Západočeská univerzita v Plzni
Fakulta aplikovaných věd
Katedra infromatiky 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

KAPITOLA 1

Úvod

KAPITOLA 2

Sociální a citační sítě

2.1 Sociální 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. Nejčastěji, zejména v poslední době s rozmachem populárních webových sociálních sítí (MySpace, Facebook, G+, Lidé), jím bývá přátelství.

2.2 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 spojená 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.

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 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 $\mathbf{A}_{uv}$ na řádku u a sloupci v určuje jestli existuje hrana od vrcholu u do vrcholu v. Pokud je hodnota $\mathbf{A}_{uv}$ 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 $\mathbf{B}$. Incidenční matice vyjadřuje vztah mezi vrcholy a hranami tak, že $\mathbf{B}_{ue} = 1$, pokud vrchol u je spojený s hranou e, a 0 v opačném případě. V orientovaném grafu rozlišujeme mezi počátečním uzlem $\mathbf{B}_{ue} = -1$ a koncovým uzlem $\mathbf{B}_{ue} = 1$.

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 ale 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).

2.3 Analýza sociálních sítí

2.3.1 Bezškálové sítě

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. Tento jednoduchý způsob způsobí rozložení stupňů vrcholů podle Poissonova rozdělení 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, 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 v sociální síti s velkým počtem přátel má velikou šanci, že bude představen novým lidem a rozšíří si tak svůj kruh přátel ještě více. Tomuto jevu v bezškálových sítí se říká „bohatší se stává bohatším“.

2.4 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.

2.4.1 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".

2.4.2 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.

2.4.3 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. 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.

2.4.4 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.

2.4.5 Souvislost a komponenty grafu

Pro neorientovaný graf je komponenta maximálně souvislý podgraf. Jinak řečeno komponenta je podgraf takový, že všechny jeho vrcholy jsou spojeny nějakou cestou. Komponentou ji i samotný vrchol.

Všechny komponenty grafu najdeme pomocí jednoduchých algoritmů prohledávání do šířky nebo do hloubky. Spuštění prohledávání najde celou komponentu, ve které se výchozí vrchol nachází. Spustíme-li prohledávání ze všech vrcholů, najdeme všechny komponenty.

Slabě souvislý orientovaný graf znamená, že neorientovaný graf, který by vznikl nahrazením orientovaných hran neorientovanými (symetrizace grafu), by byl souvislý.

Pro zachování vlastnosti souvislosti, že všechny vrcholy jsou spojené nějakou cestou, pro orientovaný graf musíme uvažovat silně souvislý graf nebo podgraf. Definice zůstává stejná jako u slabě spojitých komponent, ale protože hrany nejsou oboustranné, mezi dvěma spojenými vrcholy ne vždy existuje cesta oběma směry. U neorientovaného grafu můžeme souvislost vyjádřit tak, že pro každé dva uzly u a v existuje cesta z u do v. Protože jsou hrany symetrické, pak automaticky existuje i cesta z v do u. U orientovaného grafu musíme druhou podmínku explicitně dodat: graf je silně souvislý, pokud pro každé dva vrcholy u a v existuje cesta z u do v i z v do u.

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

2.4.6 Klika v grafu

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

V sociologii slovo klika souvisí se skupinou lidí, kteří jsou na sebe vázáni více než na jiné lidi v tomtéž prostředí. Klika je silněji spojená skupina lidí než sociální kruh.

2.5 Analýza citačních sítí

2.6 Citační databáze

Citační databáze poskytují možnost vyhledávání bibliografických citací. Většina z dnešních citačních databází se zaměřuje na jeden obor. Full-textové databáze poskytují kompletní text publikací, které indexují (cite http://library.amnh.org/research-tools/citation-full-text-databases).

2.6.1 DBLP

DBLP [?] je webová bibliografická databáze hostovaná na univerzitě Trier. Od 80. let indexovala literaturu z oblasti databází a logického programovaní, 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.

Charakteristika

Struktura sítě

Rozďělení vah

2.6.2 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.

