VERSION 1.2

UTVÄRDERING AV BIBLIOMETRISKA INDIKATORER

Rapport om metoder för utvärdering av universitet och högskolor

Av Erik Sandström (GU) & Ulf Sandström (LiU)
18 jan 2008

Citeringsfördelningar skiljer sig mellan olika ämnesområden, artikeltyper, och tidskrifter. Genom att arbeta med medelvärden inom normaliseringsgrupper (artiklar inom samma tidskriftsklass, av samma dokumenttyp och publicerade under samma år) har citeringsanalysen blivit mer tillförlitlig; ambitionen om jämförbarhet uppfylls förhållandevis väl (Schubert, Glänzel & Braun 1988, van Raan 2004).

Field Citation Score (FCS) och *Journal Citation Score* (JCS) är två numera flitigt använda referensmått beräknade på olika normaliseringsgrupper. FCS är antalet citeringar per artikel inom ett ämnesområde och JCS är motsvarande inom en tidskrift. Dessa två används för att beräkna normaliserade citeringsvärden, där antalet citat divideras med referensvärdet. Resultatet blir CPP/FCS och CPP/JCS (CPP=cites per paper). Den första, vilken brukar kallas för ”crown indicator”, är den vi främst kommer att behandla här, men de principer som läggs fram är tillämpliga även på CPP/JCS.

Framställningens syfte är att beskriva och diskutera metoder för bibliometrisk utvärdering av forskargrupper, avdelningar, institutioner och universitet. Det är således utvärdering av forskningsenheter som står i fokus och diskussionen är inriktad på de olika måttens tillämpbarhet och relevans. Vi för fram ett delvis nytt sätt att beräkna den relativa citeringsfrekvensen **Standard Citation Score** (SCS) – samt avslutar genom att beröra frågan om tidskriftsklasser. Vi föreslår även där en alternativ strategi: i stället för tidskriftsklasser föreslår vi en klustring av artiklar på deras bibliografiska kopplingar som underlag för normalisering av citeringsdata.

Beräkning av indikatorn (CPP/FCS)

Indikatorn CPP/FCS har tagits fram av Leidengruppen[1] och beräknas vanligtvis enligt följande:

$$\frac{\sum_{i=1}^{P} c_i}{\sum_{i=1}^{P} [\mu_f]_i}$$

där c är antalet erhållna citat och μ_f är medelvärdet för normaliseringsgruppen (FCS). Genom denna beräkningsmodell utförs normaliseringen på aggregerad nivå, t.ex. forskargrupp- eller universitetsnivå. I våra beräkningar har vi emellertid valt att justera beräkningen något genom att utföra normaliseringen på artikelnivå. Beräkningsformeln blir då:

$$\frac{1}{P}\sum_{i=1}^{P} \frac{c_i}{[\mu_f]_i}$$

Skillnaden är att vårt beräkningssätt behandlar alla artiklar lika medan Leidengruppens metod ger högre vikt till publiceringar inom normaliseringsgrupper med höga referensvärden.[2] För ett universitet som publicerar såväl inom medicin som inom samhällsvetenskap blir konsekvensen av Leidengruppens metod att citeringsvärdet i hög grad kommer att bero på medicinarnas publikationer. Med vår beräkningsmodell väger samhällsvetarnas publikationer lika tungt som medicinarnas

[1] Se van Raan (2004.2006). Leiden är ett holländskt universitet där gruppen CWTS underledning av Anthony van Raan, är verksam .
[2] Jfr Lundberg (2006), s. III:3.

och kommer att påverka slutresultatet i samma mån. Enligt vårt förmenande ger detta ett mer rättvisande resultat.

Trots att vår beräkning skiljer sig lite från Leidengruppens har vi av praktiska skäl valt att benämna den på samma sätt. Det bör dock påpekas att beteckningen CPP/FCS är något missvisande när vår beräkning används.

Citeringsfördelningar

Med fältnormaliseringen har citeringsanalyserna blivit mer tillförlitliga och användbara. Många ställer dock frågan om det finns möjlighet att uttala sig om signifikansnivåer av de resultat som presenteras. Är det säkert att det går att använda medelvärden och variansmått för den här typen av data?

