

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

Bakalářská práce

Měření významnosti autorů v citační síti

Plzeň, 2013

Tomáš Maršálek

Abstrakt

Prvky sociální sítě, které nemají žádné apriorní ohodnocení významnosti, jsou různě významné pouze na základě vztahů s okolními prvky. V této práci byly prozkoumány a implementovány známé metody centrality a bibliografické metody měřící významnost prvků v sociální nebo citační síti. Výsledky aplikování metod na volně dostupné citační databáze ukázaly vysokou podobnost jednotlivých metod a rovněž shodu nejvýznamnějších autorů dle těchto metod se známými oceněními v oblasti informatiky a informační vědy (Turing Award, Codd, ACM Fellows, ISI Highly Cited). Bylo zjištěno, že některé implementované metody jsou i přes použití nejrychlejších algoritmů výpočetně příliš náročné vzhledem k velikosti citačních sítí, vzniklých z těchto citačních databází. Dále bylo empiricky potvrzeno, že implementované metody měří významnost, byť může mít více interpretací.

Obsah

1	Úvod	5
1.1	Sociální sítě	5
1.1.1	Analýza sociálních sítí	5
1.2	Citační sítě	5
1.2.1	Síť publikací	5
1.2.2	Síť autorů	6
1.2.3	Vážené citační sítě	6
1.2.4	Orientované a neorientované sítě	7
1.2.5	Komponenty grafu	9
1.2.6	Klika v grafu	9
1.3	Analýza citačních sítí	9
1.4	Významnost autorů	9
1.5	Míry centrality	9
1.5.1	Radius	9
1.5.2	Degree	9
1.5.3	Eigenvector	9
1.5.4	Closeness	9
1.5.5	Betweenness	9
1.6	Ostatní používané míry významnosti autorů	9
1.6.1	H-index	9
1.6.2	Impact factor	9
1.7	Metody porovnání	9

1.7.1	Spearmanův koeficient pořadové korelace	9
1.7.2	Pearsonův korelační koeficient	9
1.8	Ocenění významných autorů	9
1.8.1	ACM A.M. Turing Award	9
1.8.2	ACM SIGMOD Edgar F. Codd Innovations Award . .	10
1.8.3	ACM Fellows	10
1.8.4	ISI Highly Cited highlighted	10
2	Implementace	11
2.1	Načtení vstupních dat	12
2.2	Vytvoření citačních sítí	12
2.2.1	Síť publikací	12
2.2.2	Síť autorů	12
2.3	Analýza struktury sítě	12
2.4	Reprezentace citační sítě	12
2.4.1	All-pair shortest path	12
2.4.2	Single source shortest path	12
2.5	Hledání nejkratších cest	12
2.6	Knihovna pro SNA	12
2.6.1	Radius	12
2.6.2	Degree	12
2.6.3	Eigenvector	12
2.6.4	Closeness	12
2.6.5	Betweenness	12
2.6.6	H-index	12
2.6.7	Paralelní výpočty	12
2.7	Citační databáze	12
2.7.1	DBLP	12
2.7.2	CiteSeer	13
3	Výsledky	14
3.1	Struktura sítě	15
3.1.1	DBLP	15
3.1.2	CiteSeer	15
3.2	Porovnání nejvýznamnějších autorů	15
3.3	Porovnání implementovaných metod	15
3.4	Žebříčky významných autorů	15
3.4.1	DBLP	15

3.4.2	CiteSeer	25
4	Diskuse	29
4.1	Podobnost výsledků jednotlivých metod	29
4.2	Shoda výsledků s oceněními	29
4.3	Vliv vah na přesnost výsledků	29
4.4	Vstupní a výstupní hrany	29
5	Závěr	30

1.1 Sociální sítě

1.1.1 Analýza sociálních sítí

1.2 Citační sítě

Citační sítě jsou podobné sociálním sítím, pouze místo uzlů, které představují osoby, v citační síti se 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.

1.2.1 Síť publikací

Uvažujeme-li první případ, kde uzly reprezentují publikace a hrany přímo citace mezi těmito publikacemi, jedná se o síť publikací. Tedy pokud publikace A odkazuje na publikaci B , pak 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".

1.2.2 Síť autorů

Druhým případem citační sítě je 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, tzn. pro jednu 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 této publikace a duplikujeme ji pro každého z autorů citované publikace. Celkově vznikne nm nových hran, pokud odkazovaná publikace obsahuje n autorů a odkazující m autorů. Stejně jako v síti publikací, i zde uvažujeme dvě opačné orientace hrany se stejnou interpretací, tedy "cituji" a "jsem citován".