2.7 Ocenění významných autorů

2.7.1 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].

2.7.2 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 [?].

2.7.3 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 [?].

2.7.4 ISI Highly Cited highlighted

ISI Highly Cited je databáze často citovaných autorů v článcích 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 3

Významnost autorů

3.1 Míry centrality

3.1.1 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^+.

$$C_{Din}(u) = deg^-(u) = \sum_{v \in V} \mathbf{A}_{uv} \tag{3.1}$$

$$C_{Dout}(u) = deg^+(u) = \sum_{v \in V} \mathbf{A}_{vu} \tag{3.2}$$

$$C_D(v) = C_{Din} + C_{Dout} \tag{3.3}$$

Kde $\mathbf{A}$ je matice sousednosti grafu. 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 skok. Tento fakt je známý problém a důvod pro zavedení dalších a složitějších metod pro výpočet významnosti.

3.1.2 Eigenvector

Eigenvector centrality, také známá jako Gould's index of accessibility of a Network (Linear Algebra with Applications: Alternate Edition by Gareth Williams), je míra vlivu vrcholu v grafu. Hodnotu vlivu získáme z vlastního vektoru x matice sousednosti grafu:

$$\mathbf{A}\mathbf{x} = \lambda\mathbf{x} \tag{3.4}$$

$\mathbf{A}$ je matice sousednosti, $\mathbf{x}$ je vlastní vektor matice $\mathbf{A}$ a řešením této rovnice o více řešeních. 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 [**?**].

Rovnici můžeme rozepsat z maticového tvaru do jednotlivých složek:

$$x_u = \frac{1}{\lambda} \sum_{v \in G} \mathbf{A}_{uv} x_v \tag{3.5}$$

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 řádku u a sloupci v.

$$x_{ui+1} = \frac{1}{\lambda} \sum_{v \in G} \mathbf{A}_{uv} x_{vi} \tag{3.6}$$

V tomhle rekurentním tvaru je vidět předpis pro iterační výpočet eigenvector centrality. Algoritmus se nazývá mocninná 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 (3.1). 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 (cite něco). 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.

$$x_{ui+1} = \frac{1-d}{|V|} + d \sum_{v \in V} \mathbf{A}_{uv} \frac{x_{vi}}{deg^{+}(v)} \tag{3.7}$$

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 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 je často ze zkušenosti nastaven na 85%.

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

Jedním problémem algoritmu PageRank jsou uzly bez výstupních hran (dangling nodes). Protože musíme v každé iteraci algoritmu zachovat vlastnost rozdělení pravděpodobnosti, že suma všech pravděpodobností je 1, je třeba zajistit, aby se přenášená hodnota mezi iteracemi neztrácela právě v uzlech bez výstupních hran. Problém se nazývá rank sink a nejčastěji se řeší přidáním zdroje PageRanku:

$$x_{ui+1} = \frac{1-d}{|V|} + d \sum_{v \in V} \mathbf{A}_{uv} \frac{x_{vi}}{deg^{+}(v)} + \frac{1}{|D|} \sum_{w \in D} x_{wi} \tag{3.8}$$

V každé iteraci předem vypočítáme součet hodnot PageRanku, které by se ztratily v uzlech bez výstupních hran. Tahle hodnota je v rámci iterace konstantní a pouze ji rovnoměrně rozdělíme mezi uzly sítě (tedy s váhou $1/|D|$, kde D je množina uzlů bez výstupních hran (dangling nodes).

Přestože je PageRank původně určený pro webovou síť, lze ho použít na jakoukoliv orientovanou váženou i neváženou síť, tedy i na sociální a citační sítě, o kterých je zde řeč. Pro neorientovaný graf je hodnota PageRanku pro jednotlivé uzly velmi blízká stupňům grafu, ale ne totožná (cite icola Perra and Santo Fortunato.; Fortunato (September 2008). "Spectral centrality measures in complex networks")

3.1.3 Míry založené na nejkratších cestách

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 matematických pracích.