En rimlig utgångspunkt för att svara på denna fråga är att se närmare på fördelningen av de värden vi utför beräkningar på. Följande figur 1 visar citeringsfördelningen för tidskriftsklassen *Dentistry, Oral Surgery and Medicine*. Analysen gäller artiklar publicerade 2003-2004 med treårigt citeringsfönster.

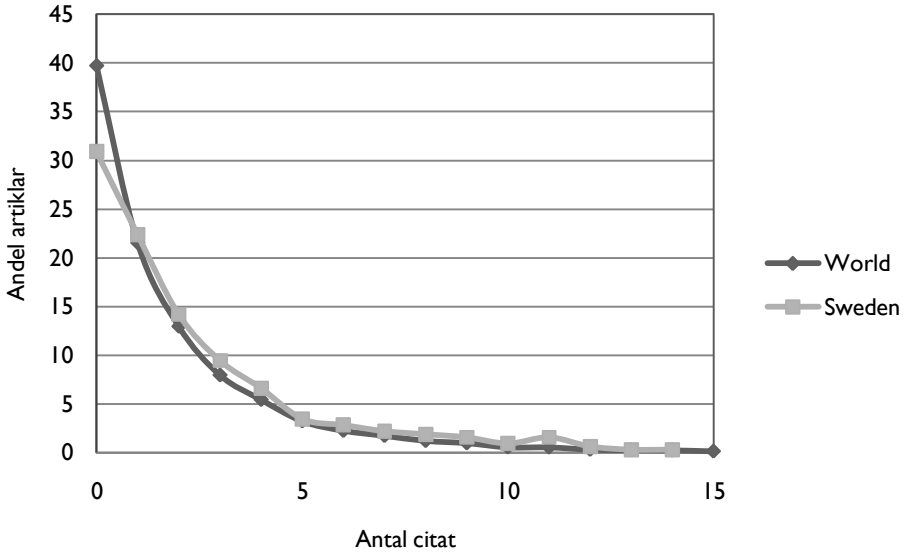


Fig 1. Citeringsfördelning Dentistry. Artiklar 2003-2004 med treårigt citeringsfönster. Anm: Figuren visar endast artiklar med <16 citeringar.

Närmare 40 procent av världens artiklar har med treårigt citeringsfönster inte refererats av andra artiklar i databasen (inklusive icke-dentistry artiklar). Drygt 20 procent har fått en citering och cirka tretton procent två citeringar o.s.v. Världen och Sveriges fördelningar följer som väntat varandra även om Sverige har färre nollciterade artiklar.

Vidareutveckling av bibliometriska indikatorer

Medelvärden och spridningsmått kan vara missvisande för skeva fördelningar, och citeringsfördelningar är ett typfall av skeva fördelningar. Van Raan (2006) har emellertid visat att tidskrifts- och fältnormaliserade citeringsindikatorer närmar sig normalfördelning på enhetsnivå, t.ex. forskargrupper. Innebörden av detta viktiga

konstaterande är att fältnormaliserade indikatorer baserade på medelvärden kan användas för att jämföra forskargrupper, universitet etc.

Kronindikatorn är därför ett relevant mått, men likväl inte utan brister. Kritiken gäller att den inte tar hänsyn till den statistiska variansen i de olika normaliseringsgrupperna. Effekten blir att det, trots normaliseringen, är svårt att göra helt rättvisande jämförelser mellan grupper som är aktiva inom skilda ämnesområden. Forskningsområdens heterogenitet är ett välkänt faktum; bl.a. har Richard Whitley (2000) och Stephen Cole (1994) pekat på strukturella skillnader i förutsättningarna att uppnå konsensus. Detta påverkar bland annat hur vetenskapliga konflikter hanteras inom olika fält. Framförallt är det graden av kollegial kontroll som skiljer sig mellan områden. Inom t.ex. fysik är den kollegiala kontrollen stark, merparten av resurserna fördelas med *peer review*, samtidigt som osäkerheten i tolkningen av ett forskningsresultat är låg. Alla vet när någon har publicerat ett för området värdefullt resultat. Forskningsfronterna inom området är välstrukturerade och så gott som alla områdets forskare känner snabbt igen viktiga resultat. Inom områden som uppvisar motsatta karakteristika, dvs. låg kollegial kontroll och hög tolkningsosäkerhet, t ex energiforskning och sociologi, blir konsekvensen att områdets forskare arbetar på forskningsfronter där man sinsemellan har svårt att värdera varandras resultat (Sandström 2003).