V síti autorů má pro naše účely smysl uvažovat ohodnocení hran. Existuje více způsobů, jak přiřadit ohodnocení (váhy) jednotlivým hranám, ale nejjednodušším způsobem, který je použitý i v implementaci knihovny, je prosté přiřazení počtu publikací, jejichž autorem nebo spoluautorem je daný autor A , které odkazují na publikace, jejichž autorem je autor B . Srozumitelnější popis poskytne obrázek:

Druhým způsobem ohodnocení hran, který rovněž využívá implementovaná knihovna pro některé metody, je převrácená hodnota prvního způsobu ohodnocení. Důvodem je přímá souvislost mezi vahou hrany a vzdáleností mezi uzly. V prvním případě, kdy silnější pouto mezi autory vyjadřuje vyšší ohodnocení hrany, v druhém případě 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.

1.2.3 Vážené citační síť

V definici grafu nebo sítě $G = (V, E)$ 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.

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.

Například síť world wide web tvořená webovými stránkami je příkladem multigrafu, protože je povoleno z jedné stránky odkazovat na jinou na více místech. Při analýze takových 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 například síť kde sledujeme města a dopravní spojení mezi nimi. V tomhle 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 a nejrychlejší spojení.

Pro citační síť můžeme uvažovat ohodnocení hran obojího typu. Například mezi dvěma autory může být silnější vztah, pokud se citují ve více publikacích. Pokud 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.

1.2.4 Orientované a neorientované sítě

Obecně můžeme uvažovat grafy s hranami s orientací či bez orientace. V obou případech se stále jedná o množinu (V, E) , pouze pro orientovaný graf je množina hran množinou uspořádaných dvojic oproti množině neuspořádaných dvojic u neorientovaného grafu.

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 hrana.

V případě sociálních sítí nejčastěji uvažujeme síť 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.

1.2.5 Komponenty grafu

Silně spojené komponenty

1.2.6 Klika v grafu

1.3 Analýza citačních sítí

1.4 Významnost autorů

1.5 Míry centrality

1.5.1 Degree

Pro orientovaný graf můžeme uvažovat vstupní (indegree) a výstupní stupeň (outdegree) vrcholu nebo obecný stupeň (degree), tedy součet těchto dvou. Pokud uvažujeme pouze vstupní stupeň, vypočtená hodnota určuje významnost uzlu, kdežto výstupní stupeň ukazuje společenskost či otevřenost uzlu. Obecný stupeň znamená celkový počet spojení s okolními prvky sítě.

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

1.5.2 Eigenvector

Eigenvector centrality je míra vlivu uzlu v síti. Oproti jednoduché degree centrality, která použila pouze počet přímých spojení, tahle míra navíc používá významnost uzlů v těchto přímých spojeních. Pokud máme dva uzly se stejným počtem přímých spojení, hodnota degree centrality bude u obou stejná. Pokud je ale první uzel spojen s významnějšími uzly než druhý uzel, v metodě typu eigenvector centrality bude první uzel hodnocen výše. Eigenvector centralitu vypočteme pomocí numerických metod pro vlastní čísla, které vyřeší rovnici

$$\mathbf{Ax} = \lambda \mathbf{x} \quad (1.1)$$

kde \mathbf{A} je matice sousednosti (adjacency matrix), \mathbf{x} je vlastní vektor matice \mathbf{A} a je řešením této rovnice, která má více řešení. Ke každému řešení náleží

vlastní číslo λ . Podle Perron-Frobeniovy věty má pouze jedno řešení všechny prvky vlastního vektoru \mathbf{x} nezáporné [?]. Tohle řešení odpovídá největšímu vlastnímu číslu λ a jako jediné nás v problému eigenvector centrality zajímá.

Pro výpočet použijeme iterační metodu

$$x_u = \frac{1}{\lambda} \sum_{v \in G} \mathbf{A}_{uv} x_v \quad (1.2)$$

$$(1.3)$$

kde x_u je prvek vlastního vektoru \mathbf{x} , který odpovídá uzlu u . G je množina uzlů sítě a \mathbf{A}_{uv} je prvek matice sousednosti na řádku u a sloupci v , který je 1, pokud existuje přímé spojení mezi uzly u a v , jinak 0.

V tomto vyjádření eigenvector centrality zjistíme, že se jedná o přímé rozšíření degree centrality. Výsledek z předchozí iterace použijeme jako vstup do další iterace. Výpočet iterujeme, dokud nedosáhneme požadované přesnosti výsledku.