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 closeness a betweeness.

Nechť 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, které spojují vrcholy u a v.

3.1.4 Closeness

Closeness neboli blízkost je definována jako převrácená hodnota míry farness, tedy dalekost. 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 $\frac{1}{n-1} \sum_{v \in V} d_G(u, v)$ a closeness podle této definice:

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

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, modelují 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). 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 $O(|V||E| + |V|^2 \log |V|)$, kde V a E jsou množiny vrcholů a hran sítě (cite JO77, FT87).

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 pro použití Dijkstrova algoritmu pro všechny páry cest.

Pro rozsáhlé sítě s miliony uzlů (sociální sítě k dnešnímu datu) je tato metoda příliš náročná. Eppstein a Wang vyvinuli aproximační algoritmus s náročností $O(\frac{\log |V|}{\epsilon}^2 (|V| \log |V| + |E|))$ s chybou $\epsilon\delta$ pro převrácenou hodnotu closeness s pravděpodobností alespoň $1 - \frac{1}{n}$, kde $\epsilon > 0$ a δ je diametr sítě (nejdelší z nejkratších cest). Na základě tohoto aproximačního algoritmu byl vytvořen jiný aproximační algoritmus pro nalezení k nejvýznamnějších uzlů hodnocených podle closeness centrality.

Algorithm 1 Closeness

```
for s ∈ V do
    f ← 0
    d[v] ← ∞, v ∈ V
    d[u] ← 0
    Q ← {s}
    while Q ≠ ∅ do
        u ← extract-min(Q)
        f ← f + d[u]
        for sousedící vrchol v ∈ Adj[u] do
            if d[v] > d[u] + w(u, v) then
                d[v] ← d[u] + w(u, v)
                Q ← Q ∪ {v}
            end if
        end for
    end while
    c[s] ← (n−1)/f
end for
```

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

Aproximace

Algoritmus TOPRANK (Okamoto, Chen, Li) najde prvních k nejvýznamnějších uzlů s vysokou pravděpodobností 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.

3.1.5 Betweeness

Betweeness je druhá metoda, která modeluje šíření informace sítí pomocí nejkratších cest. Princip betweeness spočívá v zvýhodnění uzlů s pozicí, přes kterou teče nejvíce informace. Pokud uzel A komunikuje s uzlem C, pak můžeme tvrdit, že uzel B, který leží mezi nimi, bude mít roli prostředníka. Být tímto prostředníkem mezi více takovými uzly intuitivně napovídá, že takový uzel bude centrální. „Čím více lidí na mně závisí k vytvoření spojení s jinými lidmi, tím mám větší moc“ (cite introductino to social network methods). Betweeness měří na kolika nejkratších cestách se uzel nachází. Více se setkáme s definicí, kde do sumy zahrneme poměr cest, na kterých se uzel nachází, k celkovému počtu cest mezi dvěma uzly (Freeman, 1977; Anthonisse, 1971, Brandes).

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

$b(v)$ značí hodnotu betweeness centrality pro uzel v, V množinu uzlů, 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 betweeness je hodnota v intervalu od 0 do 1, kterou získáme tak, že betweeness vydělíme celkovým počtem možných cest - $((|V|-1)(|V|-2))$ pro orientované grafy a $(\frac{(|V|-1)(|V|-2)}{2})$ pro neorientované grafy.

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

Vznik betweeness je připisován sociologovi Lintonu Freemanovi (Freeman 77).

Brandesův algoritmus

Ve své práci Ulrik Brandes zmiňuje do té doby nejrychlejší algoritmus pro výpočet betweeness centrality s časovou náročností $\theta(|V|^3)$ a $\theta(|V|^2)$ paměťovými nároky. Tento způsob přistupuje k problému nekratších cest způsobem all-pair shortest paths. Brandesův způsob využívá algoritmů pro nalezení nejkratších cest z jednoho bodu, kde výsledný algoritmus pracuje s paměťovou náročností $O(|V| + |E|)$ a běží v čase $O(|V||E|)$ nebo $O(|V||E| + |V|^2 \log |V|)$ pro nevážený, respektive vážený graf.