Poängen med denna teoretiska analys är att vi kan förklara varför kunskapsområden skiljer sig med avseende på citeringarnas fördelning. I områden som uppvisar låg tolkningsosäkerhet är det troligt att det blir stora skillnader mellan artiklars citeringsgrad. Eftersom områdets forskare är förhållandevis eniga i värderingen av ett forskningsresultat blir det också ovanligt många citeringar till ett fåtal betydelsefulla artiklar, dvs. det blir stor spridning i antalet citeringar per artikel. I områden med hög tolkningsosäkerhet blir i stället spridningen mera måttlig.

Av detta följer att det sannolikt är betydligt svårare att uppnå en hög generell citeringsgrad för en forskargrupp eller institution som arbetar inom fysik än inom sociologi. En lösning på detta problem är att använda z-score, som, förutom medelvärdet, tar hänsyn till standardavvikelsen. Fördelningarnas skevhet förhindrar emellertid användandet av standardavvikelsen som spridningsmått. Glänzel (Glänzel 1996) har försökt lösa detta problem genom att använda ett spridningsmått som baserar sig på de matematiska regler som gäller för Waringfördelningar – en typ av diskret sannolikhetsfördelning som visat sig vara användbar för bibliometriska data. En annan lösning föreslogs redan på 1970-talet av den amerikanska gruppen kring CHI, Computer Horizon Inc. (McAllister et.al. 1983, jfr. Allison 1980).[3] CHI löser skevhetsproblemet genom att logaritmera varje artikels citeringsantal. Medelvärde och standardavvikelse för dessa logaritmerade värden räknas sedan fram inom varje normaliseringsgrupp. Dessa värden används sedan i beräkningen av z-score för en artikel:

[3] CHI leds av Francis Narin och företaget har bland annat varit ledande amerikansk producent av nationella indikatorer för vetenskap och teknik.

$$\frac{1}{P}\sum_{i=1}^{P}\frac{ln(c+0{,}5)-\left[\mu_{f[ln]}\right]_i}{\left[\sigma_{f[ln]}\right]_i}$$

där c är antalet erhållna citat, $\mu_{f[ln]}$ är medelvärdet av det logaritmerade antalet erhållna citeringar (plus 0,5) för artiklar inom normaliseringsgruppen och $\sigma_{f[ln]}$ är standardavvikelsen av $\mu_{f[ln]}$fördelningen.[4]

Resultatet blir ett värde som säger hur många standardavvikelser ifrån medelvärdet det aktuella citeringsvärdet ligger.

Vi föreslår att den metod som har beskrivits här benämns **Standard Citation Score** (SCS) och att den får tjäna som huvudindikator vid bedömning av forskargrupper, institutioner och lärosäten.[5] Om vi går över till att använda denna indikator bör vi samtidigt fortsätta att använda CPP/FCS eftersom många är bekanta med den metoden och jämförelsen ger i sig en del information om den undersökta enheten.

Jämförelser mellan indikatorerna

För att visa hur kronindikatorn och SCS skiljer sig åt fördelningsmässigt kan vi se närmare på Malmö högskolas artiklar under 1998-2005. Dessa har laddats ner från Web of Science utan särskild adressrättning. Siffrorna nedan skall därför endast ses som ett exempel på skillnaden mellan kronindikatorn och SCS, inte en citeringsanalys av Malmö högskola.

