořadové číslo v tomto seznamu je nižší nebo rovné počtu jejích citací. Toto číslo je pak h -index autora.

Algoritmus 6 H-index

```
1: function H-INDEX( $P, c$ )
2:    $sort(P, key \leftarrow c)$            ▷ Seřadíme publikace sestupně podle počtu citací
3:    $h \leftarrow 1$ 
4:   for  $p \in P$  do
5:     if  $c(p) \leq h$  then           ▷ Bod, kde počet citací vyrovná  $h$ , je  $h$ -index autora
6:       break
7:     end if
8:      $h \leftarrow h + 1$ 
9:   end for
10:  return  $h$ 
11: end function
```

5 Výsledky

Výsledky této práce jsou znázorněny v tabulkách korelací mezi jednotlivými metodami (tab. A.1, A.2, A.3, A.4), tabulkami (tab.) srovnávající pořadí prvních třiceti autorů mezi jednotlivými metodami a porovnání implementovaných metod s oceněními (ACM Codd, ...)

5.1 Porovnání implementovaných metod

Zajímá nás, zda-li jsou metody mezi sebou podobné a do jaké míry. Pro porovnání metod mezi sebou je použit Spearmanův koeficient pořadové korelace.

Tab. A.1 ukazuje vypočtené Spearmanovy koeficienty pořadové korelace mezi všemi páry implementovaných metod. Do výpočtu koeficientu korelace byly zahrnuty pouze ty uzly, které mají alespoň jednu hranu. Pokud by byly zachovány, většina koeficientů korelace by byla velmi vysoká (větší než 0.95), protože většina vrcholů, zejména v databázi DBLP, jsou osamocené vrcholy. Protože jsou výsledky seřazeny podle hodnoty centrality a poté podle jména autora, všechny tyto izolované vrcholy, které nijak nepřispívají k hodnotě centrality, jsou pro všechny metody umístěny na posledních pozicích v tomtéž pořadí. Důsledkem je vysoký koeficient korelace, který nám však nic nevypoví o podobnosti výsledků metod.

Jelikož výpočet pro centralitu closeness ve všech implementovaných variantách (vstupní hrany, výstupní hrany, vážený closeness) byl proveden pouze na největší silně spojitě komponentě, nelze přímo provést srovnání pomocí koeficientu korelace s výsledky metod, které byly aplikovány na celý graf, protože se jedná o dvě neporovnatelné proměnné. Proto byli z výsledků metod pro celý graf odebráni ti autoři, kteří se nenacházejí v hlavní silně spojitě komponentě. Pro tyto žebříčky bylo provedeno stejné srovnání pomocí Spearmanova koeficientu korelace pořadí křížově mezi všemi metodami, viz tab. A.2

Tab. 5.1 popisuje zkratky metod, časy běhu pro síť DBLP a CiteSeer a velikost největší kliky (k_{DBLP} pro DBLP a k_{CS} pro CiteSeer) v symetrizovaném grafu autorů. Největší klika je nalezena z top 20 autorů pro každou metodu. Výpočty byly vykonány na Intel®Core™2 Quad Q8400 přetaktovaném na 3.1 GHz s 4 GB fyzické paměti. Betweenness centrality bylo vypočteno paralelně na čtyřech jádrech, zbylé metody byly relativně časově nenáročné, proto nebylo nutné paralelizovat výpočet.

5.1.1 Spearmanův koeficient pořadové korelace

Spearmanův koeficient je klasický Pearsonův koeficient korelace, který je aplikovaný na proměnné s pořadím [37]. Koeficient korelace obecně dosahuje hodnot od -1 do

⁰Největší klika mezi prvními 20 autory

Tabulka 5.1: Implementované metody

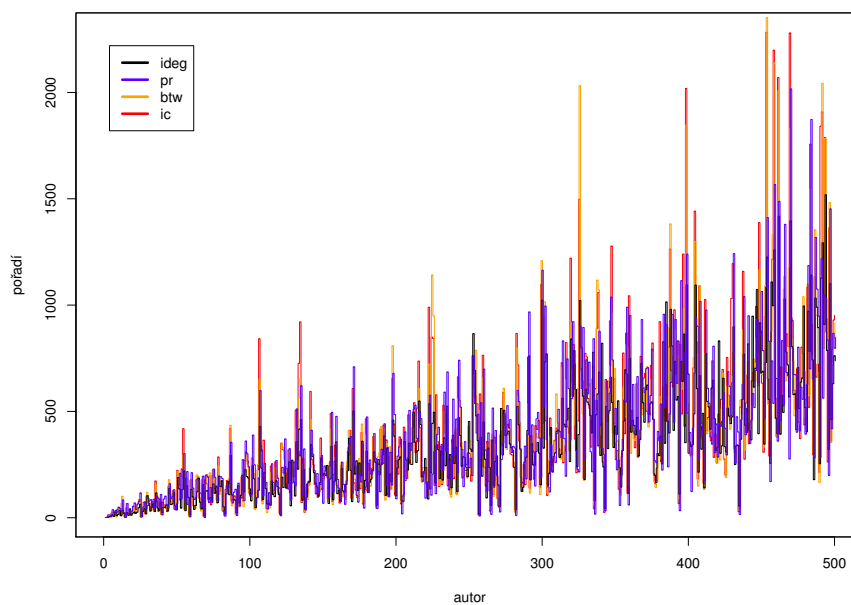
zkratka	metoda	t_{DBLP}	t_{CS}	k_{DBLP}	k_{CS}
hi	H-index	0:04.627	0:00:15.534	17	8
ideg	indegree	0:00.049	0:00:20.372	19	13
odeg	outdegree	0:00.036	0:00:00.343	16	8
deg	degree	0:00.036	0:00:00.060	19	12
wideg	vážený indegree	0:00.099	0:00:21.899	18	13
wodeg	vážený outdegree	0:00.046	0:00:01.171	17	11
wdeg	vážený degree	0:00.062	0:00:00.214	17	10
pr	PageRank	0:00.376	0:00:02.758	18	7
btw	betweenness	24:56.814	7:45:25.904	19	13
btwA	betweenness aproximace	6:51.216	1:50:31.412	19	12
wBtwA	vážený betweenness aproximace	8:54.073	13:01:51.078	16	18
ic	closeness vstupní hrany	0:05.583	1:12:32.345	19	11
oc	closeness výstupní hrany	0:11.313	2:23:36.778	16	6
wic	vážený closeness vstupní hrany	1:14.054	33:44:56.355	18	15

1, přičemž hodnota 1 znamená naprostou lineární závislost mezi porovnávanými proměnnými tak, že s rostoucí první proměnnou roste i druhá proměnná. Hodnota -1 znamená rovněž naprostou lineární závislost, ale při rostoucí jedné proměnné druhá proměnná klesá. Hodnota 0 znamená kompletní náhodnost či nezávislost mezi měřeními proměnnými.

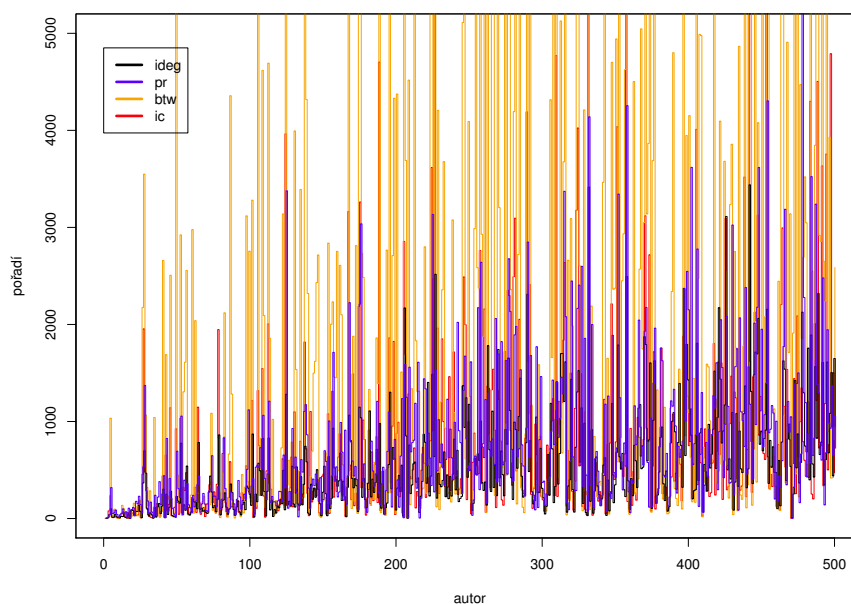
5.2 Žebříčky významných autorů

Tabulky B.1 a B.2 ukazují prvních 30 autorů v pořadí podle metody h-index. Pro každého autora je zobrazeno jeho umístění podle všech zbylých implementovaných metod. Poslední řádek u každé metody označuje součet pořadí uvedených 30ti autorů. Stejná data jsou zobrazena v grafech 5.1 a 5.2 pro prvních 500 autorů z největší silně spojitě komponenty, kde porovnáváme metody indegree, PageRank, betweenness a closeness.

Výstupy programu, tj. jednotlivá pořadí autorů a jejich ocenění jsou uvedeny v příloze C, v sekci C.1 pro databázi DBLP a v sekci C.2 pro databázi CiteSeer. Hodnoty u metody PageRank jsou v žebříčcích transformovány z intervalu $[0; 1]$ na $[0; |V|]$, protože při zaokrouhlení na tři desetinná místa jsou všechny hodnoty PageRanku zanedbatelně malé.



Obrázek 5.1: Pořadí prvních 500 autorů DBLP dle H-indexu z největší silně spojené komponenty



Obrázek 5.2: Pořadí prvních 500 autorů CiteSeer dle H-indexu z největší silně spojené komponenty

5.3 Porovnání metod s oceněními

Pro zjištění, jestli se výsledky implementovaných metod shodují s uvedenými oceněními, použijeme metodu součtu pořadí oceněných autorů, viz tab. 5.2. Tzn. že pro každé ocenění a pro každou metodu sečteme pořadí všech autorů, kteří byli oceněni danou cenou. Tuto jednoduchou míru můžeme porovnat pouze mezi jednotlivými metodami pro jedno ocenění, ale ne mezi různými oceněními pro jednu metodu. Jednoduše protože např. Turingova cena je ve výsledcích udělena pouze několika autorům, z čehož plyne malý součet pořadí, kdežto velké množství autorů je členy ACM Fellows, tím pádem velký součet pořadí oceněných.