PageRank

1.5.3 Closeness

Algoritmus

Aproximace

1.5.4 Betweenness

Brandesův algoritmus

Aproximace

1.5.5 Radius

1.6 Ostatní používané míry významnosti autorů

1.6.1 H-index

1.6.2 Impact factor

1.7 Metody porovnání

1.7.1 Spearmanův koeficient pořadové korelace

1.7.2 Pearsonův korelační koeficient

1.8 Ocenění významných autorů

1.8.1 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. [?].

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

1.8.2 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 [?].

1.8.3 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 vzhlížejí jako k autoritám v době rozvoje informačních technologií [?].

1.8.4 ISI Highly Cited highlighted

ISI Highly Cited je databáze často citovaných autorů v člancích posledního desetiletí, které byly vydány 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 [?].

KAPITOLA 2

Implementace

2.1 Načtení vstupních dat

2.2 Vytvoření citačních sítí

2.2.1 Sít' publikací

2.2.2 Sít' autorů

2.3 Analýza struktury sítě

2.4 Reprezentace citační sítě

2.4.1 All-pair shortest path

2.4.2 Single source shortest path

2.5 Hledání nejkratších cest

2.6 Knihovna pro SNA

2.6.1 Radius

2.6.2 Degree

2.6.3 Eigenvector

2.6.4 Closeness

2.6.5 Betweenness

používáme verzi z roku 2004.

Charakteristika

2.7.2 CiteSeer

CiteSeer (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.

Charakteristika

KAPITOLA 3

Výsledky

3.1 Struktura sítě

3.1.1 DBLP

Rozdělení vah

3.1.2 CiteSeer

Rozdělení vah

3.2 Porovnání nejvýznamnějších autorů

3.3 Porovnání implementovaných metod

3.4 Žebříčky významných autorů

3.4.1 DBLP

H-index

	Autor	H-index	Turing	Codd	Fellows	ISI
1	MICHAEL STONEBRAKER	28		•	•	
2	DAVID J. DEWITT	24		•	•	
3	JEFFREY D. ULLMAN	24		•	•	•
4	PHILIP A. BERNSTEIN	22		•	•	•
5	RAKESH AGRAWAL	21		•	•	
6	WON KIM	21		•	•	
7	YEHOSHUA SAGIV	20				
8	CATRIEL BEERI	20			•	•

Nevážený indegree

	Autor	indegree	Turing	Codd	Fellows	ISI
1	MICHAEL STONEBRAKER	1909		•	•	
2	DAVID J. DEWITT	1484		•	•	
3	JIM GRAY	1400		•	•	
4	RAYMOND A. LORIE	1276			•	
5	JEFFREY D. ULLMAN	1180		•	•	•
6	WON KIM	1146			•	
7	PHILIP A. BERNSTEIN	1145		•	•	•
8	E. F. CODD	1110	•		•	
9	MICHAEL J. CAREY	1110		•	•	
10	UMESHWAR DAYAL	1076		•	•	
11	HECTOR GARCIA-MOLINA	1020		•	•	•
12	DAVID MAIER	1017		•	•	•
13	DONALD D. CHAMBERLIN	966		•	•	
14	RAKESH AGRAWAL	907		•	•	
15	PETER P. CHEN	906			•	
16	SERGE ABITEBOUL	848		•	•	•
17	KAPALI P. ESWARAN	847				
18	MORTON M. ASTRAHAN	846				
19	FRANCCEDILOIS BANCILHON	840				
20	NATHAN GOODMAN	819				•
21	BRUCE G. LINDSAY	806			•	
22	HAMID PIRAHESH	803			•	
23	IRVING L. TRAIGER	785			•	
24	EUGENE WONG	762				
25	JEFFREY F. NAUGHTON	729			•	