Brandes ve své práci o algoritmu uvádí pseudokód pro nevážený graf, který je následně snadné pozměnit pro vážený graf zaměněním obyčejné fronty za prioritní frontu; kompletní důkaz správnosti algoritmu a porovnání standardního algoritmu s tímto (TODO cite Brandes).

Algoritmus lze paralelizovat stejně jako algoritmus pro closeness centralitu, pokud zajistíme synchronizaci vláken při přístupu k hodnotám betweeness. Narozdíl od closeness, kde výpočet pro jeden uzel ovlivňuje hodnotu closeness pouze pro tento uzel, u betweeness výpočet vycházející z jednoho uzlu ovlivňuje hodnotu betweeness potencionálně i pro všechny ostatní uzly.

Algorithm 2 Brandesův algoritmus

$b[v] \leftarrow 0,\ v \in V$
for $s \in V$ **do**
    $S \leftarrow$prázdný zásobník
    $P[w] \leftarrow$prázdný seznam, $w \in V$
    $\sigma[t] \leftarrow 0,\ t \in V$
    $\sigma[s] \leftarrow 1$
    $d[t] \leftarrow -1,\ t \in V$
    $d[s] \leftarrow 0$
    $Q \leftarrow \{s\}$
    while $Q \neq \varnothing$ **do**
        $v \leftarrow dequeue(Q)$
        $push(S, v)$
        for sousedící vrchol $w \in Adj[v]$ **do**
            if $d[w] < 0$ **then**
                $enqueue(Q, w)$
                $d[w] \leftarrow d[v] + 1$
            end if
            if $d[w] = d[v] + 1$ **then**
                $\sigma[w] \leftarrow \sigma[w] + \sigma[v]$
                $append(P[w], v)$
            end if
        end for
    end while
    $\delta[v] \leftarrow 0, v \in V$
    while $S \neq \varnothing$ **do**
        $w \leftarrow pop(S)$
        for $v \in P[w]$ **do**
            $\delta[v] + \frac{\sigma[v]}{\sigma[w]}(1 + \delta[w])$
        end for
        if $w \neq s$ **then**
            $b[w] \leftarrow b[w] + \delta[w]$
        end if
    end while
end for

Aproximace

I přes rychlejší Brandesův algoritmus je výpočet betweeness centrality příliš náročný výpočet pro sítě reálného světa (např. biologické, dopravní nebo webové sítě) a pokud nám jde více o relativní pořadí uzlů podle hodnoty betweeness 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 betweeness a odhadem chyby. Myšlenkou je jednoduchá lineární extrapolace Brandesova algoritmu, pokud do výpočtu zahrneme pouze náhodný vzorek namísto celé množiny vrcholů. Nechť k je velikost vzorku množiny vrcholů, se kterým počítáme, pak extrapolovaná hodnota betweeness je $\frac{|V|S}{k}$, kde S je vypočtená přibližná hodnota.

3.2 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 řešících metod různými lidmi (Shimbel [1955], Ford [1956], Dantzig [1958], Bellman [1958], Moore[1959], Dijkstra [1959]).(cite schrijver alexander ON THE HISTORY OF THE SHORTEST PATH PROBLEM)

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

3.2.1 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.