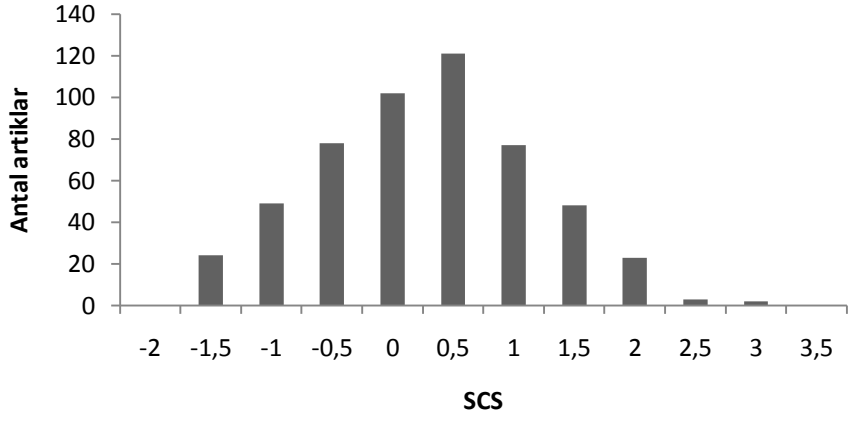
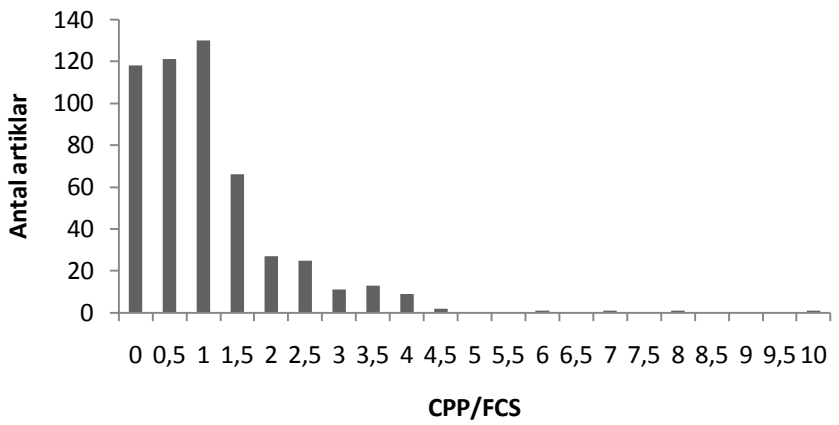


Fig 2. Fördelning av artiklar över grupper av z-score. Figuren är baserad på 527 artiklar från Malmö Högskola under tidsperioden 1998-2005

[4] Allison (1980) föreslår att lägga till 0,5 vid logaritmering av citeringsdata. Vi använder här Lundbergs (2006) föreslagna notationer.
[5] CHI:s förslag var att benämna indikatorn Standardized Citation Score (McAlister et. al. 1983).



.

Fig 3. Fördelning av artiklar över grupper av CPP/FCS. Figuren är baserad på 527 artiklar från Malmö Högskola under tidsperioden 1998-2005.

Förutsatt att det finns en spridning av normaliseringsgrupper inom materialet så närmar sig den på z-score baserade SCS-indikatorn normalfördelning redan på artikelnivå. Kronindikatorn (CPP/FCS) blir normalfördelad först på aggregerad nivå, t.ex. lärosätes- eller institutionsnivå (se fig 3). En anledning till att kurvorna ser annorlunda ut är att lägsta värdet för CPP/FCS alltid är noll. Det gör att artiklarna travas upp i noll-stapeln.

I fig 4 och 5 visas indikatorernas fördelningar på lärosätesnivå inom tidskriftsklassen *Dentistry, Oral Surgery & Medicine*. Båda fördelningarna närmar sig normal-fördelning, med viss fördel för SCS.