Nevážený outdegree

	Autor	indegree	Turing	Codd	Fellows	ISI
1	GERHARD WEIKUM	872			•	
2	HECTOR GARCIA-MOLINA	856		•	•	•
3	RAKESH AGRAWAL	761		•	•	
4	MICHAEL J. CAREY	758		•	•	
5	DAVID J. DEWITT	758		•	•	
6	H. V. JAGADISH	717			•	
7	MICHAEL STONEBRAKER	677		•	•	
8	RAGHU RAMAKRISHNAN	652			•	
9	YANNIS E. IOANNIDIS	649			•	
10	ABRAHAM SILBERSCHATZ	636			•	
11	ELISA BERTINO	635			•	
12	SHAMKANT B. NAVATHE	629				
13	PHILIP S. YU	622			•	•
14	STEFANO CERI	611				
15	CHRISTOS FALOUTSOS	607			•	
16	MATTHIAS JARKE	586				
17	GULTEKIN OUMLZSOYOGLU	582				
18	SERGE ABITEBOUL	575		•	•	•
19	NICK ROUSSOPOULOS	568			•	
20	MIRON LIVNY	559				
21	STANLEY Y. W. SU	558				
22	HANS-JOUMLRG SCHEK	557			•	
23	PATRICK VALDURIEZ	547			•	
24	GOETZ GRAEFE	546				
25	CLEMENT T. YU	542				

Vážený indegree

	Autor	indegree	Turing	Codd	Fellows	ISI
1	MICHAEL STONEBRAKER	5946		•	•	
2	DAVID J. DEWITT	5733		•	•	
3	JEFFREY D. ULLMAN	4429		•	•	•
4	JIM GRAY	3982		•	•	
5	MICHAEL J. CAREY	3583		•	•	
6	RAYMOND A. LORIE	3501			•	
7	HECTOR GARCIA-MOLINA	3275		•	•	•
8	PHILIP A. BERNSTEIN	3225		•	•	•
9	SERGE ABITEBOUL	3177		•	•	•
10	RAKESH AGRAWAL	3152		•	•	
11	WON KIM	2993			•	
12	DAVID MAIER	2772		•	•	•
13	E. F. CODD	2736	•		•	
14	YEHOSHUA SAGIV	2575				
15	CATRIEL BEERI	2491			•	•
16	UMESHWAR DAYAL	2465		•	•	
17	RAGHU RAMAKRISHNAN	2426			•	
18	CHRISTOS FALOUTSOS	2413			•	
19	JENNIFER WIDOM	2354		•	•	
20	DONALD D. CHAMBERLIN	2269		•	•	
21	FRANCCEDILOIS BANCILHON	2264				
22	JEFFREY F. NAUGHTON	2186			•	
23	NATHAN GOODMAN	2176				•
24	HAMID PIRAHESH	2135			•	
25	BRUCE G. LINDSAY	2013			•	

Vážený outdegree

	Autor	outdegree	Turing	Codd	Fellows	ISI
1	MICHAEL J. CAREY	3239		•	•	
2	GERHARD WEIKUM	3071			•	
3	DAVID J. DEWITT	2818		•	•	
4	PHILIP S. YU	2614			•	•
5	HECTOR GARCIA-MOLINA	2512		•	•	•
6	MICHAEL STONEBRAKER	2316		•	•	
7	SERGE ABITEBOUL	2297		•	•	•
8	H. V. JAGADISH	2263			•	
9	RAKESH AGRAWAL	2240		•	•	
10	RAGHU RAMAKRISHNAN	2059			•	
11	CHRISTOS FALOUTSOS	2042			•	
12	WON KIM	1902			•	
13	ABRAHAM SILBERSCHATZ	1867			•	
14	MIRON LIVNY	1806				
15	GOETZ GRAEFE	1789				
16	STEFANO CERI	1775				
17	YANNIS E. IOANNIDIS	1775			•	
18	RICHARD HULL	1692			•	
19	HAMID PIRAHESH	1685			•	
20	HANS-JOUMLRG SCHEK	1661			•	
21	STANLEY Y. W. SU	1651				
22	CLEMENT T. YU	1630				
23	JEFFREY F. NAUGHTON	1587			•	
24	RICHARD T. SNODGRASS	1558			•	
25	SHAMKANT B. NAVATHE	1538				

PageRank

Hodnoty PageRanku dosahují hodnot mezi 0 a 1. Pro účely přehlednosti byly v této tabulce normalizovány na interval 0 až $|V|$, tedy počet uzlů sítě.

	Autor	PageRank	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		•	•	•

Nevážený closeness

	Autor	Closeness	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				

Vážený closeness

	Autor	Closeness	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	YEHOSHUA 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		•	•	