Tabulka 5.2: Součty pořadí oceněných autorů pro DBLP

	Codd	ACM Fellows	ISI HC	Turing
hi	819	6 791 886	2 555 487	725 373
ideg	716	801 295	376 586	61 665
odeg	4 605	3 904 308	2 367 669	257 133
deg	741	1 077 282	525 324	84 191
wideg	571	802 148	374 344	61 757
wodeg	4 491	3 902 718	2 365 245	256 620
wdeg	692	1 065 801	511 523	82 652
pr	867	805 267	403 109	41 735
btw	796	828 571	399 720	70 342
btwA	821	1 024 010	428 133	68 996
wBtwA	567	1 057 788	473 146	77 168
ic ¹	805	63 753	34 004	1 384
oc ¹	6 607	129 810	59 701	4 098
wic ¹	512	73 729	38 045	1 482

¹Platí pro největší silně spojitou komponentu

¹Platí pro největší silně spojitou komponentu

Tabulka 5.3: Součty pořadí oceněných autorů pro CiteSeer

	Codd	ACM Fellows	ISI HC	Turing
hi	174 280	10 388 626	5 679 672	1 071 346
ideg	46 997	11 345 011	6 812 044	1 013 904
odeg	153 490	19 741 913	13 189 294	1 883 554
deg	35 156	13 989 257	8 822 049	1 074 213
wideg	52 467	11 589 445	6 857 195	1 048 460
wodeg	134 992	19 751 885	12 877 820	1 829 697
wdeg	43 918	14 153 909	8 728 725	1 115 118
pr	81 704	11 469 485	6 426 334	938 757
btw	95 920	11 074 054	5 879 992	848 269
btwA	96 643	11 691 668	6 058 579	1 007 215
wBtwA	103 043	12 899 982	6 365 899	1 224 576
ic ¹	55 490	6 835 926	4 047 426	560 873
oc ¹	124 724	9 108 426	6 074 616	903 529
wic ¹	86 773	7 576 984	4 585 710	710 976

5.4 Aproximace betweenness centrality

Výpočet betweenness je časově nejnáročnější ze všech implementovaných metod. Pro citační síť autorů databáze DBLP není exaktní výpočet problém, ale pro rozsáhlou síť databáze CiteSeer je výpočet váženého betweenness časově velmi náročný úkol, jak můžeme nahlédnout jen z doby běhu aproximovaného betweenness v tab. 5.1. Tab. 5.4 znázorňuje Spearmanovo koeficienty korelace pořadí autorů DBLP mezi exaktním betweenness a aproximovanými, kde velikost množiny vrcholů, kterou uvažujeme ve výpočtu, je postupně $\frac{|V|}{2}, \frac{|V|}{4}, \frac{|V|}{8}, \dots$

Tabulka 5.4: Tabulka Spearmanova koeficientů korelace mezi exaktním a aproximovaným betweenness pro DBLP

zlomek velikosti množiny V	Spearmanova korelace	doba běhu [min:sec.msec]
1	1.000 000	23:55.038
2	0.998 151	12:05.370
4	0.996 191	06:11.010
8	0.990 510	03:03.723
16	0.979 103	01:30.228
32	0.976 887	00:44.870
64	0.976 745	00:22.630

6 Diskuse

V této kapitole se pokusíme přiblížit, proč jsou jednotlivé metody podobné samy sobě a co naznačuje, že skutečně měříme významnost a ne nějakou jinou vlastnost.

Pokud bychom v praktických aplikacích chtěli měřit významnost autorů nebo obecně uzlů v jakékoliv komplexní síti, metody založené na nejkratších cestách nepřipadají v úvahu, jestliže hledíme na časovou náročnost. Ta se zvyšuje i se zvyšující se hustotou grafu, viz doba běhu closeness pro DBLP a CiteSeer (tab. 5.1).

I použití aproximací, které výrazně urychlí výpočet, může být stejně nebo méně přesné než použití jiné metody. V implementaci byla aproximována pouze centralita betweenness, ale pro druhou náročnou metodu closeness byl zmíněn aproximační algoritmus. Protože tato aproximace nebyla implementována a vyzkoušena, nemůžeme o ní poskytnout jakýkoliv závěr.

Z časového hlediska je výhodnější používat metody degree nebo eigenvector.

6.1 Podobnost výsledků jednotlivých metod

Množina autorů je pro obě databáze větší než 300 000. Pro proměnné o této velikosti můžeme uvažovat statistickou signifikanci Spearmanovo koeficientu pořadové korelace od hodnoty 0.01 a výše (cite Yan, Ding - Applying centrality measures to impact analysis: A coauthorship network analysis). Záporné korelace by poukazovaly na opačné pořadí, které je nežádoucí a korelace menší než 0.01 budeme považovat za nevýznamné.

Všechny korelace, tj. mezi všemi porovnávanými metodami pro obě databáze a pro obě varianty (celý graf, největší silně spojitá komponenta), vyšly větší než 0.1 s minimem 0.102 (pr - odeg), 0.340 (pr - oc), 0.220 (pr - odeg), 0.207 (pr - oc) pro DBLP celý graf, DBLP největší silně spojitou komponentu, CiteSeer celý graf a CiteSeer největší silně spojitou komponentu, respektive.

Protože je eigenvector centralita rozšířením degree centrality (jednoduše řečeno iterovaný indegree), vskutku si povšimneme vysokých korelací mezi indegree a PageRankem ve všech čtyřech případech (. . .).

Metody založené na nejkratších cestách intuitivně rovněž dosahují vysokých korelací, přestože je význam odlišný (btw - ic, atd.)

H-index je zcela očekávaně podobnější všem metodám než odeg, wodeg a oc, které nemají přímo měřit významnost, ale poukázat na fakt, že významní autoři navíc mají tendence ovlivňovat i jiné autory tím, že je citují. Přestože nepatří do metod centrality jako ostatní metody, relativně vysoká korelace s mírami centrality poukazuje na jistou podobnost.

Z vypočtených koeficientů korelací si můžeme dovolit tvrdit, že všechny implementované metody měří zhruba stejnou vlastnost - významnost, vyjma metod

odeg, wodeg a oc, které se shodují s ostatními, jelikož jsou spíše důsledkem vysoké významnosti autora.

6.2 Vliv vah na výsledky

Intuitivně bychom očekávali, že metody na vážené síti (wideg, wodeg, wBtwA, wic) budou vzájemně korelovat výše než ekvivalentní metody na nevážené síti (ideg, odeg, btwA, ic). U DBLP například dosahuje korelace mezi wBtwA a wideg (0.849), wodeg (0.140) a wdeg (0.549) výše než mezi wBtwA a ideg (0.830), odeg (0.134) a deg (0.529), viz tab. A.1. U výsledků pro DBLP pro největší silně spojitou komponentu (tab. A.2) dostáváme stejnou podobnost, tj. korelace mezi wBtwA a wideg (0.889), wodeg (0.529) a wdeg (0.766) je vyšší než mezi ideg (0.865), odeg (0.505) a deg (0.753). Korelace mezi váženým closeness a váženým betweenness dominují nad korelacemi mezi váženým betweenness a neváženým closeness a naopak, tj. korelace mezi wic a wBtwA (0.930), wideg (0.887), wodeg (0.527) a wdeg (0.763) jsou vyšší než mezi wic a btwA (0.864), ideg (0.884), odeg (0.493) a deg (0.734). Mohli bychom proto dedukovat, že všechny výsledky jsou si podobnější, pokud se aplikují výhradně na vážené síti nebo výhradně na nevážené síti, než pokud porovnáváme vážené s neváženými.

Nemůžeme ale najisto tvrdit, že metody pro váženou síť jsou přesnější než metody pro neváženou síť. Logicky by měly být vážené metody přesnější, protože pracují s větších množstvím informace, ale protože není významnost jednoznačně definována, nemůžeme tvrdit, která z variant je přesnější.

6.3 Vstupní a výstupní hrany

Přestože vstupní a výstupní hrany mají opačný význam, existuje nezanedbatelná korelace mezi jejich interpretacemi „cituji“ a „jsem citován“, tzn. pokud je autor často citován, pak je velká šance, že zároveň cituje ostatní autory. Metody odeg, wodeg a oc, které zastupují výstupní hrany „cituji“, jsou spíše důsledkem významnosti autora než její příčinou. Pokud je autor dlouho ve své branži a vybudoval si významnou pozici, pak můžeme očekávat, že zároveň referencoval práci mnoho jiných autorů.

Zajímavým jevem je použití vstupních a výstupních hran u closeness centrality a jejích variantách. Zatímco u betweenness centrality počítáme s přirozenými výstupními hranami a získáme vyšší korelace s metodami H-index, indegree, PageRank, než s metodami outdegree nebo vážený outdegree, které intuitivně ani objektivně neměří významnost autora (rozdíl mezi „cituji“ u outdegree a „jsem citován“ u ostatních, tj. H-index, indegree, PageRank), u closeness centrality získáme vyšší korelace naopak při použití vstupních hran namísto výstupních.

6.4 Shoda výsledků s oceněními

Tab. 5.2 pro DBLP a tab. 5.3 pro CiteSeer, srovnávají jednotlivé metody s oceněními tak, že menší hodnota znamená méně oceněných autorů na nižších pozicích. Tedy oceněné autory nalezneme na vyšších pozicích, pokud máme v tabulce nízkou hodnotu. Všechny metody se přibližně v hodnotách shodují, vyjma metod odeg, wodeg a oc. Tyto metody ukazují na významnost autora, ale jsou více důsledkem než příčinou. Můžeme je použít ke srovnání ostatních metod, které mají obecně nižší hodnotu součtu oceněných pozic, a znázorňují větší shodu s oceněními.

V žebříčcích nejvýznamnějších autorů jednotlivých implementovaných metod si povšimneme koncentraci ocenění na vyšších pozicích. Zejména u databáze DBLP a ocenění ACM SIGMOD Edgar F. Codd Innovations Award (Codd), protože DBLP je databáze autorů se zaměřením na databázové systémy a ocenění ACM SIGMOD Edgar F. Codd Innovations Award je rovněž z oblasti databází a databázových systémů.

Srovnání s Turingovo cenou bychom ani neměli používat pro objektivní závěr, protože v databázích DBLP a CiteSeer je autorů, kteří jí byli oceněni, minimum. ACM Fellows je na druhou stranu hojně zastoupena mezi autory DBLP a CiteSeer, a proto ji s konfidencí můžeme použít jako referenci pro naše pozorování.

6.5 Aproximace betweenness

Jednoduchá lineární extrapolace u aproximace betweenness centrality se zdá být dostatečně přesnou, pokud zmenšujeme velikost množiny vrcholů, se kterou pracuje Brandesův algoritmus.

Zcela očekávaně dostáváme vysoké korelace mezi exaktním a aproximovaným betweenness a mezi metodami a jejich alternativami pro váženou síť (btwA - wBtwA, deg - wdeg, apod.), viz tab. A.1, A.2, A.3, A.4. Exaktní betweenness se rovněž shoduje (korelace > 0.9) s aproximacemi, kde používáme různé velikosti množiny vrcholů (tab. 5.4).

7 Závěr

Literatura

- [1] ANTHONISSE, J. – BESLISKUNDE, S. M. C. A. A. M. *The rush in a directed graph*. Chicago : SMC, 1973.
- [2] BELLMAN, R. On a Routing Problem. *Quarterly of Applied Mathematics*. 1958, 16, s. 87–90.
- [3] BONACICH, P. Factoring and weighting approaches to status scores and clique identification. *Journal of Mathematical Sociology*. 1972, 2, 1, s. 113–120.
- [4] BRANDES, U. A Faster Algorithm for Betweenness Centrality. *Journal of Mathematical Sociology*. 2001, 25, 2, s. 163–177.
- [5] BRIN, S. – PAGE, L. The anatomy of a large-scale hypertextual Web search engine. *Computer Networks and ISDN Systems*. 1998, 30, s. 107–117.
- [6] BURT, R. S. *Toward a Structural Theory of Action*. New York : Academic Press, 1982.
- [7] COLEMAN, J. S. *The Mathematics of Collective Action*. Chicago : Aldine Pub. Co., 1973.
- [8] CORMEN, T. H. et al. *Introduction to Algorithms*. Cambridge, MA : The MIT Press, 2 edition, 2001.
- [9] DANTZIG, G. B. Discrete-Variable Extremum Problems. *Operations Research*. 1957, 5, 2.
- [10] DASGUPTA, S. – PAPADIMITRIOU, C. – VAZIRANI, U. *Algorithms*. New York, NY : McGraw-Hill, 2008.
- [11] DIJKSTRA, E. W. A note on two problems in connexion with graphs. *Numerische Mathematik*. December 1959, 1, 1, s. 269–271.
- [12] COMPUTING MACHINERY, A. *ACM Fellows* [online]. [cit. 15.03.2013]. Dostupné z: <http://fellows.acm.org/>.