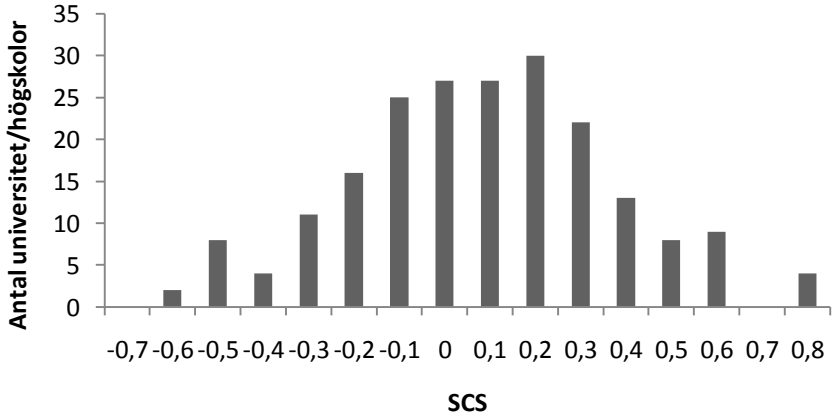


Fig 4. Fördelning av universitet/högskolor över grupper av SCS (standard citation score). Figuren är baserad på värden för 208 universitet/högskolor inom ämnesområdet Dentistry, Oral Surgery & Medicine och tidsperioden 2003-04.

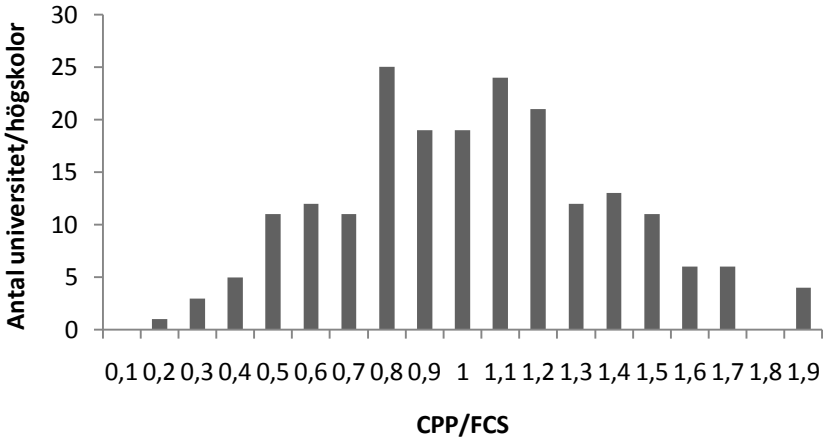


Fig 5. Fördelning av universitet/högskolor över grupper av CPP/FCS. Figuren är baserad på värden för 208 universitet/högskolor inom ämnesområdet Dentistry, Oral Surgery & Medicine och tidsperioden 2003-04.

Sammanfattningsvis kan konstateras att SCS har fördelar. Indikatorn tar hänsyn till citeringsfördelningarnas varians och minskar extremvärdenas betydelse för slutvärdet. Samtidigt bör man komma ihåg att det är de högciterade artiklarna som inverkar på den aktuella forskningsfronten och som, inom de flesta områden, är mest intressanta i de sammanhang när policy och utvärdering står i fokus. Ett sätt att fånga in denna aspekt vore att ange antalet artiklar inom topp 5 %, vilket kan göras för avdelning, institution eller lärosäte.

I tabell 1 visas SCS-värden för två institutioner vid ett svenskt universitet under perioden 2004-2005. Skillnaden mellan de två beräkningsgrunderna framgår med önskvärd tydlighet. Till följd av logartimeringen av citeringar reduceras den effekt som extremvärdena har på medelvärdet och fysik hamnar över det internationella genomsnittet. Fysikämnet har egenskaper som gör att skevheten i citeringsfördelningen ger ett högre medelvärde. Därför blir SCS-värdet högre.

Tabell1. Relativ citeringsgrad mätt med två metoder: Exempel från ett svenskt universitet 2004-2005 med öppet citeringsfönster till maj 2007.

test	SumFRAC	SCS	CPP/FCS
fysik	>100	0,06	0,96
datavetenskap	>50	0,03	1,12

Författarfraktionaliserade data.

Standardfelet

Standardfelet *d(x)* för medelvärdet för SCS *(x)* för en forskningsenhet beror på storleken på enheten och fördelningens spridning enligt följande:

$$d(x)=D/\sqrt{n}$$

där n är antalet artiklar publicerade vid enheten och D är standardavvikelsen hos SCS-fördelningen (Schubert & Glänzel 1983).