Navážený betweeness

	Autor	Betweeness	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		•	•	

Vážený betweeness

	Autor	Betweeness	Turing	Codd	Fellows	ISI
1	MICHAEL STONEBRAKER	51920270.604		•	•	
2	DAVID J. DEWITT	47407633.796		•	•	
3	JIM GRAY	46202513.744		•	•	
4	JEFFREY D. ULLMAN	43255880.998		•	•	•
5	MICHAEL J. CAREY	40700171.932		•	•	
6	RAYMOND A. LORIE	37350006.114			•	
7	PHILIP A. BERNSTEIN	36361815.206		•	•	•
8	LAWRENCE A. ROWE	35035718.542		•		
9	EUGENE WONG	34499259.565				
10	MIRON LIVNY	34488016.860				
11	YEHOASHUA SAGIV	32954711.980				
12	DONALD D. CHAMBERLIN	32704585.298			•	
13	C. MOHAN	32417070.578		•	•	
14	DAVID MAIER	32132542.674		•	•	•
15	NATHAN GOODMAN	31719323.894		•		
16	HECTOR GARCIA-MOLINA	31080457.354		•	•	•
17	RANDY H. KATZ	30977276.737		•		
18	JENNIFER WIDOM	30793150.109		•	•	
19	RAKESH AGRAWAL	30763284.150		•	•	
20	E. F. CODD	29634524.870	•		•	
21	JEFFREY F. NAUGHTON	29267110.900			•	
22	HAMID PIRAHESH	29233790.884			•	
23	CATRIEL BEERI	28857797.364			•	•
24	BRUCE G. LINDSAY	28566231.162			•	
25	RAGHU RAMAKRISHNAN	28057226.481			•	

3.4.2 CiteSeer

PageRank

	Autor	PageRank	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				

Nevážený closeness

	Autor	Closeness	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			•	•

Nevážený betweeness

	Autor	Betweeness	Turing	Codd	Fellows	ISI
1	M. FRANS KAASHOEK	10112159061.330			•	
2	SCOTT SHENKER	9785892051.377			•	
3	SENIOR MEMBER	8845140725.909				
4	VAN JACOBSON	8813158813.753				
5	SALLY FLOYD	8690842977.231			•	•
6	LARRY L. PETERSON	8630281410.114			•	
7	HARI BALAKRISHNAN	8544868651.846			•	
8	JENNIFER WIDOM	8512314665.858		•	•	
9	DEBORAH ESTRIN	8414557973.610			•	
10	MONICA S. LAM	8394649393.784			•	
11	LIXIA ZHANG	8350916122.773			•	
12	STEVEN MCCANNE	8263572085.250				
13	M. SATYANARAYANAN	8087193503.971			•	
14	THOMAS E. ANDERSON	8078380316.694			•	
15	DON TOWSLEY	8053219720.880			•	•
16	JOHN K. OUSTERHOUT	8039121635.247			•	
17	PETER B. DANZIG	7986359949.439				
18	SERGE ABITEBOUL	7924035872.338		•	•	•
19	CHRISTOS FALOUTSOS	7915737482.408			•	
20	STUDENT MEMBER	7854847262.677				
21	KEN KENNEDY	7846470147.904			•	•
22	Y H. KATZ	7746528995.356			•	
23	DAVID B. JOHNSON	7657508456.644				
24	RAKESH AGRAWAL	7615283090.821		•	•	
25	HUI ZHANG	7588067468.401			•	

	Autor	H-index	Turing	Codd	Fellows	ISI
1	SCOTT SHENKER	37			•	
2	DEBORAH ESTRIN	34			•	
3	KEN KENNEDY	33			•	•
4	DOUGLAS C. SCHMIDT	33				
5	DON TOWSLEY	32			•	•
6	HECTOR GARCIA-MOLINA	31		•	•	•
7	THOMAS A. HENZINGER	30			•	•
8	RAKESH AGRAWAL	29		•	•	
9	M. FRANS KAASHOEK	29			•	
10	JENNIFER WIDOM	29		•	•	
11	WILLY ZWAENEPOEL	28			•	
12	HUI ZHANG	27			•	
13	IAN FOSTER	27			•	•
14	MONI NAOR	27				
15	SALLY FLOYD	26			•	•
16	LUCA CARDELLI	26			•	
17	SERGE ABITEBOUL	26		•	•	•
18	SENIOR MEMBER	26				
19	BART SELMAN	26			•	
20	SEBASTIAN THRUN	25				
21	OREN ETZIONI	24				
22	DAVID J. DEWITT	24		•	•	
23	DAPHNE KOLLER	24				
24	RAJEEV ALUR	24			•	•
25	HARI BALAKRISHNAN	24			•	

KAPITOLA 4

Diskuse

- 4.1 Podobnost výsledků jednotlivých metod
- 4.2 Shoda výsledků s oceněními
- 4.3 Vliv vah na přesnost výsledků
- 4.4 Vstupní a výstupní hrany

KAPITOLA 5

Závěr