-
- [13] COMPUTING MACHINERY, A. *A.M. Turing Award* [online]. [cit. 15.03.2013]. Dostupné z: <http://amturing.acm.org>.
- [14] FORD, L. R. *Network Flow Theory*. Report P-923, The Rand Corporation, 1956.
- [15] FREDMAN, M. L. – TARJAN, R. E. Fibonacci heaps and their uses in improved network optimization algorithms. *J. ACM*. July 1987, 34, 3, s. 596–615.
- [16] FREEMAN, L. C. A Set of Measures of Centrality Based on Betweenness. *Sociometry*. March 1977, 40, 1, s. 35–41.
- [17] FREEMAN, L. C. Centrality in social networks: Conceptual clarification. *Social Networks*. January 1979, 1, 3, s. 215–239.
- [18] FRIEDKIN, N. E. Theoretical Foundations for Centrality Measures. *Academic Journal of Sociology*. 1991, 96, 6, s. 1478–1504.
- [19] HANNEMAN, R. A. – RIDDLE, M. *Introduction to Social Network Methods* [online]. 2005. [cit. 11.03.2013]. Dostupné z: <http://faculty.ucr.edu/~hanneman/nettext/>.
- [20] HUBBELL, C. H. An Input-Output Approach to Clique Identification. *Sociometry*. 1965, 28, s. 377–399.
- [21] INTERNATIONAL, H. *Author impact metrics* [online]. 2013. [cit. 20.04.2013]. Dostupné z: http://hlwiki.slais.ubc.ca/index.php/Author_impact_metrics.
- [22] LANGVILLE, A. N. – MEYER, C. D. *Google's PageRank and Beyond: The Science of Search Engine Rankings*. Princeton, NJ : Princeton University Press, 2006.
- [23] LIBRARY, A. R. *Citation & Full-text Databases* [online]. [cit. 23.03.2013]. Dostupné z: <http://library.amnh.org/research-tools/citation-full-text-databases>.
- [24] MALSEED, M. – VISE, D. *The Google Story*. New York, NY : Delacorte Press, 2005.
- [25] MOORE, E. F. The shortest path through a maze. In *Proceedings of the International Symposium on the Theory of Switching, and Annals of the Computation Laboratory of Harvard University*, s. 285–292. Harvard University Press, 1959.
- [26] OKAMOTO, K. – CHEN, W. – LI, X.-Y. Ranking of Closeness Centrality for Large-Scale Social Networks. In *Proceedings of the 2nd annual international workshop on Frontiers in Algorithmics*, FAW '08, s. 186–195, Berlin, Heidelberg, 2008. Springer-Verlag.

-
- [27] PAGE, L. *PageRank: Bringing Order to the Web* [online]. 1997. [cit. 20.04.2013]. Dostupné z: <http://web.archive.org/web/20020506051802/www-diglib.stanford.edu/cgi-bin/WP/get/SIDL-WP-1997-0072?1>.
- [28] PERRA, N. – FORTUNATO, S. Spectral centrality measures in complex networks. *Phys. Rev. E*. Sep 2008, 78, s. 036107.
- [29] REUTERS, T. *Highly Cited Research* [online]. [cit. 15.03.2013]. Dostupné z: <http://researchanalytics.thomsonreuters.com/highlycited/>.
- [30] SABIDUSSI, G. The centrality index of a graph. 1966, 31, 4, s. 581–603.
- [31] SALKIND, N. J. *Encyclopedia of educational psychology*. Thousand Oaks, CA : Sage publications, 2008.
- [32] SCHRIJVER, A. On the history of the shortest path problem. *Documenta Mathematica*. 2012, extra volume, s. 155–167.
- [33] SHIMBEL, A. Structure in communication nets. In *Proceedings of the Symposium on Information Networks (New York, 1954)*, s. 199–203, 1955.
- [34] SIGMOD, A. *SIGMOD Awards* [online]. [cit. 15.03.2013]. Dostupné z: <http://www.sigmod.org/sigmod-awards>.
- [35] TRIER, U. *DBLP Bibliography* [online]. [cit. 19.03.2013]. Dostupné z: <http://www.informatik.uni-trier.de/~ley/db>.
- [36] UNIVERSITY, T. P. S. *CiteSeerX* [online]. [cit. 23.03.2013]. Dostupné z: <http://citeseerx.ist.psu.edu>.
- [37] WELL, A. – MYERS, J. *Research Design & Statistical Analysis*. Taylor & Francis, 2002.
- [38] WHITE, D. R. – BORGATTI, S. P. Betweenness centrality measures for directed graphs. *Social Networks*. 1994, 16, s. 335–346.
- [39] WILLIAMS, G. *Linear Algebra With Applications*. Burlington, MA : Jones & Bartlett Learning, 2007.
- [40] ZWEIG, K. – IYENGAR, S. *An introduction to Centrality measures - An Introductory Course on Network Analysis* [online]. 2010. [cit. 20.04.2013]. Dostupné z: <https://sites.google.com/site/networkanalysisacourse/schedule/an-introduction-to-centrality-measures>.

A Tabulky korelací

Tabulka A.1: Tabulka korelací pro neizolované uzly pro DBLP

	hi	ideg	odeg	deg	wideg	wodeg	wdeg	pr	btw	btwA	wBtwA
hi	-	0.571	0.252	0.508	0.596	0.258	0.525	0.533	0.517	0.513	0.520
ideg	0.571	-	0.173	0.669	0.988	0.178	0.669	0.909	0.880	0.873	0.830
odeg	0.252	0.173	-	0.670	0.186	0.998	0.662	0.102	0.121	0.131	0.134
deg	0.508	0.669	0.670	-	0.673	0.670	0.992	0.570	0.566	0.573	0.529
wideg	0.596	0.988	0.186	0.673	-	0.192	0.687	0.898	0.877	0.870	0.849
wodeg	0.258	0.178	0.998	0.670	0.192	-	0.665	0.108	0.126	0.136	0.140
wdeg	0.525	0.669	0.662	0.992	0.687	0.665	-	0.570	0.573	0.579	0.549
pr	0.533	0.909	0.102	0.570	0.898	0.108	0.570	-	0.818	0.810	0.754
btw	0.517	0.880	0.121	0.566	0.877	0.126	0.573	0.818	-	0.993	0.929
btwA	0.513	0.873	0.131	0.573	0.870	0.136	0.579	0.810	0.993	-	0.924
wBtwA	0.520	0.830	0.134	0.529	0.849	0.140	0.549	0.754	0.929	0.924	-

Tabulka A.2: Tabulka korelací pro hlavní komponentu pro DBLP. Zahrnuty i variace closeness

	hi	ideg	odeg	deg	wideg	wodeg	wdeg	pr	btw	btwA	wBtwA	ic	oc	wic
hi	-	0.718	0.551	0.698	0.752	0.579	0.721	0.674	0.657	0.656	0.679	0.626	0.486	0.674
ideg	0.718	-	0.509	0.813	0.984	0.527	0.804	0.940	0.918	0.917	0.865	0.890	0.467	0.844
odeg	0.551	0.509	-	0.877	0.532	0.980	0.865	0.389	0.458	0.457	0.505	0.435	0.892	0.493
deg	0.698	0.813	0.877	-	0.821	0.873	0.984	0.708	0.747	0.747	0.753	0.720	0.783	0.734
wideg	0.752	0.984	0.532	0.821	-	0.557	0.831	0.922	0.905	0.904	0.889	0.871	0.487	0.887
wodeg	0.579	0.527	0.980	0.873	0.557	-	0.889	0.410	0.475	0.475	0.529	0.453	0.884	0.527
wdeg	0.721	0.804	0.865	0.984	0.831	0.889	-	0.701	0.740	0.739	0.766	0.712	0.780	0.763
pr	0.674	0.940	0.389	0.708	0.922	0.410	0.701	-	0.894	0.894	0.815	0.891	0.340	0.803
btw	0.657	0.918	0.458	0.747	0.905	0.475	0.740	0.894	-	0.998	0.918	0.961	0.422	0.866
btwA	0.656	0.917	0.457	0.747	0.904	0.475	0.739	0.894	0.998	-	0.917	0.960	0.421	0.864
wBtwA	0.679	0.865	0.505	0.753	0.889	0.529	0.766	0.815	0.918	0.917	-	0.876	0.463	0.930
ic	0.626	0.890	0.435	0.720	0.871	0.453	0.712	0.891	0.961	0.960	0.876	-	0.412	0.862
oc	0.486	0.467	0.892	0.783	0.487	0.884	0.780	0.340	0.422	0.421	0.463	0.412	-	0.472
wic	0.674	0.844	0.493	0.734	0.887	0.527	0.763	0.803	0.866	0.864	0.930	0.862	0.472	-

Tabulka A.3: Tabulka korelací pro neizolované uzly pro CiteSeer

	hi	ideg	odeg	deg	wideg	wodeg	wdeg	pr	btw	btwA	wBtwA
hi	-	0.770	0.315	0.589	0.789	0.354	0.623	0.738	0.731	0.744	0.738
ideg	0.770	-	0.314	0.693	0.993	0.344	0.704	0.954	0.928	0.954	0.929
odeg	0.315	0.314	-	0.833	0.320	0.981	0.807	0.220	0.319	0.297	0.312
deg	0.589	0.693	0.833	-	0.690	0.833	0.983	0.602	0.660	0.651	0.650
wideg	0.789	0.993	0.320	0.690	-	0.354	0.714	0.950	0.928	0.954	0.938
wodeg	0.354	0.344	0.981	0.833	0.354	-	0.834	0.257	0.346	0.327	0.343
wdeg	0.623	0.704	0.807	0.983	0.714	0.834	-	0.620	0.674	0.667	0.673
pr	0.738	0.954	0.220	0.602	0.950	0.257	0.620	-	0.886	0.920	0.886
btw	0.731	0.928	0.319	0.660	0.928	0.346	0.674	0.886	-	0.980	0.957
btwA	0.744	0.954	0.297	0.651	0.954	0.327	0.667	0.920	0.980	-	0.960
wBtwA	0.738	0.929	0.312	0.650	0.938	0.343	0.673	0.886	0.957	0.960	-

Tabulka A.4: Tabulka korelací pro hlavní komponentu pro CiteSeer. Zahrnutý i variace closeness

	hi	ideg	odeg	deg	wideg	wodeg	wdeg	pr	btw	btwA	wBtwA	ic	oc	wic
hi	-	0.706	0.452	0.623	0.749	0.503	0.663	0.652	0.605	0.608	0.629	0.562	0.379	0.650
ideg	0.706	-	0.484	0.825	0.973	0.505	0.815	0.868	0.823	0.829	0.790	0.783	0.420	0.775
odeg	0.452	0.484	-	0.849	0.491	0.966	0.821	0.294	0.416	0.416	0.423	0.457	0.841	0.457
deg	0.623	0.825	0.849	-	0.815	0.838	0.972	0.640	0.698	0.700	0.682	0.710	0.729	0.701
wideg	0.749	0.973	0.491	0.815	-	0.528	0.841	0.853	0.817	0.822	0.830	0.768	0.426	0.840
wodeg	0.503	0.505	0.966	0.838	0.528	-	0.859	0.336	0.435	0.435	0.456	0.465	0.806	0.493
wdeg	0.663	0.815	0.821	0.972	0.841	0.859	-	0.649	0.697	0.700	0.713	0.699	0.703	0.746
pr	0.652	0.868	0.294	0.640	0.853	0.336	0.649	-	0.743	0.748	0.699	0.694	0.207	0.680
btw	0.605	0.823	0.416	0.698	0.817	0.435	0.697	0.743	-	0.996	0.879	0.907	0.421	0.797
btwA	0.608	0.829	0.416	0.700	0.822	0.435	0.700	0.748	0.996	-	0.883	0.907	0.420	0.799
wBtwA	0.629	0.790	0.423	0.682	0.830	0.456	0.713	0.699	0.879	0.883	-	0.777	0.392	0.831
ic	0.562	0.783	0.457	0.710	0.768	0.465	0.699	0.694	0.907	0.907	0.777	-	0.510	0.827
oc	0.379	0.420	0.841	0.729	0.426	0.806	0.703	0.207	0.421	0.420	0.392	0.510	-	0.479
wic	0.650	0.775	0.457	0.701	0.840	0.493	0.746	0.680	0.797	0.799	0.831	0.827	0.479	-