Standardfelet kan sedan användas för att göra ett signifikanstest. Testet visar om x avviker signifikant från det internationella genomsnittet (0,00) om $(x-a)/d(x)$ är större än 1,96 eller mindre än -1,96 förutsatt att vi tillämpar ett konfidensintervall på 95 %. Institutionerna i tabell 1 ligger båda alltför nära 0 och antalet artiklar är för få för att värdena skall uppfylla signifkansvillkoren.

Alternativ till tidskriftsklasser som normaliseringsgrupp

Utvecklingen av avancerade metoder för beräkning av citeringsgrad har en uppenbar brist. Normaliseringen görs mot tidskriftsklass och den ordningen följer Thomson/ISI:s ”subject categories”. Många har velat sätta frågetecken för ISI:s sätt att klassificera tidskrifter. På *Sigmetrics* diskussionsforum förekommer det inte sällan inlägg som påvisar fel och brister i den klassning som Thomson/ISI tillämpar. Loet Leydesdorff[6] hör till dem som brukar framföra kritiska åsikter (se Leydesdorff 2008). I ett antal undersökningar baserade på Journal Citation Reports (JCR) från ISI har han visat att ISI-klassningarna kan ifrågasättas (se särskilt Leydesdorff 2006).

Det bör dock framhållas att klassificeringen är betydligt mera intrikat och genomtänkt än vad som normalt framkommer av diskussionsinläggen: alla som deltar i debatten är nog inte riktigt införstådda med klassningens finesser. Undersökningar[7] visar också att den på vissa områden har betydligt bättre träffsäkerhet än vad som vanligtvis medges, men att det på andra områden inte finns särskilt hög korrelation mellan forskningsfält och tidskriftsfält.[8]

Klart är att det i de flesta fall inte går att avgränsa ett forskningsområde med hjälp av tidskriftskategorier. Den engelske bibliometrikern Grant Lewison är kanske den som tydligast visat att avgränsningen av ett forskningsområde kräver en genomtänkt strategi som bygger på nyckeltermer, specialtidskrifter och expertomdömen (Lewison 1999, 1996). Om man i stället skulle utgå från ISIs tidskrifter för ett område skulle en stor del av materialet inte komma med. Ofta är det dessutom centrala artiklar som publiceras i mer generella tidskrifter som *PNAS, Science och Nature*, vilka hamnar utanför. Dessa tidskrifter klassificeras som ”Multidisciplinary”, vilket kan hävdas vara oegentligt eftersom det snarare handlar om att tidskrifterna

[6] Leydesdorff är verksam vid Amsterdam School of Communications Research (ASCoR)-

[7] Se t.ex. Basu & Lewison (2005) som visar att 77% av ett kalibrerat antal astronomiartiklar förekommer i tidskrifter som klassats som astronomitidskrifter. Dock bör sägas att detta är representativt för alla tidskriftsklasser.

[8] I våra arbeten vid KTH och Linköpings universitet har vi ofta understrukit att tidskriftskategorierna kan användas för att ange ett universitets styrka inom ett angivet forskningsfält, men att det inte går att utifrån detta sluta sig till vilka forskargrupper som bidragit till denna styrka.

publicerar artiklar från många områden, inte att varje artikel i sig är fler- eller tvärvetenskaplig. Det faktum att ett antal viktiga artiklar automatiskt hamnar utanför är också av betydelse för normaliseringen av citeringsdata. *Nature* och *Science* innehåller artiklar som hör till områden som biomedicin, biologi, fysik, kemi och geologi. Det är alltså delar av flera helt åtskilda *research communities* som förs samman och tilldelas samma normaliseringsgrupp, vilket sannolikt gynnar forskare inom kemi och biomedicin men missgynnar forskare inom geologi. Det vore möjligt att fördela artiklarna i dessa ”multidisciplinära” tidskrifter till respektive tidskriftsklass t.ex. utifrån de tidskrifter som används i referenserna, vilket föreslagits av Glänzel, Schubert och Czerwon (1998), och det är vår avsikt att använda denna metodik. I anslutning till att är arbetat med denna problematik har det blivit tydligt att det behövs ett alternativ till ISI:s tidskriftsklasser som grund för normalisering av citeringsdata.

