

Bibliometri

– aktuella trender

av Olle Persson

Olle Persson är professor vid Sociologiska institutet vid Umeå universitet och en auktoritet inom området bibliometri.



Recent trends in bibliometrics

Some recent developments within bibliometrics are outlined, with special reference to research evaluation and bibliometric indicators of research performance. A few examples from Nordic library and information science context are presented.

Aldrig förr har väl forskning och högre utbildning granskats och jämförts som idag. Ansökningar om anslag och tjänster skall gallras och rangordnas, universiteten skall profilera sig och satsa på starka forskargrupper, fakultetsmedel skall fördelas. Allt skall mätas på längden och bredden. Ingen kommer undan och rangordnar vi inte oss själva så görs det av andra - kineser, holländare och amerikaner. Vad beror denna utveckling på?

En viktig omständighet är konkurrens om resurser. Att placera sig väl på topplistorna blir alltmer viktigt i en situation då resurserna inte växer längre. Högt rankade universitet, forskargrupper eller forskare, torde kunna locka till sig mer pengar, fler studenter, bättre forskare och samarbetspartners än de som hamnar längre ned på listan. Det handlar om en skarp tävlan mellan aktörer inom forskarvärlden såväl lokalt och nationellt som internationellt. Därför har behovet och betydelsen av mätbara prestationer ökat och därmed har också bibliometrin hamnat i fokus.

Bibliometri betyder helt enkelt konsten att räkna eller mäta böcker. Man borde kanske hellre tala om artikelmetri eftersom man idag företrädesvis räknar tidskriftsartiklar och i undantagsfall böcker. Artiklarna är lättare att hitta och är för syftet bättre beskrivna än böcker i de befintliga databaserna. Man tager vad man haver.

Den som skriver fler artiklar än andra, inte nödvändigtvis fler sidor, anses prestera mer. Kan vi dessutom räkna citaten till artiklarna så kan vi dessutom säga att vissa presterar bättre än andra. Artiklar likställs med aktivitet och citat med genomslag, impact och ibland med kvalitet. På det hela taget är detta logiskt, åtminstone statistiskt sett. Det omvända vore orimligt. Förvisso finns det en och annan fuskande författare och nog är en del citat simpelt smicker. Men det statistiska sambandet är positivt och tämligen starkt mellan bibliometriska prestationsmått och kvalitativa bedömningar gjorda av forskarkollegor (peer-reviews). Ungefär så går det bibliometriska tänkandet. Idealet är när bibliometriska data kombineras med peer reviews. Pekar de åt samma håll kan vi vara säkrare i våra slutomdömen, pekar de åt olika håll behöver vi fundera ytterligare.

Utvecklingen av de bibliometriska teknikerna har gått framåt de senaste tjugo åren. Till stor del sammanfaller det med bildandet av bibliometriska forskargrupper runt om i världen, för Europas del ungefär vid mitten av

1980-talet. Numera anordnas minst en större internationell konferens kring bibliometriska frågor varje år och det finns ett antal tidskrifter som regelmässigt publicerar bibliometriska studier. Scientometrics, Research Evaluation och JASIS&T är några exempel. Området har till och med ett särskilt pris som delas ut till de mest framstående forskarna – Derek John De Solla Price Award. Av de tio mest citerade författarna i tidskriften Scientometrics har samtliga fått detta pris – det säger något om sambandet mellan citat och kollegial värdering (Tabell 1)

Tabell 1. De författare som citerats mest av tidskriften Scientometrics under åren 1978-2004
Notera: Price fick naturligtvis inte sitt eget pris. Pristagarna utses genom slutna omröstning.

Citerad forskare	Citerad i antal artiklar
Price D	427
Garfield E	373
Narin F	270
Schubert A	239
Small H	219
Braun T	192
Moed H	188
Egghe L	156
Glanzel W	143
Leydesdorff L	121
Vanraan A	121

Täckningsproblem kvarstår

Även om de bibliometriska teknikerna gått framåt så är det fortfarande i stort sett samma databas man studerar, dvs den som numera kallas Web of Science (WoS) och som innefattar de tre citeringsindexen SCI, SSCI och A&HCI. Fortfarande täcker WoS bara in cirka 8000 tidskrifter av det totala utbudet som kanske är dubbelt eller tredubbelt så stort. Täckningen har inte blivit särskilt mycket bättre under åren. En tydlig underrepresentation finns för icke-engelskspråkiga tidskrifter, företrädesvis inom samhällsvetenskap och hu-

maniora samt också nystartade tidskrifter som inte ackumulerat tillräckligt många citat från de tidskrifter som redan ingår i WoS. Så vad vi studerar när vi räknar papers och citat i WoS är olika aktörers visibilitet på ett, förvisso viktigt, segment av publiceringsmarknaden.

Vid många av våra universitet och högskolor byggs nu egna publiceringsdatabaser, där i stort sett allt som publicerats registreras, först som en bibliografisk databas och längre fram med länkar till hela texten. Då får vi en väsentligt större täckning av vad som skrivs. Men å andra sidan får vi inga citat och det är svårare att jämföra lärosäten med varandra, för att inte tala om länder.