B Porovnání pořadí implementovaných metod

B.1 DBLP

Tabulka B.1: Top 30 autorů DBLP podle H-indexu a pořadí podle ostatních metod

	hi	ideg	odeg	deg	wideg	wodeg	wdeg	pr	btw	btwA	wBtwA
MICHAEL STONEBRAKER	1	1	7	1	1	6	2	2	2	2	1
DAVID J. DEWITT	2	2	5	2	2	3	1	14	3	3	2
JEFFREY D. ULLMAN	3	5	65	9	3	48	5	12	9	9	4
PHILIP A. BERNSTEIN	4	7	119	10	8	87	13	6	1	1	7
RAKESH AGRAWAL	5	14	3	5	10	9	7	40	27	24	19
WON KIM	6	6	39	6	11	12	8	21	11	12	42
CATRIEL BEERI	7	28	66	36	15	57	24	37	18	20	23
UMESHWAR DAYAL	8	10	36	8	16	39	16	27	4	5	47
SERGE ABITEBOUL	9	16	18	11	9	7	6	53	22	23	30
YEHOSHUA SAGIV	10	34	125	47	14	61	21	47	46	48	11
MICHAEL J. CAREY	11	9	4	4	5	1	3	36	12	8	5
CHRISTOS FALOUTSOS	12	36	15	22	18	11	10	87	101	90	53
NATHAN GOODMAN	13	20	71	27	23	47	27	23	16	16	15
JIM GRAY	14	3	279	7	4	256	11	3	6	4	3
JEFFREY F. NAUGHTON	15	25	35	24	22	23	17	70	67	58	22
HECTOR GARCIA-MOLINA	16	11	2	3	7	5	4	32	21	19	16
RONALD FAGIN	17	42	378	73	29	207	57	25	20	21	51
DAVID MAIER	18	12	84	15	12	82	23	31	13	14	14
HAMID PIRAHESH	19	22	33	17	24	19	15	57	33	28	20
RAGHU RAMAKRISHNAN	20	27	8	13	17	10	9	73	47	45	25
BRUCE G. LINDSAY	21	21	38	18	25	32	29	41	23	22	24
JENNIFER WIDOM	22	26	29	23	19	30	19	67	57	57	18
C. MOHAN	23	47	61	44	33	31	34	80	71	68	13
YANNIS E. IOANNIDIS	24	63	9	29	34	17	25	123	81	72	40
RAYMOND A. LORIE	25	4	854	14	6	550	20	5	8	6	6
SHAMKANT B. NAVATHE	26	30	12	16	44	25	35	54	42	49	154
RICHARD HULL	27	44	26	32	31	18	26	78	55	59	46
FRANCCEILOIS BANCILHON	28	19	245	42	21	212	43	45	15	15	41
ARIE SHOSHANI	29	104	280	132	80	199	110	60	120	124	137
ALBERTO O. MENDELZON	30	68	34	43	43	40	38	81	90	95	43
suma	465	756	2980	733	586	2144	658	1330	1041	1017	923

Tabulka B.2: Top 30 autorů největší komponenty DBLP podle H-indexu a pořadí podle ostatních metod

	hi	ideg	odeg	deg	wideg	wodeg	wdeg	pr	btw	btwA	wBtwA	ic	oc	wic
MICHAEL STONEBRAKER	1	1	7	1	1	6	2	2	2	2	1	1	24	1
AVID J. DEWITT	2	2	5	2	2	3	1	14	3	3	2	3	13	4
JEFFREY D. ULLMAN	3	5	65	9	3	48	5	12	9	9	4	5	58	5
PHILIP A. BERNSTEIN	4	7	119	10	8	87	13	6	1	1	7	6	217	7
RAKESH AGRAWAL	5	14	3	5	10	9	7	38	27	24	16	22	2	25
WON KIM	6	6	39	6	11	12	8	21	11	12	43	9	47	49
CATRIEL BEERI	7	28	66	36	15	57	24	35	18	20	22	24	41	17
UMESHWAR DAYAL	8	10	36	8	16	39	16	26	4	5	48	10	23	37
SERGE ABITEBOUL	9	16	18	11	9	7	6	50	22	23	32	39	14	51
YEHOSHUA SAGIV	10	34	125	47	14	61	21	44	46	48	11	38	97	14
MICHAEL J. CAREY	11	9	4	4	5	1	3	34	12	8	5	11	9	8
CHRISTOS FALOUTSOS	12	36	15	22	18	11	10	84	101	90	45	70	38	83
NATHAN GOODMAN	13	20	71	27	23	47	27	23	16	16	15	15	166	13
JIM GRAY	14	3	279	7	4	256	11	3	6	4	3	2	798	2
JEFFREY F. NAUGHTON	15	25	35	24	22	23	17	67	67	58	20	46	53	20
HECTOR GARCIA-MOLINA	16	11	2	3	7	5	4	31	21	19	18	18	4	15
RONALD FAGIN	17	42	378	73	29	207	57	25	20	21	50	23	701	29
DAVID MAIER	18	12	84	15	12	82	23	30	13	14	14	13	109	9
HAMID PIRAHESH	19	22	33	17	24	19	15	54	33	28	23	32	28	26
RAGHU RAMAKRISHNAN	20	27	8	13	17	10	9	70	47	45	24	52	11	22
BRUCE G. LINDSAY	21	21	38	18	25	32	29	39	23	22	25	20	50	18
JENNIFER WIDOM	22	26	29	23	19	30	19	64	57	57	19	55	19	21
C. MOHAN	23	47	61	44	33	31	34	77	71	68	13	67	116	27
YANNIS E. IOANNIDIS	24	63	9	29	34	17	25	117	81	72	35	85	6	40
RAYMOND A. LORIE	25	4	852	14	6	550	20	5	8	6	6	4	2024	6
SHAMKANT B. NAVATHE	26	30	12	16	44	25	35	51	42	49	150	34	8	327
RICHARD HULL	27	44	26	32	31	18	26	75	55	59	51	73	22	64
FRANCESCO BANCILHON	28	19	245	42	21	212	43	42	15	15	41	19	294	34
ARIE SHOSHANI	29	104	280	132	80	199	110	57	120	124	143	102	281	232
ALBERTO O. MENDELZON	30	68	34	43	43	40	38	78	90	95	42	90	30	33
suma	465	756	2978	733	586	2144	658	1274	1041	1017	923	988	5303	1239

B.2 CiteSeer

Tabulka B.3: Top 30 autorů CiteSeer podle H-indexu a pořadí podle ostatních metod

	hi	ideg	odeg	deg	wideg	wodeg	wdeg	pr	btw	btwA	wBtwA
SCOTT SHENKER	1	2	29	4	1	10	2	7	2	2	3
DEBORAH ESTRIN	2	7	35	7	3	8	4	20	9	9	2
KEN KENNEDY	3	44	404	53	21	244	25	47	20	21	183
DOUGLAS C. SCHMIDT	4	103	92	74	25	4	13	335	1068	1013	1125
DON TOWSLEY	5	13	11	10	8	5	5	58	16	15	12
HECTOR GARCIA-MOLINA	6	14	17	14	15	18	15	66	39	41	21
THOMAS A. HENZINGER	7	46	275	47	13	26	14	95	100	97	55
JENNIFER WIDOM	8	15	109	16	9	43	12	49	8	8	18
RAKESH AGRAWAL	9	10	398	15	4	534	10	19	23	24	39
M. FRANS KAASHOEK	10	6	18	6	6	23	8	44	1	1	9
HUI ZHANG	11	22	163	23	22	153	24	61	25	25	14
WILLY ZWAENEPOEL	12	20	54	18	14	46	17	112	150	161	49
LUCA CARDELLI	13	61	1806	104	67	1532	134	56	114	113	1059
IAN FOSTER	14	12	21	13	7	15	6	77	45	44	76
SALLY FLOYD	15	5	521	9	2	293	7	10	6	5	5
SERGE ABITEBOUL	16	53	637	64	23	165	26	119	18	18	41
SENIOR MEMBER	17	1	1	1	12	1	1	4	3	3	17
MONI NAOR	18	92	231	93	38	65	34	104	69	68	177
BART SELMAN	19	85	1091	122	31	813	59	127	26	27	193
DAVID J. DEWITT	20	38	641	48	28	719	52	33	172	184	276
SEBASTIAN THRUN	21	78	49	54	32	6	18	192	144	149	168
OREN ETZIONI	22	75	110	62	97	351	122	162	36	35	162
DAPHNE KOLLER	23	140	101	103	99	131	94	244	29	31	120
RAJEEV ALUR	24	77	312	83	29	102	29	156	204	200	64
HARI BALAKRISHNAN	25	9	23	8	10	14	9	64	7	7	8
ROBERT HARPER	26	380	306	276	118	74	86	564	2260	2219	832
MAURIZIO LENZERINI	27	713	680	580	124	40	69	1435	3710	3625	197
ODED GOLDBREICH	28	481	1033	493	100	73	77	302	1102	1144	198
MARK D. HILL	29	99	211	97	89	174	95	195	334	338	361
HENRY M. LEVY	30	26	59	22	45	114	44	62	146	151	212
suma	465	2727	9438	2519	1092	5796	1111	4819	9886	9778	5696

Tabulka B.4: Top 30 autorů největší komponenty CiteSeer podle H-indexu a pořadí podle ostatních metod

	hi	ideg	odeg	deg	wideg	wodeg	wdeg	pr	btw	btwA	wBtwA	ic	oc	wic
SCOTT SHENKER	1	2	29	4	1	10	2	7	2	2	3	3	104	1
DEBORAH ESTRIN	2	7	35	7	3	8	4	15	9	9	2	8	35	4
KEN KENNEDY	3	41	404	53	21	244	25	41	20	21	183	79	1272	426
DOUGLAS C. SCHMIDT	4	100	92	74	25	4	13	319	1034	979	1106	235	89	1261
DON TOWSLEY	5	13	11	10	8	5	5	52	16	15	12	26	21	15
HECTOR GARCIA-MOLINA	6	14	17	14	15	18	15	59	39	41	21	9	59	29
THOMAS A. HENZINGER	7	43	275	47	13	26	14	87	97	94	55	92	920	191
JENNIFER WIDOM	8	15	109	16	9	43	12	43	8	8	18	20	375	27
RAKESH AGRAWAL	9	10	398	15	4	534	10	14	23	24	39	6	127	18
M. FRANS KAASHOEK	10	6	18	6	6	23	8	38	1	1	9	4	244	6
HUI ZHANG	11	22	163	23	22	153	24	55	25	25	14	36	242	13
WILLY ZWAENEPOEL	12	20	54	18	14	46	17	100	136	147	49	49	377	62
LUCA CARDELLI	13	58	1798	101	67	1532	134	50	101	100	1041	70	1758	1150
IAN FOSTER	14	12	21	13	7	15	6	70	45	44	76	17	147	96
SALLY FLOYD	15	5	520	9	2	293	7	9	6	5	5	13	638	2
SERGE ABITEBOUL	16	50	636	64	23	165	26	107	18	18	41	62	1491	43
SENIOR MEMBER	17	1	1	1	12	1	1	4	3	3	17	1	1	236
MONI NAOR	18	89	231	90	38	65	34	93	66	65	177	72	939	111
BART SELMAN	19	82	1089	119	31	813	59	115	26	27	193	57	1604	232
DAVID J. DEWITT	20	35	640	48	28	719	52	27	158	170	274	46	1147	73
SEBASTIAN THRUN	21	75	49	54	32	6	18	178	130	135	168	108	30	432
OREN ETZIONI	22	72	110	62	97	351	122	150	36	35	162	30	142	268
DAPHNE KOLLER	23	137	101	100	99	131	94	230	29	31	120	37	114	129
RAJEEV ALUR	24	74	312	80	29	102	29	144	190	186	64	177	282	235
HARI BALAKRISHNAN	25	9	23	8	10	14	9	57	7	7	8	7	155	7
ROBERT HARPER	26	366	306	273	118	74	86	534	2175	2136	814	454	853	732
MAURIZIO LENZERINI	27	698	679	567	124	40	69	1372	3549	3469	197	1955	1610	385
ODED GOLDBREICH	28	467	1031	480	100	73	77	286	1067	1109	198	479	3936	155
MARK D. HILL	29	96	211	94	89	174	95	181	320	324	358	107	376	211
HENRY M. LEVY	30	26	59	22	45	114	44	56	132	137	212	41	494	76
suma	465	2645	9422	2472	1092	5796	1111	4493	9468	9367	5636	4300	19582	6626

C Žebříčky významných autorů

Příloha C obsahuje tabulky pro každou implementovanou metodu s prvními třiceti autory a vyznačenými oceněními (Codd - ACM SIGMOD Edgar F. Codd Innovations Award, Fellows - ACM Fellows, Turing - ACM A.M. Turing Award, ISI - ISI Highly Cited highlighted), pokud ji autor obdržel.