Algorithm 3 Prohledávání do šířky

```
function BFS(G, s)
    d[v] ← ∞,v ∈ V
    d[s] ← 0
    p[v] ← NIL,v ∈ V
    Q ← V
    while Q ≠ ∅ do
        u ← dequeue(Q)
        for sousedící vrchol v ∈ Adj[u] do
            if d[v] = −1 then
                d[v] ← d[u] + 1
                Q ← Q ∪ {v}
                p[v] ← u
            end if
        end for
    end while
    return p
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 problém 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 (cite CLRS, Fredman & Tarjan). 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 (cite CLRS):

Algorithm 4 Dijkstrův algoritmus

function DIJKSTRA(G, w, s) ▷ s je zdrojový vrchol a w je váhová funkce
    $d[v] \leftarrow \infty, v \in V$ ▷ Vrcholy jsou zpočátku nedosažitelné
    $d[s] \leftarrow 0$
    $p[v] \leftarrow$ NIL,$v \in V$ ▷ Předchůdci vrcholů zpočátku neexistují
    $Q \leftarrow V$ ▷ Prioritní fronta Q obsahuje všechny vrcholy
    while $Q \neq \emptyset$ **do** ▷ Postupujeme, dokud není prázdná
        $u \leftarrow$ *extract-min*(Q) ▷ Vytáhneme z fronty vrchol s nejmenší hodnotou $d[u]$
        for sousedící vrchol $v \in Adj[u]$ **do**
            if $d[v] > d[u] + w(u, v)$ **then**
                $d[v] \leftarrow d[u] + w(u, v)$ ▷ Relaxace hrany $\{u, v\}$
                $p[v] \leftarrow u$
            end if
        end for
    end while
    return p ▷ Strom předchůdců
end function

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

3.2.2 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 [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:

$$
\begin{aligned}
x + y &\equiv \min(x, y) \\
xy &\equiv x + y
\end{aligned}
$$

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:

$$
\begin{aligned}
d_0(u, v) &= \mathbf{A}_{uv} \\
d_{k+1}(u, v) &= \min(d_k(u, v), d_k(u, k) + d_k(k, v))
\end{aligned}
$$

Jednodušše zkoušíme, 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 všechny vrcholy k ($|V|^2|V|$).

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

3.3 Ostatní používané míry významnosti autorů

3.3.1 H-index

H-index je metoda pro měření produktivity a významu 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 sestupně podle počtu citací a poté od začátku tohoto seřazeného seznamu hledáme tu publikaci, jejíž počet citací je nižší než pořadové číslo v tomto seznamu. Toto číslo je pak h-index autora.

3.4 Porovnání výsledků

Pro zjištění, jestli jsou výsledky implementovaných metod shodné s uvedenými oceněními, použijeme metodu součtu pořadí oceněných autorů. Tzn. pro jedno ocenění sečteme pořadí všech autorů, kteří byli oceněni touhle cenou. Tuhle jednoduchou míru můžeme porovnat pouze mezi jednotlivými metodami pro jedno ocenění, ale ne mezi různými oceněními pro jednu metodu. Jednodušše protože například Turingova cena je ve výsledcích udělena pouze několika autorům, z čehož plyne malý součet pořadí, kdežto velké množství autorů na prvních pozicích je členy ACM Fellows, tím pádem velký součet pořadí oceněných.

Dále nás zajímá, zda-li jsou metody mezi sebou podobné či nikoliv. Pro porovnání metod mezi sebou je použit Spearmanův koeficient pořadové korelace.

3.4.1 Spearmanův koeficient pořadové korelace

Spearmanův koeficient je klasický Pearsonův koeficient korelace, který je aplikovaný na proměnné s pořadím (cite Myers, Jerome L.; Well, Arnold D. (2003), Research Design and Statistical Analysis (2nd ed.), Lawrence Erlbaum, pp. 508).