Data kan rättas till

Vidare är det fortfarande svårt att identifiera publikationer från forskare enbart med ledning av deras namn, och någon tydlig koppling mellan författarfältet och adressfältet i posterna från WoS finns inte. Sedan ändrar ju forskare också namn, precis som vanliga människor. Så gör förresten också institutioner och forskargrupper, och även tidskrifter. Till och med universitet ändrar namn och ibland är det tveksamt vad som ingår i ett universitet och inte. Dylika fel och variationer bör tas i beaktande. Men de går att rätta till även om det ibland kan vara jobbigt. I mindre studier där enskilda forskare eller grupper jämförs kan de vara realistiskt att be forskarna om hjälp att identifiera relevanta arbeten. Detta underlättas av att de flesta svenska forskare själva har tillgång till WoS direkt via nätet.

Storskaliga studier

Även om det vi räknar ifrån inte blivit särskilt mycket bättre, så har sättet att räkna och bilda mått utvecklats. Det har i sin tur lett till bättre och mer varierade tolkningar av bibliometriska data. Det är nämligen nästan omöjligt att på ett tillfredsställande sätt fånga allt i ett enda mått. Ledande i denna utveckling har CWTS

i Leiden varit under ledning av Anthony Van Raan, och liknande metoder tillämpas nu av andra forskargrupper och även av det svenska Vetenskapsrådets analysavdelning.

Institute for Scientific Information (ISI) som bygger WoS har under en lång rad av år producerat publicerings- och citeringstatistik för länder och olika ämnesområden definierade via tidskriftskategorier. För Sveriges del har citeringsnivån sjunkit i dessa data, framförallt som en följd av en vikande citeringstrend inom flera stora medicinska områden. Inom andra områden har citatnivån legat konstant eller förbättras något.

Dessa data säger något om i vilken grad svenska forskare finns med i de mest citerade artiklarna. Så måste vi uttrycka oss, eftersom drygt 40 procent av de svenska artiklarna är samförfattade med andra länder idag, vilket är en fördubbling på tjugo år. Detta är en generell trend som gäller de flesta länder och som gör det allt svårare att bestämma vilket inflytande ett enskilt land har. Samma sak gäller på författarnivån där vi idag har i genomsnitt fyra författare per artikel.

En lösning på detta problem är att fraktionalisera såväl artiklarna som citaten till dem, beroende på hur många forskare som skrivit artiklarna, hur många institutioner eller länder som varit inblandade osv. I viss mån gynnas ensamvargar av ett sådant räkningsätt, då deras artiklar och citat tillfaller dem odelat. Å andra sidan rättar fraktionaliseringen till den tilltagande författarinflationen i artiklarna. Vad som är bäst att göra kan diskuteras – och i ett sådant läge kan man välja att presentera såväl sk ”whole counts” som ”fractional counts”. Är skillnaden stor kan den indikera en stark samarbets effekt, som inte nödvändigtvis är ett tecken på svaghet.

Självcitater är svårare att eliminera eller hålla under kontroll. Numera kan självcitater elimineras i analyserna, åtminstone om man har full information om citerade och citerande

författare, dvs vem som citerar vem. Ofta är det så att självcitaten infaller kort tid efter publicering och kanske särskilt hos framgångsrika forskargrupper där ”hela gänget” bjuds in till författarlistan. Eliminering av självcitater erbjuder en möjlighet att renodla genomslaget utanför den närmaste gruppen. I Vetenskapsrådets nyligen publicerade studie, Hur mycket citeras svenska publikationer? tycks effekten av självciteringar vara märkbar på den långsiktiga citeringstrenden för Sverige. Den höga citeringsnivån Sverige hade i början av 1980-talet syns inte när självcitaten eliminerats. Den logiska slutsatsen, att vi var bra förr på grund att vi citerade oss själva mer än idag, framstår dock som substantiellt svårtolkad. Forskningens volym och internationalisering kan vara en orsak till att effekten av självcitater minskar, men varför skall det vara särskilt tydligt för Sveriges del? Är självcitater effekten beroende av delområde? Detta bör studeras närmare.

Det tar olika lång tid innan artiklarna citeras, på vissa områden går det snabbt medan det kan ta betydligt längre tid inom andra fält. Det är också så att en del områden är större och andra mindre, vilket påverkar den aktiva mängden citat som delas ut. Vissa typer av artiklar, t e x översikter (reviews) får mycket fler citat än notiser och sk genuina artiklar. För att skapa mer lika tävlingsförhållanden kan man jämföra lika med lika, dvs jämföra samma artikeltyp inom samma område inom samma tidsperiod. Vanligt är att man väljer en tidsperiod på tre år, dvs man startar under publiceringsåret och följer de tävlande artiklarna också under de två följande åren. I tävlingen ingår antingen bara de artiklar som kommer från samma tidskrift (tidskriftsnormering) eller inom samma grupp av tidskrifter (fältnormalisering). Citaten till den enskilda artikeln jämförs då t ex med motsvarande genomsnitt för fältet, där värdet 1.0 betyder att man ligger precis på fältnormen, medan värden över betyder att man är bättre än normen.