C.1 DBLP

Tabulka C.1: Top 30 autorů podle metody hi

	Autor	hi	Turing	Codd	Fellows	ISI
1	MICHAEL STONEBRAKER	27.000		•	•	
2	DAVID J. DEWITT	25.000		•	•	
3	JEFFREY D. ULLMAN	24.000		•		
4	PHILIP A. BERNSTEIN	21.000		•		
5	RAKESH AGRAWAL	20.000		•		
6	WON KIM	20.000				
7	CATRIEL BEERI	20.000			•	•
8	UMESHWAR DAYAL	20.000		•		
9	SERGE ABITEBOUL	19.000		•	•	•
10	YEHOASHUA SAGIV	19.000				
11	MICHAEL J. CAREY	19.000		•		
12	CHRISTOS FALOUTSOS	19.000				
13	NATHAN GOODMAN	18.000				
14	JIM GRAY	18.000		•		
15	JEFFREY F. NAUGHTON	18.000				
16	HECTOR GARCIA-MOLINA	18.000		•		
17	RONALD FAGIN	18.000		•	•	•
18	DAVID MAIER	17.000		•		
19	HAMID PIRAHESH	17.000			•	
20	RAGHU RAMAKRISHNAN	17.000				
21	BRUCE G. LINDSAY	17.000				
22	JENNIFER WIDOM	17.000		•	•	
23	C. MOHAN	16.000		•		
24	YANNIS E. IOANNIDIS	16.000				
25	RAYMOND A. LORIE	16.000				
26	SHAMKANT B. NAVATHE	15.000				
27	RICHARD HULL	15.000			•	
28	FRANCCEDILOIS BANCILHON	15.000				
29	ARIE SHOSHANI	15.000				
30	ALBERTO O. MENDELZON	15.000				

Tabulka C.2: Top 30 autorů podle metody ideg

	Autor	ideg	Turing	Codd	Fellows	ISI
1	MICHAEL STONEBRAKER	1909.000		•	•	
2	DAVID J. DEWITT	1484.000		•	•	
3	JIM GRAY	1400.000		•		
4	RAYMOND A. LORIE	1276.000				
5	JEFFREY D. ULLMAN	1180.000		•		
6	WON KIM	1146.000				
7	PHILIP A. BERNSTEIN	1145.000		•		
8	E. F. CODD	1110.000		•		
9	MICHAEL J. CAREY	1110.000		•		
10	UMESHWAR DAYAL	1076.000		•		
11	HECTOR GARCIA-MOLINA	1020.000		•		
12	DAVID MAIER	1017.000		•		
13	DONALD D. CHAMBERLIN	966.000		•	•	
14	RAKESH AGRAWAL	907.000		•		
15	PETER P. CHEN	906.000				
16	SERGE ABITEBOUL	848.000		•	•	•
17	KAPALI P. ESWARAN	847.000				
18	MORTON M. ASTRAHAN	846.000				
19	FRANCCEDILOIS BANCILHON	840.000				
20	NATHAN GOODMAN	819.000				
21	BRUCE G. LINDSAY	806.000				
22	HAMID PIRAHESH	803.000			•	
23	IRVING L. TRAIGER	785.000			•	
24	EUGENE WONG	762.000				
25	JEFFREY F. NAUGHTON	729.000				
26	JENNIFER WIDOM	727.000		•	•	
27	RAGHU RAMAKRISHNAN	724.000				
28	CATRIEL BEERI	722.000			•	•
29	NICK ROUSSOPOULOS	702.000				
30	SHAMKANT B. NAVATHE	694.000				

Tabulka C.3: Top 30 autorů podle metody odeg

	Autor	odeg	Turing	Codd	Fellows	ISI
1	GERHARD WEIKUM	872.000			•	
2	HECTOR GARCIA-MOLINA	856.000		•		
3	RAKESH AGRAWAL	761.000		•		
4	MICHAEL J. CAREY	758.000		•		
5	DAVID J. DEWITT	758.000		•	•	
6	H. V. JAGADISH	717.000			•	
7	MICHAEL STONEBRAKER	677.000		•	•	
8	RAGHU RAMAKRISHNAN	652.000				
9	YANNIS E. IOANNIDIS	649.000				
10	ABRAHAM SILBERSCHATZ	636.000				
11	ELISA BERTINO	635.000			•	
12	SHAMKANT B. NAVATHE	629.000				
13	PHILIP S. YU	622.000				
14	STEFANO CERI	611.000				
15	CHRISTOS FALOUTSOS	607.000				
16	MATTHIAS JARKE	586.000				
17	GULTEKIN OUMLZSOYOGLU	582.000				
18	SERGE ABITEBOUL	575.000		•	•	•
19	NICK ROUSSOPOULOS	568.000				
20	MIRON LIVNY	559.000				
21	STANLEY Y. W. SU	558.000				
22	HANS-JOUMLRG SCHEK	557.000			•	
23	PATRICK VALDURIEZ	547.000				
24	GOETZ GRAEFE	546.000				
25	CLEMENT T. YU	542.000				
26	RICHARD HULL	537.000			•	
27	MICHAEL J. FRANKLIN	526.000				
28	RICHARD T. SNODGRASS	513.000			•	
29	JENNIFER WIDOM	510.000		•	•	
30	DENNIS SHASHA	508.000				

Tabulka C.4: Top 30 autorů podle metody deg

	Autor	deg	Turing	Codd	Fellows	ISI
1	MICHAEL STONEBRAKER	2586.000		•	•	
2	DAVID J. DEWITT	2242.000		•	•	
3	HECTOR GARCIA-MOLINA	1876.000		•		
4	MICHAEL J. CAREY	1868.000		•		
5	RAKESH AGRAWAL	1668.000		•		
6	WON KIM	1626.000				
7	JIM GRAY	1582.000		•		
8	UMESHWAR DAYAL	1562.000		•		
9	JEFFREY D. ULLMAN	1554.000		•		
10	PHILIP A. BERNSTEIN	1440.000		•		
11	SERGE ABITEBOUL	1423.000		•	•	•
12	H. V. JAGADISH	1411.000			•	
13	RAGHU RAMAKRISHNAN	1376.000				
14	RAYMOND A. LORIE	1364.000				
15	DAVID MAIER	1357.000		•		
16	SHAMKANT B. NAVATHE	1323.000				
17	HAMID PIRAHESH	1302.000			•	
18	BRUCE G. LINDSAY	1287.000				
19	ABRAHAM SILBERSCHATZ	1287.000				
20	GERHARD WEIKUM	1280.000			•	
21	NICK ROUSSOPOULOS	1270.000				
22	CHRISTOS FALOUTSOS	1264.000				
23	JENNIFER WIDOM	1237.000		•	•	
24	JEFFREY F. NAUGHTON	1216.000				
25	STEFANO CERI	1214.000				
26	PATRICK VALDURIEZ	1192.000				
27	NATHAN GOODMAN	1184.000				
28	DONALD D. CHAMBERLIN	1181.000		•	•	
29	YANNIS E. IOANNIDIS	1179.000				
30	MIRON LIVNY	1178.000				

Tabulka C.5: Top 30 autorů podle metody wideg

	Autor	wideg	Turing	Codd	Fellows	ISI
1	MICHAEL STONEBRAKER	5946.000		•	•	
2	DAVID J. DEWITT	5733.000		•	•	
3	JEFFREY D. ULLMAN	4429.000		•		
4	JIM GRAY	3982.000		•		
5	MICHAEL J. CAREY	3583.000		•		
6	RAYMOND A. LORIE	3501.000				
7	HECTOR GARCIA-MOLINA	3275.000		•		
8	PHILIP A. BERNSTEIN	3225.000		•		
9	SERGE ABITEBOUL	3177.000		•	•	•
10	RAKESH AGRAWAL	3152.000		•		
11	WON KIM	2993.000				
12	DAVID MAIER	2772.000		•		
13	E. F. CODD	2736.000		•		
14	YEHOASHUA SAGIV	2575.000				
15	CATRIEL BEERI	2491.000			•	•
16	UMESHWAR DAYAL	2465.000		•		
17	RAGHU RAMAKRISHNAN	2426.000				
18	CHRISTOS FALOUTSOS	2413.000				
19	JENNIFER WIDOM	2354.000		•	•	
20	DONALD D. CHAMBERLIN	2269.000		•	•	
21	FRANCCEDILOIS BANCILHON	2264.000				
22	JEFFREY F. NAUGHTON	2186.000				
23	NATHAN GOODMAN	2176.000				
24	HAMID PIRAHESH	2135.000			•	
25	BRUCE G. LINDSAY	2013.000				
26	MORTON M. ASTRAHAN	1985.000				
27	IRVING L. TRAIGER	1820.000			•	
28	ABRAHAM SILBERSCHATZ	1791.000				
29	RONALD FAGIN	1773.000		•	•	•
30	EUGENE WONG	1764.000				

Tabulka C.6: Top 30 autorů podle metody wodeg

	Autor	wodeg	Turing	Codd	Fellows	ISI
1	MICHAEL J. CAREY	3239.000		•		
2	GERHARD WEIKUM	3071.000			•	
3	DAVID J. DEWITT	2818.000		•	•	
4	PHILIP S. YU	2614.000				
5	HECTOR GARCIA-MOLINA	2512.000		•		
6	MICHAEL STONEBRAKER	2316.000		•	•	
7	SERGE ABITEBOUL	2297.000		•	•	•
8	H. V. JAGADISH	2263.000			•	
9	RAKESH AGRAWAL	2240.000		•		
10	RAGHU RAMAKRISHNAN	2059.000				
11	CHRISTOS FALOUTSOS	2042.000				
12	WON KIM	1902.000				
13	ABRAHAM SILBERSCHATZ	1867.000				
14	MIRON LIVNY	1806.000				
15	GOETZ GRAEFE	1789.000				
16	STEFANO CERI	1775.000				
17	YANNIS E. IOANNIDIS	1775.000				
18	RICHARD HULL	1692.000			•	
19	HAMID PIRAHESH	1685.000			•	
20	HANS-JOUMLRG SCHEK	1661.000			•	
21	STANLEY Y. W. SU	1651.000				
22	CLEMENT T. YU	1630.000				
23	JEFFREY F. NAUGHTON	1587.000				
24	RICHARD T. SNODGRASS	1558.000			•	
25	SHAMKANT B. NAVATHE	1538.000				
26	ELISA BERTINO	1500.000			•	
27	ALON Y. LEVY	1487.000				
28	MICHAEL J. FRANKLIN	1454.000				
29	NICK ROUSSOPOULOS	1406.000				
30	JENNIFER WIDOM	1396.000		•	•	