Koeficient korelace obecně dosahuje hodnot od -1 do 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 měřenými proměnnými.

KAPITOLA 4

Výsledky

4.1 Žebříčky významných autorů

Viz příloha A A. Hodnoty u metody PageRank jsou transformovány z intervalu $[0; 1]$ na $[0; |V|]$.

4.2 Porovnání nejvýznamnějších autorů

4.3 Porovnání implementovaných metod

KAPITOLA 5

Diskuse

5.1 Podobnost výsledků jednotlivých metod

DBLP

pagerank	0.520	0.930	0.173	0.830	0.134	0.850	0.140	0.923	-
pagerank	0.514	0.993	0.210	0.871	0.130	0.869	0.136	-	
w outdegree	0.258	0.127	0.568	0.179	0.999	0.193	-		
w indegree	0.596	0.878	0.217	0.989	0.186	-			
outdegree	0.253	0.121	0.567	0.173	-				
indegree	0.571	0.880	0.203	-					
radius	0.300	0.203	-						
betweeness	0.528	-							
h-index	-								
-									

5.2 Shoda výsledků s oceněními

5.3 Vliv vah na přesnost výsledků

5.4 Vstupní a výstupní hrany

KAPITOLA 6

Závěr

PŘÍLOHA A

Žebříčky významných autorů

A.1 DBLP

	Autor	outdegree	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				

	Autor	**indegree**	**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				

	Autor	**wOutdegree**	**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		•	•	

	Autor	**wIndegree**	**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	YEHOSHUA 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				

	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		•	•	•
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				

	Autor	**inCloseness**	**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				

	Autor	**wCloseness**	**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		•		
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				

	Autor	**approximatedbetweenessnover4**	**Turing**	**Codd**	**Fellows**	**IS**
1	PHILIP A. BERNSTEIN	62833222.856		•		
2	MICHAEL STONEBRAKER	61060269.416		•	•	
3	DAVID J. DEWITT	59157247.917		•	•	
4	UMESHWAR DAYAL	58213170.228		•		
5	JIM GRAY	58075515.166		•		
6	RAYMOND A. LORIE	56617190.909				
7	DONALD D. CHAMBERLIN	56486140.230		•	•	
8	MICHAEL J. CAREY	56041230.549		•		
9	KAPALI P. ESWARAN	55785110.823				
10	JEFFREY D. ULLMAN	55779941.492		•		
11	E. F. CODD	55520481.855		•		
12	WON KIM	54705881.971				
13	DAVID MAIER	53742407.716		•		
14	MORTON M. ASTRAHAN	52845731.275				
15	FRANCCEDILOIS BANCILHON	52230591.856				
16	NATHAN GOODMAN	52118900.717				
17	EUGENE WONG	49805297.281				
18	IRVING L. TRAIGER	49604533.387			•	
19	HECTOR GARCIA-MOLINA	48896884.351		•		
20	CATRIEL BEERI	48804707.627			•	•
21	RONALD FAGIN	47993987.430		•	•	•
22	BRUCE G. LINDSAY	47844195.026				
23	SERGE ABITEBOUL	47245747.572		•	•	•
24	RAKESH AGRAWAL	46003606.416		•		
25	DENNIS MCLEOD	44749242.164				
26	PATRICIA G. SELINGER	44308278.984		•	•	
27	JOHN MILES SMITH	44222906.647				
28	HAMID PIRAHESH	44026421.580			•	
29	THOMAS G. PRICE	43911467.842				
30	HENRY F. KORTH	43801765.543			•	

	Autor	**betweenessParallel**	**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				

	Autor	**wBetweeness**	**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	YEHOSHUA 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				
26	GOETZ GRAEFE	27015757.722				
27	IRVING L. TRAIGER	26609996.395			•	
28	LAURA M. HAAS	26327991.827				
29	MORTON M. ASTRAHAN	26134436.913				
30	H. V. JAGADISH	24773687.470			•	

	Autor	hindex	Turing	Codd	Fellows	ISI
1	MICHAEL STONEBRAKER	28.000		•	•	
2	DAVID J. DEWITT	24.000		•	•	
3	JEFFREY D. ULLMAN	24.000		•		