Det finns en rad förslag att ta hänsyn till, var och en med sina för- och nackdelar, och de skiljer sig med avseende på om man använder data på artikel- eller tidskriftsnivå samt vilken metod som används för att ta fram underliggande strukturer, antingen co-citations eller bibliografisk koppling (bibliograhical coupling). Metoder som utgår från tidskrifter lägger vi här åt sidan eftersom vi anser att de inte är tillfyllest när det gäller att representera forskningens olika specialiteter. Kvar blir då att utgå från artikelnivå för att bygga upp ett system över vetenskapernas struktur. På grund av att det är frågan om stora datamängder, ungefär 1 miljon artiklar med cirka 30 miljoner referenser tillkommer per år i ISI-databasen, har det varit praktiskt svårt att bearbeta så stora datamängder, men numera finns mjukvara som hanterar detta med rimlig tidsåtgång.

Det finns, som framhållits ovan, två likhetsmått (proximity) som brukar användas i bibliometriska analyser. Den första är bibliografisk koppling. Två dokument är kopplade om deras respektive referenslistor innehåller en eller flera gemensamma referenser. Styrkan i relationen uttrycks i antalet gemensamma referenser (Egghe & Rousseau 1990). Det kan riktas kritik mot denna metod eftersom det faktum att två artiklar har gemensamma referenser inte innebär att de refererar till samma information i den citerade artikeln (Verbeek et al 2002). I bibliometrisk forskning har co-citeringsanalys blivit mer använd som metod för etablera kopplingar mellan dokument. Två dokument är ”co-cited” om de samtidigt citeras i dokument utgivna under en avgränsad period. Styrkan i relationen bestäms av frekvensen samciteringar. Bibliografisk koppling fokuserar således på grupper av artiklar som refererar till gemensamma dokument medan samciteringar grupperar artiklar som parvist förekommer i artiklarnas referenser. Självklart kan samma kritik riktas mot co-citeringar som mot bibliografisk koppling. Det är inte säkert att varje förekomst av samcitering indikeras av samma orsak.

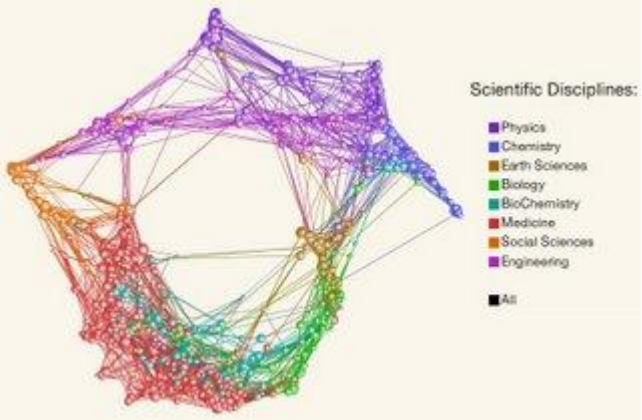
> *In co-citation earlier documents become linked because they are later cited together; in bibliographic coupling later documents become linked because they cite the same earlier documents. The difference is that bibliographic coupling is an association intrinsic to the documents (static), while co-citation is a linkage extrinsic to the documents, and one that is valid only so long as they continue to be co-cited (dynamic). [Smith 1981]*

Kopplingsmåtten – likhetsmåtten – kan konstrueras så att de normaliseras med hänsyn till hur ofta de förekommer, referenslistornas längd och citeringsfrekvensen.

Av likhetsmåttens egenskaper följer att bibliografisk koppling går att genomföra fram till föregående års artiklar. Samcitering som mått blir sannolikt beroende av ett tre- till fyraårigt citeringsfönster om vi tänker oss att hela databasen skall struktureras efter likhetsmåttet. Av den anledningen föredrar vi att använda bibliografisk koppling, av hela databasen, som alternativ till ISI:s tidskriftsklasser.