Tabulka C.7: Top 30 autorů podle metody wdeg

	Autor	wdeg	Turing	Codd	Fellows	ISI
1	DAVID J. DEWITT	8551.000		•	•	
2	MICHAEL STONEBRAKER	8262.000		•	•	
3	MICHAEL J. CAREY	6822.000		•		
4	HECTOR GARCIA-MOLINA	5787.000		•		
5	JEFFREY D. ULLMAN	5643.000		•		
6	SERGE ABITEBOUL	5474.000		•	•	•
7	RAKESH AGRAWAL	5392.000		•		
8	WON KIM	4895.000				
9	RAGHU RAMAKRISHNAN	4485.000				
10	CHRISTOS FALOUTSOS	4455.000				
11	JIM GRAY	4374.000		•		
12	GERHARD WEIKUM	4193.000			•	
13	PHILIP A. BERNSTEIN	4091.000		•		
14	H. V. JAGADISH	3991.000			•	
15	HAMID PIRAHESH	3820.000			•	
16	UMESHWAR DAYAL	3778.000		•		
17	JEFFREY F. NAUGHTON	3773.000				
18	PHILIP S. YU	3765.000				
19	JENNIFER WIDOM	3750.000		•	•	
20	RAYMOND A. LORIE	3723.000				
21	YEHOSHUA SAGIV	3672.000				
22	ABRAHAM SILBERSCHATZ	3658.000				
23	DAVID MAIER	3657.000		•		
24	CATRIEL BEERI	3621.000			•	•
25	YANNIS E. IOANNIDIS	3470.000				
26	RICHARD HULL	3451.000			•	
27	NATHAN GOODMAN	3399.000				
28	MIRON LIVNY	3374.000				
29	BRUCE G. LINDSAY	3371.000				
30	GOETZ GRAEFE	3265.000				

Tabulka C.8: Top 30 autorů podle metody pr

	Autor	pr	Turing	Codd	Fellows	ISI
1	E. F. CODD	179.324		•		
2	MICHAEL STONEBRAKER	137.371		•	•	
3	JIM GRAY	115.364		•		
4	DONALD D. CHAMBERLIN	114.010		•	•	
5	RAYMOND A. LORIE	107.204				
6	PHILIP A. BERNSTEIN	99.575		•		
7	MORTON M. ASTRAHAN	87.673				
8	KAPALI P. ESWARAN	87.167				
9	PETER P. CHEN	84.098				
10	IRVING L. TRAIGER	79.313			•	
11	JOHN MILES SMITH	78.833				
12	JEFFREY D. ULLMAN	74.323		•		
13	EUGENE WONG	68.319				
14	DAVID J. DEWITT	67.701		•	•	
15	MIKE W. BLASGEN	62.185				
16	GIANFRANCO R. PUTZOLU	61.585				
17	BRADFORD W. WADE	60.731				
18	RUDOLF BAYER	60.706		•		
19	JAMES W. MEHL	58.499				
20	PATRICIA P. GRIFFITHS	58.215				
21	WON KIM	57.946				
22	W. FRANK KING III	57.169				
23	NATHAN GOODMAN	56.791				
24	PAUL R. MCJONES	55.967			•	
25	RONALD FAGIN	54.766		•	•	•
26	RAYMOND F. BOYCE	54.475				
27	UMESHWAR DAYAL	54.099		•		
28	DIANE C. P. SMITH	53.677				
29	VERA WATSON	53.085				
30	MICHAEL HAMMER	52.687				

Tabulka C.9: Top 30 autorů podle metody btw

	Autor	btw	Turing	Codd	Fellows	ISI
1	PHILIP A. BERNSTEIN	62655703.293		•		
2	MICHAEL STONEBRAKER	61738362.921		•	•	
3	DAVID J. DEWITT	60335509.092		•	•	
4	JIM GRAY	58452724.132		•		
5	UMESHWAR DAYAL	58105048.655		•		
6	RAYMOND A. LORIE	57606842.228				
7	DONALD D. CHAMBERLIN	57435250.431		•	•	
8	MICHAEL J. CAREY	56191915.811		•		
9	JEFFREY D. ULLMAN	56098986.122		•		
10	KAPALI P. ESWARAN	55953909.624				
11	E. F. CODD	55595773.178		•		
12	WON KIM	55485910.707				
13	MORTON M. ASTRAHAN	53967137.730				
14	DAVID MAIER	53884993.441		•		
15	FRANCCEDILOIS BANCILHON	52436978.786				
16	NATHAN GOODMAN	51776071.388				
17	EUGENE WONG	50457002.386				
18	IRVING L. TRAIGER	50067735.663			•	
19	HECTOR GARCIA-MOLINA	49279794.248		•		
20	CATRIEL BEERI	49031169.516			•	•
21	RONALD FAGIN	48476621.189		•	•	•
22	BRUCE G. LINDSAY	47956637.448				
23	SERGE ABITEBOUL	47196023.670		•	•	•
24	RAKESH AGRAWAL	46621125.945		•		
25	PATRICIA G. SELINGER	45312957.343		•	•	
26	THOMAS G. PRICE	44961579.565				
27	DENNIS MCLEOD	44846630.893				
28	HAMID PIRAHESH	44408421.808			•	
29	HENRY F. KORTH	44365555.952			•	
30	RANDY H. KATZ	44264843.771				

Tabulka C.10: Top 30 autorů podle metody btwA

	Autor	btwA	Turing	Codd	Fellows	ISI
1	PHILIP A. BERNSTEIN	159954377.951		•		
2	MICHAEL STONEBRAKER	154562285.765		•	•	
3	DAVID J. DEWITT	151226543.383		•	•	
4	UMESHWAR DAYAL	148199498.491		•		
5	DONALD D. CHAMBERLIN	146188880.058		•	•	
6	JIM GRAY	145464683.525		•		
7	E. F. CODD	144527193.507		•		
8	RAYMOND A. LORIE	143126778.632				
9	JEFFREY D. ULLMAN	142742025.957		•		
10	KAPALI P. ESWARAN	141829887.829				
11	WON KIM	141617051.174				
12	MICHAEL J. CAREY	139609158.987		•		
13	DAVID MAIER	136756549.030		•		
14	MORTON M. ASTRAHAN	135873540.410				
15	FRANCCEDILOIS BANCILHON	133327793.505				
16	NATHAN GOODMAN	131716143.987				
17	EUGENE WONG	127621249.542				
18	CATRIEL BEERI	126349161.825			•	•
19	IRVING L. TRAIGER	125392610.648			•	
20	RONALD FAGIN	124989300.213		•	•	•
21	HECTOR GARCIA-MOLINA	122564518.399		•		
22	SERGE ABITEBOUL	122057559.819		•	•	•
23	BRUCE G. LINDSAY	119996828.360				
24	DENNIS MCLEOD	118304803.594				
25	PETER P. CHEN	116123796.130				
26	JOHN MILES SMITH	115771141.296				
27	RAKESH AGRAWAL	115204381.342		•		
28	MICHAEL HAMMER	113451632.828				
29	NICK ROUSSOPOULOS	113024824.395				
30	PATRICIA G. SELINGER	112877895.367		•	•	

Tabulka C.11: Top 30 autorů podle metody wBtwA

	Autor	wBtwA	Turing	Codd	Fellows	ISI
1	MICHAEL STONEBRAKER	52317849.461		•	•	
2	DAVID J. DEWITT	48312413.783		•	•	
3	JIM GRAY	46557285.992		•		
4	JEFFREY D. ULLMAN	43551886.027		•		
5	MICHAEL J. CAREY	41234479.322		•		
6	RAYMOND A. LORIE	37835019.206				
7	PHILIP A. BERNSTEIN	37185666.397		•		
8	LAWRENCE A. ROWE	35451644.493				
9	EUGENE WONG	35190409.740				
10	MIRON LIVNY	34782881.586				
11	YEHOASHUA SAGIV	33417737.415				
12	DONALD D. CHAMBERLIN	33378865.916		•	•	
13	C. MOHAN	32875071.666		•		
14	NATHAN GOODMAN	32569687.965				
15	DAVID MAIER	32530494.480		•		
16	HECTOR GARCIA-MOLINA	31382496.051		•		
17	RAKESH AGRAWAL	31297218.188		•		
18	RANDY H. KATZ	31284072.254				
19	JENNIFER WIDOM	31154675.393		•	•	
20	E. F. CODD	30937491.000		•		
21	HAMID PIRAHESH	29513838.180			•	
22	JEFFREY F. NAUGHTON	29442109.817				
23	CATRIEL BEERI	29080749.828			•	•
24	BRUCE G. LINDSAY	28637340.939				
25	RAGHU RAMAKRISHNAN	28386450.924				
26	GOETZ GRAEFE	27502892.430				
27	MORTON M. ASTRAHAN	26910055.956				
28	IRVING L. TRAIGER	26615723.912			•	
29	LAURA M. HAAS	26607613.707				
30	DONOVAN A. SCHNEIDER	25170370.292				

Tabulka C.12: Top 30 autorů podle metody ic

	Autor	ic	Turing	Codd	Fellows	ISI
1	MICHAEL STONEBRAKER	0.593		•	•	
2	JIM GRAY	0.560		•		
3	DAVID J. DEWITT	0.556		•	•	
4	RAYMOND A. LORIE	0.556				
5	JEFFREY D. ULLMAN	0.546		•		
6	PHILIP A. BERNSTEIN	0.546		•		
7	E. F. CODD	0.543		•		
8	DONALD D. CHAMBERLIN	0.539		•	•	
9	WON KIM	0.537				
10	UMESHWAR DAYAL	0.535		•		
11	MICHAEL J. CAREY	0.532		•		
12	MORTON M. ASTRAHAN	0.531				
13	DAVID MAIER	0.529		•		
14	KAPALI P. ESWARAN	0.529				
15	NATHAN GOODMAN	0.527				
16	EUGENE WONG	0.526				
17	IRVING L. TRAIGER	0.525			•	
18	HECTOR GARCIA-MOLINA	0.523		•		
19	FRANCCEDILOIS BANCILHON	0.520				
20	BRUCE G. LINDSAY	0.519				
21	PETER P. CHEN	0.518				
22	RAKESH AGRAWAL	0.518		•		
23	RONALD FAGIN	0.517		•	•	•
24	CATRIEL BEERI	0.517			•	•
25	THOMAS G. PRICE	0.514				
26	PATRICIA G. SELINGER	0.514		•	•	
27	JOHN MILES SMITH	0.513				
28	MIKE W. BLASGEN	0.512				
29	RANDY H. KATZ	0.512				
30	GIO WIEDERHOLD	0.512				

Tabulka C.13: Top 30 autorů podle metody oc

	Autor	oc	Turing	Codd	Fellows	ISI
1	H. V. JAGADISH	0.475			•	
2	RAKESH AGRAWAL	0.473		•		
3	GERHARD WEIKUM	0.471			•	
4	HECTOR GARCIA-MOLINA	0.468		•		
5	GULTEKIN OUMLZSOYOGLU	0.467				
6	YANNIS E. IOANNIDIS	0.467				
7	STEFANO CERI	0.466				
8	SHAMKANT B. NAVATHE	0.466				
9	MICHAEL J. CAREY	0.465		•		
10	ELISA BERTINO	0.464			•	
11	RAGHU RAMAKRISHNAN	0.463				
12	RICHARD T. SNODGRASS	0.463			•	
13	DAVID J. DEWITT	0.462		•	•	
14	SERGE ABITEBOUL	0.462		•	•	•
15	CLEMENT T. YU	0.460				
16	GOETZ GRAEFE	0.460				
17	HANS-JOUMLRG SCHEK	0.459			•	
18	ABRAHAM SILBERSCHATZ	0.459				
19	JENNIFER WIDOM	0.459		•	•	
20	PATRICK VALDURIEZ	0.458				
21	NICK ROUSSOPOULOS	0.457				
22	RICHARD HULL	0.457			•	
23	UMESHWAR DAYAL	0.457		•		
24	MICHAEL STONEBRAKER	0.454		•	•	
25	DENNIS SHASHA	0.454				
26	MATTHIAS JARKE	0.453				
27	MIRON LIVNY	0.451				
28	HAMID PIRAHESH	0.451			•	
29	CHRISTIAN S. JENSEN	0.451				
30	ALBERTO O. MENDELZON	0.450				