4	PHILIP A. BERNSTEIN	22.000		•		
5	RAKESH AGRAWAL	21.000		•		
6	WON KIM	21.000				
7	YEHOSHUA SAGIV	20.000				
8	CATRIEL BEERI	20.000			•	•
9	MICHAEL J. CAREY	20.000		•		
10	SERGE ABITEBOUL	19.000		•	•	•
11	HECTOR GARCIA-MOLINA	19.000		•		
12	UMESHWAR DAYAL	19.000		•		
13	CHRISTOS FALOUTSOS	19.000				
14	NATHAN GOODMAN	18.000				
15	JIM GRAY	18.000		•		
16	JEFFREY F. NAUGHTON	18.000				
17	RAGHU RAMAKRISHNAN	18.000				
18	RONALD FAGIN	18.000		•	•	•
19	JENNIFER WIDOM	18.000		•	•	
20	DAVID MAIER	17.000		•		
21	BRUCE G. LINDSAY	17.000				
22	SHAMKANT B. NAVATHE	16.000				
23	C. MOHAN	16.000		•		
24	HAMID PIRAHESH	16.000			•	
25	H. V. JAGADISH	16.000			•	
26	YANNIS E. IOANNIDIS	16.000				
27	MOSHE Y. VARDI	15.000		•	•	•
28	FRANCCEDILOIS BANCILHON	15.000				
29	ALBERTO O. MENDELZON	15.000				
30	CARLO ZANIOLO	15.000				

A.2 CiteSeer

	Autor	outdegree	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				

	Autor	indegree	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 ZWAENEPOEL	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				

	Autor	wOutdegree	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				

	Autor	**wIndegree**	**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				

	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				
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				

	Autor	**inCloseness**	**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				

	Autor	wCloseness	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				

	Autor	betweenessParallel	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				

	Autor	wBetweeness	Turing	Codd	Fellows	ISI
1	RAMESH GOVINDAN	9236624144.764				
2	DEBORAH ESTRIN	9204684837.308			•	
3	SCOTT SHENKER	9093192865.090				
4	JOHN HEIDEMANN	7995791211.560				
5	SALLY FLOYD	7541647035.535				
6	LIXIA ZHANG	7036608526.217				
7	VAN JACOBSON	6839488033.728				
8	HARI BALAKRISHNAN	6839435524.752				
9	M. FRANS KAASHOEK	6613771252.670				
10	STEVEN MCCANNE	6467785219.402				
11	ROBERT MORRIS	6366861686.720				
12	DON TOWSLEY	5982456216.269				
13	JIM KUROSE	5527191589.744				
14	HUI ZHANG	5176531717.047				
15	CHING-GUNG LIU	5120183722.781				
16	DAVID CULLER	5088337083.240				
17	SENIOR MEMBER	4986684628.204				
18	JENNIFER WIDOM	4904364960.579		•		
19	HECTOR GARCIA-MOLINA	4891302563.725		•		
20	LEWIS GIROD	4865909407.605				
21	MARK H	4862908859.075				
22	STUDENT MEMBER	4836314123.276				
23	VICTOR FIROIU	4787506549.742				
24	JEREMY ELSON	4748827894.239				
25	DOMENICO FERRARI	4711946517.104				
26	ION STOICA	4650529706.620				
27	VERN PAXSON	4614740987.807				
28	JITENDRA PADHYE	4425596738.761				
29	ANDREW S. TANENBAUM	4421498874.688				
30	HAOBO YU	4420381445.078				

	Autor	hindex	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	30.000				
8	RAKESH AGRAWAL	29.000		•		
9	M. FRANS KAASHOEK	29.000				
10	JENNIFER WIDOM	29.000		•		
11	WILLY ZWAENEPOEL	28.000				
12	HUI ZHANG	27.000				
13	IAN FOSTER	27.000				
14	MONI NAOR	27.000				
15	SALLY FLOYD	26.000				
16	LUCA CARDELLI	26.000			•	
17	SERGE ABITEBOUL	26.000		•		
18	SENIOR MEMBER	26.000				
19	BART SELMAN	26.000				
20	SEBASTIAN THRUN	25.000				
21	OREN ETZIONI	24.000				
22	DAVID J. DEWITT	24.000		•	•	
23	DAPHNE KOLLER	24.000				
24	RAJEEV ALUR	24.000			•	•
25	HARI BALAKRISHNAN	24.000				
26	MAURIZIO LENZERINI	24.000				
27	ODED GOLDREICH	24.000				
28	STUDENT MEMBER	23.000				
29	MARK D. HILL	23.000				
30	ROBERT HARPER	23.000				