Vi följer ett tillvägagångssätt som tagits fram vid Sandia Research Labs. Den bibliometriska kopplingen beräknas mellan alla artiklar det aktuella året. Varje artikels tio starkaste kopplingar skickas sedan in i ett program som utvecklats av Sandia (VxOrd). Detta program använder en force-directed layout algorithm för att beräkna koordinater för artiklarna, baserat på deras likhet. Koordinaterna kan sedan användas för klustring. I detta fall används en modifierad single-link-algoritm (se illustration nedan för ett av resultaten från Sandia).

Denna procedur upprepas (klustren klustras) tills de cirka en miljon artiklarna fördelats på cirka 300-400 kluster. Poängen är att dessa kluster – ämnesområden – kan tas som utgångspunkt för en normalisering av citeringsdata som är oberoende av tidskriftsklasserna.



Obs! Bilden bör skrivas ut med färgskrivare.

Illustration: Boyack och Klavans Map of Science
(baserad på co-citations av 820 000 citerade artiklar, ISI-data 2003-04)

Litteratur

Allison, P, (1980), Inequality and scientific productivity, *Social Studies of Science*, 10 163-179.

Basu, A. & Lewison, G. (2005) Going beyond journal classification for evaluation of research outputs, *Journal of Documentation*, 57, (3): 232–246.

Boyack, K. W., Klavans, R., Borner, K. (2005), Mapping the backbone of science, Scientometrics, 64: 351–374.

Cole, S. (1994) *Making science: between nature and society*. Cambridge, Mass.: Harvard University Press.

Egghe, L & Rousseau, R. (1990) Introduction to bibliometrics. Amsterdam: Elsevier.

Glänzel, W. (1996), The need for standards in bibliometric research and technology. *Scientometrics* 35:167 – 176.

Lewison, G. (1996) The definition of biomedical research subfields with title keywords and application to the analysis of research outputs.Research Evaluation, 6: 25–36.

Lewison, G. (1999), The definition and calibration of biomedical subfields. *Scientometrics* Vol. 46, No. 3 (1999) 529-537.

Lundberg, J. (2006) *Bibliometrics as a research assessment tool – impact beyond the impact factor*. PhD-thesis, Karolinska Institute. Stockholm

McAllister, P. R., Narin, F., Corrigan, J. G. (1983), Programmatic evaluation and comparison based on standardized citation scores, *IEEE Transactions on Engineering Management*, 30 : 205–211.

Sandström, U. (2003), Eftertankar om forskningspolitik. *Arkus verksamhets-berättelse* 2002/03. Stockholm: Arkus. (www.forskningspolitik.se)

Schubert, A. Glänzel, W. Braun, T. (1988), Against absolute methods: relative scientometric indicators and relational charts as evaluation tools. In: van Raan, de Bruin, Moed, Nederhof, Tijssen (Eds.) Leiden Conference on Science and Technology in a Policy Context. Leiden: DSWO Press, pp. 35-50.

Schubert, A. Glänzel, W. (1983) Statistical reliability of comparisons based on the citation impact of scientific publications. *Scientometrics* vol 5:59 – 74.

van Raan, A.J.F. (2004) Measuring science: Capita selecta of current main issues. In: H.F. Moed, W, Glänzel and U. Schmoch, editors: *Handbook of Quantitative Science and Technology Research*, Dordrecht: Kluwer Academic Publishers.

Verbeek, A. Debackere, K. Luwel, M. & Zimmermann, E. (2002) Measuring progress and evolution in science and technology–1. The multiple uses of bibliometric indicators. *International Hournal of Management Reviews* vol 4. Iss. 2, pp.179–211.

Visser, M.S. Nederhof, A.J. (2007) Bibliometric study of the Uppsala University, Sweden, 2002–2006. In: Quality and renewal 2007: An overall evaluation of research at Uppsala University 2006/2007. Uppsala: Uppsala University.

Whitley, R. (2000) *The intellectual and social organization of the sciences.* Second Edition (First ed. 1984). Oxford: Oxford University Press.