Tabulka C.14: Top 30 autorů podle metody wic

	Autor	wic	Turing	Codd	Fellows	ISI
1	MICHAEL STONEBRAKER	2.055		•	•	
2	JIM GRAY	2.047		•		
3	E. F. CODD	2.043		•		
4	DAVID J. DEWITT	2.017		•	•	
5	JEFFREY D. ULLMAN	2.014		•		
6	RAYMOND A. LORIE	2.012				
7	PHILIP A. BERNSTEIN	2.011		•		
8	MICHAEL J. CAREY	1.993		•		
9	DAVID MAIER	1.979		•		
10	EUGENE WONG	1.973				
11	DONALD D. CHAMBERLIN	1.969		•	•	
12	LAWRENCE A. ROWE	1.967				
13	NATHAN GOODMAN	1.966				
14	YEHOASHUA SAGIV	1.966				
15	HECTOR GARCIA-MOLINA	1.960		•		
16	IRVING L. TRAIGER	1.955			•	
17	CATRIEL BEERI	1.949			•	•
18	BRUCE G. LINDSAY	1.944				
19	MORTON M. ASTRAHAN	1.943				
20	JEFFREY F. NAUGHTON	1.942				
21	JENNIFER WIDOM	1.937		•	•	
22	RAGHU RAMAKRISHNAN	1.936				
23	MIRON LIVNY	1.936				
24	RANDY H. KATZ	1.934				
25	RAKESH AGRAWAL	1.933		•		
26	HAMID PIRAHESH	1.933			•	
27	C. MOHAN	1.921		•		
28	DONOVAN A. SCHNEIDER	1.910				
29	RONALD FAGIN	1.910		•	•	•
30	LAURA M. HAAS	1.906				

C.2 CiteSeer

Tabulka C.15: Top 30 autorů podle metody hi

	Autor	hi	Turing	Codd	Fellows	ISI
1	SCOTT SHENKER	37.000				
2	DEBORAH ESTRIN	34.000			•	
3	KEN KENNEDY	33.000				
4	DOUGLAS C. SCHMIDT	33.000				
5	DON TOWSLEY	32.000				
6	HECTOR GARCIA-MOLINA	31.000		•		
7	THOMAS A. HENZINGER	31.000				
8	JENNIFER WIDOM	30.000		•		
9	RAKESH AGRAWAL	29.000		•		
10	M. FRANS KAASHOEK	29.000				
11	HUI ZHANG	28.000				
12	WILLY ZWAENEPOEL	28.000				
13	LUCA CARDELLI	27.000			•	
14	IAN FOSTER	27.000				
15	SALLY FLOYD	26.000				
16	SERGE ABITEBOUL	26.000		•		
17	SENIOR MEMBER	26.000				
18	MONI NAOR	26.000				
19	BART SELMAN	26.000				
20	DAVID J. DEWITT	25.000		•	•	
21	SEBASTIAN THRUN	25.000				
22	OREN ETZIONI	24.000				
23	DAPHNE KOLLER	24.000				
24	RAJEEV ALUR	24.000			•	•
25	HARI BALAKRISHNAN	24.000				
26	ROBERT HARPER	24.000				
27	MAURIZIO LENZERINI	24.000				
28	ODED GOLDBREICH	24.000				
29	MARK D. HILL	23.000				
30	HENRY M. LEVY	23.000				

Tabulka C.16: Top 30 autorů podle metody ideg

	Autor	ideg	Turing	Codd	Fellows	ISI
1	SENIOR MEMBER	5024.000				
2	SCOTT SHENKER	4798.000				
3	VAN JACOBSON	4112.000				
4	STUDENT MEMBER	4082.000				
5	SALLY FLOYD	3963.000				
6	M. FRANS KAASHOEK	3854.000				
7	DEBORAH ESTRIN	3747.000			•	
8	LIXIA ZHANG	3626.000				
9	HARI BALAKRISHNAN	3495.000				
10	RAKESH AGRAWAL	3463.000		•		
11	JOHN K. OUSTERHOUT	3401.000				
12	IAN FOSTER	3135.000				
13	DON TOWSLEY	3087.000				
14	HECTOR GARCIA-MOLINA	2917.000		•		
15	JENNIFER WIDOM	2880.000		•		
16	STEVEN MCCANNE	2781.000				
17	FACHBEREICH INFORMATIK	2645.000				
18	ROBERT MORRIS	2625.000				
19	VERN PAXSON	2480.000				
20	WILLY ZWAENEPOL	2478.000				
21	CARL KESSELMAN	2453.000				
22	HUI ZHANG	2449.000				
23	TAKEO KANADE	2397.000				
24	RANDAL E. BRYANT	2370.000				
25	THORSTEN VON EICKEN	2358.000				
26	HENRY M. LEVY	2282.000				
27	DAVID CULLER	2241.000				
28	THOMAS E. ANDERSON	2240.000				
29	DAVID E. CULLER	2231.000				
30	RAMAKRISHNAN SRIKANT	2224.000				

Tabulka C.17: Top 30 autorů podle metody odeg

	Autor	odeg	Turing	Codd	Fellows	ISI
1	SENIOR MEMBER	6846.000				
2	STUDENT MEMBER	6055.000				
3	PH. D	4143.000				
4	FACHBEREICH INFORMATIK	2413.000				
5	PROF DR	1943.000				
6	ARTHUR C. SMITH	1848.000				
7	GRADUATE SCHOOL—NEW BRUNSWICK	1620.000				
8	KARSTEN SCHWAN	1437.000				
9	KANG G. SHIN	1364.000				
10	KLARA NAHRSTEDT	1324.000				
11	DON TOWSLEY	1260.000				
12	IEEE COMPUTER SOCIETY	1251.000				
13	COPYRIGHT STICHTING	1225.000				
14	MATHEMATISCH CENTRUM	1225.000				
15	JARMO T. ALANDER	1217.000				
16	TZI-CKER CHIUEH	1097.000				
17	HECTOR GARCIA-MOLINA	1082.000		•		
18	M. FRANS KAASHOEK	1066.000				
19	AMIN VAHDAT	1066.000				
20	ASSOCIATE MEMBER	1063.000				
21	IAN FOSTER	1055.000				
22	INRIA ROCQUENCOURT	1048.000				
23	HARI BALAKRISHNAN	1047.000				
24	JOHN A. STANKOVIC	1030.000				
25	DEPARTMENT CHAIR	1014.000				
26	CAMBRIDGE CB FD	1009.000				
27	DAVID KOTZ	1007.000				
28	CALTON PU	1001.000				
29	SCOTT SHENKER	1000.000				
30	MARIO GERLA	999.000				

Tabulka C.18: Top 30 autorů podle metody deg

	Autor	deg	Turing	Codd	Fellows	ISI
1	SENIOR MEMBER	11870.000				
2	STUDENT MEMBER	10137.000				
3	PH. D	5969.000				
4	SCOTT SHENKER	5798.000				
5	FACHBEREICH INFORMATIK	5058.000				
6	M. FRANS KAASHOEK	4920.000				
7	DEBORAH ESTRIN	4699.000			•	
8	HARI BALAKRISHNAN	4542.000				
9	SALLY FLOYD	4372.000				
10	DON TOWSLEY	4347.000				
11	LIXIA ZHANG	4302.000				
12	VAN JACOBSON	4276.000				
13	IAN FOSTER	4190.000				
14	HECTOR GARCIA-MOLINA	3999.000		•		
15	RAKESH AGRAWAL	3919.000		•		
16	JENNIFER WIDOM	3537.000		•		
17	JOHN K. OUSTERHOUT	3531.000				
18	WILLY ZWAENEPOEL	3265.000				
19	ROBERT MORRIS	3232.000				
20	STEVEN MCCANNE	3204.000				
21	DAVID CULLER	3140.000				
22	HENRY M. LEVY	3058.000				
23	HUI ZHANG	3045.000				
24	CARL KESSELMAN	3034.000				
25	TAKEO KANADE	2976.000				
26	JACK DONGARRA	2925.000				
27	KANG G. SHIN	2924.000				
28	CHRISTOS FALOUTSOS	2895.000				
29	DAVID E. CULLER	2870.000				
30	VERN PAXSON	2820.000				

Tabulka C.19: Top 30 autorů podle metody wideg

	Autor	wideg	Turing	Codd	Fellows	ISI
1	SCOTT SHENKER	16274.000				
2	SALLY FLOYD	13866.000				
3	DEBORAH ESTRIN	13416.000			•	
4	RAKESH AGRAWAL	13302.000		•		
5	VAN JACOBSON	13281.000				
6	M. FRANS KAASHOEK	11422.000				
7	IAN FOSTER	11393.000				
8	DON TOWSLEY	10956.000				
9	JENNIFER WIDOM	10582.000		•		
10	HARI BALAKRISHNAN	10189.000				
11	LIXIA ZHANG	10182.000				
12	SENIOR MEMBER	9568.000				
13	THOMAS A. HENZINGER	9484.000				
14	WILLY ZWAENEPOEL	9121.000				
15	HECTOR GARCIA-MOLINA	9066.000		•		
16	CARL KESSELMAN	8492.000				
17	STEVEN MCCANNE	8308.000				
18	STUDENT MEMBER	8154.000				
19	ROBERT MORRIS	7920.000				
20	VERN PAXSON	7751.000				
21	KEN KENNEDY	7722.000				
22	HUI ZHANG	7423.000				
23	SERGE ABITEBOUL	7197.000		•		
24	RAMAKRISHNAN SRIKANT	6711.000				
25	DOUGLAS C. SCHMIDT	6705.000				
26	RANDAL E. BRYANT	6680.000				
27	DAVID B. JOHNSON	6491.000				
28	DAVID J. DEWITT	6323.000		•	•	
29	RAJEEV ALUR	6286.000			•	•
30	JOHN K. OUSTERHOUT	6124.000				

Tabulka C.20: Top 30 autorů podle metody wodeg

	Autor	wodeg	Turing	Codd	Fellows	ISI
1	SENIOR MEMBER	13741.000				
2	STUDENT MEMBER	12104.000				
3	PH. D	7010.000				
4	DOUGLAS C. SCHMIDT	6893.000				
5	DON TOWSLEY	5951.000				
6	SEBASTIAN THRUN	5890.000				
7	KARSTEN SCHWAN	5760.000				
8	DEBORAH ESTRIN	5737.000			•	
9	KLARA NAHRSTEDT	5552.000				
10	SCOTT SHENKER	5263.000				
11	FACHBEREICH INFORMATIK	5180.000				
12	KANG G. SHIN	5087.000				
13	THOMAS EITER	4807.000				
14	HARI BALAKRISHNAN	4728.000				
15	IAN FOSTER	4699.000				
16	AZER BESTAVROS	4277.000				
17	AMIN VAHDAT	4216.000				
18	HECTOR GARCIA-MOLINA	4210.000		•		
19	MARTIN RINARD	4197.000				
20	WOLFRAM BURGARD	4053.000				
21	JARMO T. ALANDER	3975.000				
22	MARIO GERLA	3972.000				
23	M. FRANS KAASHOEK	3933.000				
24	JOHN HEIDEMANN	3878.000				
25	JIAWEI HAN	3863.000				
26	THOMAS A. HENZINGER	3798.000				
27	IAN HORROCKS	3739.000				
28	ANGELOS D. KEROMYTIS	3719.000				
29	GEORGE KARYPIS	3672.000				
30	DIETER FOX	3664.000				

Tabulka C.21: Top 30 autorů podle metody wdeg

	Autor	wdeg	Turing	Codd	Fellows	ISI
1	SENIOR MEMBER	23309.000				
2	SCOTT SHENKER	21537.000				
3	STUDENT MEMBER	20258.000				
4	DEBORAH ESTRIN	19153.000			•	
5	DON TOWSLEY	16907.000				
6	IAN FOSTER	16092.000				
7	SALLY FLOYD	15556.000				
8	M. FRANS KAASHOEK	15355.000				
9	HARI BALAKRISHNAN	14917.000				
10	RAKESH AGRAWAL	14579.000		•		
11	VAN JACOBSON	13839.000				
12	JENNIFER WIDOM	13696.000		•		
13	DOUGLAS C. SCHMIDT	13598.000				
14	THOMAS A. HENZINGER	13282.000				
15	HECTOR GARCIA-MOLINA	13276.000		•		
16	LIXIA ZHANG	12818.000				
17	WILLY ZWAENEPOEL	12164.000				
18	SEBASTIAN THRUN	11838.000				
19	FACHBEREICH INFORMATIK	10918.000				
20	CARL KESSELMAN	10851.000				
21	ROBERT MORRIS	10156.000				
22	STEVEN MCCANNE	10133.000				
23	PH. D	10024.000				
24	HUI ZHANG	9575.000				
25	KEN KENNEDY	9536.000				
26	SERGE ABITEBOUL	9286.000		•		
27	KANG G. SHIN	9013.000				
28	VERN PAXSON	8976.000				
29	RAJEEV ALUR	8725.000			•	•
30	RAMESH GOVINDAN	8704.000				

Tabulka C.22: Top 30 autorů podle metody pr

	Autor	pr	Turing	Codd	Fellows	ISI
1	JOHN K. OUSTERHOUT	413.275				
2	MARTIN E. HELLMAN	336.552				
3	WHITFIELD DIFFIE	289.814				
4	SENIOR MEMBER	280.899				
5	JACK J. DONGARRA	279.157				
6	VAN JACOBSON	259.762				
7	SCOTT SHENKER	225.241				
8	S. KENT	224.260				
9	RANDAL E. BRYANT	197.141				
10	SALLY FLOYD	196.267				
11	LIXIA ZHANG	194.574				
12	STUDENT MEMBER	182.926				
13	S. KIRKPATRICK	181.416				
14	C. D. GELATT	181.416				
15	M. P. VECCHI	181.416				
16	TAKEO KANADE	178.883				
17	RANDOLPH BENTSON	178.869				
18	GEORGE W. FURNAS	177.145				
19	RAKESH AGRAWAL	175.358		•		
20	DEBORAH ESTRIN	173.030			•	
21	STEPHEN C. JOHNSON	168.657				
22	EDWARD H. ADELSON	162.659				
23	KEN THOMPSON	159.405				
24	ADI SHAMIR	155.899				
25	MICHAEL J. KARELS	153.567				
26	DENNIS M. RITCHIE	152.912				
27	H. ADELSON	149.576				
28	BUTLER W. LAMPSON	149.033				
29	MICHAEL BURROWS	148.711				
30	DAVID HAREL	147.132				

Tabulka C.23: Top 30 autorů podle metody btw

	Autor	btw	Turing	Codd	Fellows	ISI
1	M. FRANS KAASHOEK	10112159061.329				
2	SCOTT SHENKER	9785892051.378				
3	SENIOR MEMBER	8845140725.908				
4	VAN JACOBSON	8813158813.752				
5	SALLY FLOYD	8690842977.232				
6	LARRY L. PETERSON	8630281410.114				
7	HARI BALAKRISHNAN	8544868651.847				
8	JENNIFER WIDOM	8512314665.857		•		
9	DEBORAH ESTRIN	8414557973.610			•	
10	MONICA S. LAM	8394649393.784				
11	LIXIA ZHANG	8350916122.774				
12	STEVEN MCCANNE	8263572085.250				
13	M. SATYANARAYANAN	8087193503.970				
14	THOMAS E. ANDERSON	8078380316.694				
15	DON TOWSLEY	8053219720.882				
16	JOHN K. OUSTERHOUT	8039121635.247				
17	PETER B. DANZIG	7986359949.438				
18	SERGE ABITEBOUL	7924035872.337		•		
19	CHRISTOS FALOUTSOS	7915737482.408				
20	STUDENT MEMBER	7854847262.676				
21	KEN KENNEDY	7846470147.904				
22	Y H. KATZ	7746528995.356				
23	DAVID B. JOHNSON	7657508456.644				
24	RAKESH AGRAWAL	7615283090.820		•		
25	HUI ZHANG	7588067468.401				
26	VERN PAXSON	7489340159.080				
27	BART SELMAN	7483765335.090				
28	JOSEPH M. HELLERSTEIN	7439523002.678				
29	CHAU-WEN TSENG	7394297185.197				
30	MICHAEL J. KARELS	7379419075.434				

Tabulka C.24: Top 30 autorů podle metody btwA

	Autor	btwA	Turing	Codd	Fellows	ISI
1	M. FRANS KAASHOEK	10490039542.119				
2	SCOTT SHENKER	10163084814.946				
3	SENIOR MEMBER	9176669511.323				
4	VAN JACOBSON	9121102934.920				
5	LARRY L. PETERSON	8974774762.984				
6	SALLY FLOYD	8957931516.810				
7	HARI BALAKRISHNAN	8851040160.073				
8	JENNIFER WIDOM	8802268069.570		•		
9	DEBORAH ESTRIN	8724682679.922			•	
10	MONICA S. LAM	8715062909.044				
11	LIXIA ZHANG	8670366072.416				
12	STEVEN MCCANNE	8584234209.524				
13	THOMAS E. ANDERSON	8432183908.739				
14	M. SATYANARAYANAN	8414467674.058				
15	JOHN K. OUSTERHOUT	8354169106.912				
16	DON TOWSLEY	8339250719.641				
17	PETER B. DANZIG	8321261780.421				
18	SERGE ABITEBOUL	8259168008.370		•		
19	CHRISTOS FALOUTSOS	8185711166.805				
20	KEN KENNEDY	8105163842.735				
21	STUDENT MEMBER	8098909193.573				
22	Y H. KATZ	7997677289.242				
23	RAKESH AGRAWAL	7930247980.534		•		
24	DAVID B. JOHNSON	7906188409.970				
25	HUI ZHANG	7853246606.979				
26	BART SELMAN	7776288153.848				
27	JOSEPH M. HELLERSTEIN	7774964587.035				
28	VERN PAXSON	7755479916.742				
29	DAPHNE KOLLER	7689235456.579				
30	CHAU-WEN TSENG	7683843108.991				

Tabulka C.25: Top 30 autorů podle metody wBtwA

	Autor	wBtwA	Turing	Codd	Fellows	ISI
1	RAMESH GOVINDAN	9211451555.875				
2	DEBORAH ESTRIN	9164755029.757			•	
3	SCOTT SHENKER	9053073939.621				
4	JOHN HEIDEMANN	7964028583.955				
5	SALLY FLOYD	7526125219.850				
6	LIXIA ZHANG	7014651851.876				
7	VAN JACOBSON	6832457164.379				
8	HARI BALAKRISHNAN	6810856958.887				
9	M. FRANS KAASHOEK	6586757433.120				
10	STEVEN MCCANNE	6457726203.586				
11	ROBERT MORRIS	6345222543.733				
12	DON TOWSLEY	5988132213.598				
13	JIM KUROSE	5530165637.113				
14	HUI ZHANG	5178930791.300				
15	CHING-GUNG LIU	5115755173.821				
16	DAVID CULLER	5061015727.524				
17	SENIOR MEMBER	4997968885.661				
18	JENNIFER WIDOM	4856418323.023		•		
19	MARK H	4856146508.601				
20	STUDENT MEMBER	4850477255.558				
21	HECTOR GARCIA-MOLINA	4849921599.732		•		
22	LEWIS GIROD	4845180273.052				
23	VICTOR FIROIU	4792278150.290				
24	JEREMY ELSON	4729078672.706				
25	DOMENICO FERRARI	4718328412.856				
26	ION STOICA	4639232859.234				
27	VERN PAXSON	4611622099.486				
28	JITENDRA PADHYE	4423979043.162				
29	HAOBO YU	4416107685.632				
30	ANDREW S. TANENBAUM	4390683564.792				

Tabulka C.26: Top 30 autorů podle metody ic

	Autor	ic	Turing	Codd	Fellows	ISI
1	SENIOR MEMBER	0.393				
2	JOHN K. OUSTERHOUT	0.392				
3	SCOTT SHENKER	0.384				
4	M. FRANS KAASHOEK	0.383				
5	STUDENT MEMBER	0.380				
6	RAKESH AGRAWAL	0.380		•		
7	HARI BALAKRISHNAN	0.377				
8	DEBORAH ESTRIN	0.377			•	
9	HECTOR GARCIA-MOLINA	0.376		•		
10	FACHBEREICH INFORMATIK	0.375				
11	VAN JACOBSON	0.375				
12	RAJEEV MOTWANI	0.374			•	
13	SALLY FLOYD	0.373				
14	DAVID CULLER	0.370				
15	LIXIA ZHANG	0.370				
16	CHRISTOS FALOUTSOS	0.370				
17	IAN FOSTER	0.370				
18	STEVEN MCCANNE	0.370				
19	PRABHAKAR RAGHAVAN	0.369				
20	JENNIFER WIDOM	0.369		•		
21	ROBERT E. SCHAPIRE	0.368				
22	ROBERT MORRIS	0.368				
23	M. SATYANARAYANAN	0.368				
24	PETER B. DANZIG	0.367				
25	VERN PAXSON	0.367				
26	DON TOWSLEY	0.367				
27	RICHARD KARP	0.367				
28	THOMAS E. ANDERSON	0.366				
29	ERIC A. BREWER	0.366				
30	OREN ETZIONI	0.366				

Tabulka C.27: Top 30 autorů podle metody oc

	Autor	oc	Turing	Codd	Fellows	ISI
1	SENIOR MEMBER	0.393				
2	STUDENT MEMBER	0.388				
3	PH. D	0.378				
4	PROF DR	0.357				
5	FACHBEREICH INFORMATIK	0.357				
6	ARTHUR C. SMITH	0.351				
7	GRADUATE SCHOOL—NEW BRUNSWICK	0.351				
8	KARSTEN SCHWAN	0.339				
9	THME RSEAUX ET SYSTMES	0.339				
10	IEEE COMPUTER SOCIETY	0.338				
11	COPYRIGHT STICHTING	0.337				
12	MATHEMATISCH CENTRUM	0.337				
13	SUPERVISOR PROF	0.336				
14	KANG G. SHIN	0.335				
15	CAMBRIDGE CB FD	0.334				
16	DEPARTMENT CHAIR	0.334				
17	KLARA NAHRSTEDT	0.334				
18	INRIA ROCQUENCOURT	0.333				
19	ASSOCIATE MEMBER	0.333				
20	TZI-CKER CHIUEH	0.332				
21	DON TOWSLEY	0.331				
22	BONGKI MOON	0.331				
23	JOHN A. STANKOVIC	0.331				
24	JOEL SALTZ	0.331				
25	MIODRAG POTKONJAK	0.330				
26	COMMITTEE CHAIR	0.330				
27	KRITHI RAMAMRITHAM	0.330				
28	MING LI	0.330				
29	SHARAD MEHROTRA	0.329				
30	SEBASTIAN THRUN	0.328				

Tabulka C.28: Top 30 autorů podle metody wic

	Autor	wic	Turing	Codd	Fellows	ISI
1	SCOTT SHENKER	1.319				
2	SALLY FLOYD	1.315				
3	VAN JACOBSON	1.312				
4	DEBORAH ESTRIN	1.312			•	
5	LIXIA ZHANG	1.309				
6	M. FRANS KAASHOEK	1.307				
7	HARI BALAKRISHNAN	1.306				
8	ROBERT MORRIS	1.306				
9	STEVEN MCCANNE	1.306				
10	RAMESH GOVINDAN	1.304				
11	JOHN HEIDEMANN	1.303				
12	ION STOICA	1.302				
13	HUI ZHANG	1.297				
14	VERN PAXSON	1.294				
15	DON TOWSLEY	1.292				
16	DAVID KARGER	1.291				
17	DOMENICO FERRARI	1.290				
18	RAKESH AGRAWAL	1.290		•		
19	SYLVIA RATNASAMY	1.290				
20	JOHN KUBIATOWICZ	1.289				
21	CHING-GUNG LIU	1.287				
22	DAVID CULLER	1.287				
23	ANDREW S. TANENBAUM	1.285				
24	SRINIVASAN SESHAN	1.285				
25	LEWIS GIROD	1.285				
26	MARK HANDLEY	1.284				
27	JENNIFER WIDOM	1.284		•		
28	JEREMY ELSON	1.284				
29	HECTOR GARCIA-MOLINA	1.283		•		
30	RICHARD KARP	1.282				