Det finns givetvis också invändningar mot dessa former av normalisering. T ex kan man misstänka att de som väljer att publicera på områden med låga citeringstal har det lättare att nå upp över genomsnittet och där slumpmässiga variationer också kan spela in. På områden med höga citeringstal krävs ett rejält antal citat för att komma upp till världssnittet och signifikant fler för att komma ännu längre upp. Tidsfönstret på tre år kan också diskuteras även om rangordningen tycks bestå tämligen orubbad efter fem år.

Småskaliga studier

De storskaliga studierna har ofta syftet att rangordna olika aktörer efter aktivitet och impact. Fältnormalisering verkar vara modet för dagen och bygger på tillgång till hela WoS med dess gigantiska mängd artiklar och citat. Men det finns andra och mindre arbetskrävande värderingsmetoder. Benchmarking är en annan metodik där en aktör kan jämföra sig med en eller några andra aktörer av liknande slag. En enskild forskare kanske vill jämföra sig med en annan inom samma snäva specialitet över en längre följd av år. Här har det sk Hirsch-indexet fått ett starkt genomslag, kanske mest för att det är så lätt att räkna fram. Om man söker igenom alla artiklar av en forskare i WoS och sedan sorterar dem i fallande ordning efter antalet citat, så kan man sedan leta sig fram till den artikel vars rangtal (nummerordning) är det samma som antalet citat. Vi kan då t ex säga att en forskare har 20 artiklar med minst 20 citat vilket också blir värdet på Hirsch-indexet.

Låt oss göra ett experiment inom biblioteks- och informationsvetenskap som var och en kan upprepa. Vi går in i WoS med startåret 1986 och letar fram alla artiklar i SSCI från Sverige, Norge, Finland och Danmark vilket gav totalt 65.644 source items. Av dessa klassades 1015 inom området "information science & library science". I Tabell 2 listas de

författare som skrivit flest papers. B Hjörland noteras för totalt 42 source items, bland dessa har 11 fått minst 11 citat. Rangordnar vi efter H-index så leder Hjörland knappt före Ingwersen, Jarvelin, Vakkari, Persson, Savolainen och Wormell. Det rör sig således om en relativt liten grupp som producerat ungefär en artikel per år i genomsnitt inom just detta tidskriftsområde (de kan ha publicerat en hel del annat högciterat). Det bör noteras att självcitater är inkluderade vilket kan påverka rankingen starkt. Vi räknar heller inte citat till böcker och andra publikationer, som också kan ha stor betydelse på detta fält.

Numera har WoS en analysknapp som leder fram till möjligheter att studera sökresultatet efter mest förekommande länder, institutioner, författare, ämnesområden, publiceringsår etc. På några sekunder har man svaret och det kan många gånger räcka för att besvara enkla frågor. Listar vi de 1015 papren ovan efter land så leder Danmark med 323 artiklar, följt av 288 för Finland och 266 för Sverige. Sådana listor kan också användas i sökningen där enskilda egenskaper kan bockas för och studeras vidare. Begränsar vi de 1015 artiklarna till sk genuina artiklar blir antalet artiklar per författare annorlunda (Se Tabell 2).

Tabell 2. H-index för de mest produktiva artikelförfattarna inom nordisk biblioteks- och informationsvetenskap 1986-2006.

Notera: När rangordning och citeringstal inte överensstämmer har ett intervall för H-indexet angivits

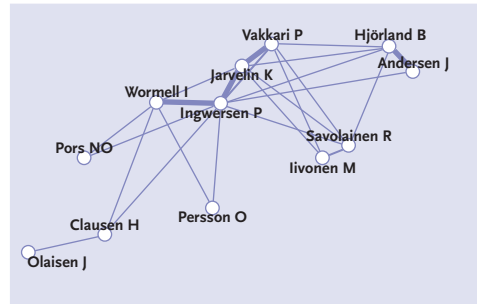
Författare	Antal source items	H-index baserat på alla source items	Antal genuina artiklar
Hjörland, B	42	11	17
Ingwersen, P	38	11 - 10	30
Jarvelin, K	34	9 - 8	34
Savolainen, R	23	7 - 6	18
Wormell, I	21	6 - 5	17
Vakkari, P	20	9	15
Persson, O	19	8	19

H-indexet är lättillgängligt men samtidigt lite bedrägligt, då det inte tar hänsyn till antalet författare per paper, självciteringar och det uppenbara faktum att många forskare publicerar sig utanför WoS täckning och där de också blir citerade. Google scholar och Scopus är två nya databaser som möjliggör en annorlunda och ibland kompletterande uppfångning av såväl artiklar som citat. Det kan vara av särskild vikt inom samhällsvetenskap och humaniora där WoS täckning är mycket begränsad. En rad jämförande studier pågår på olika håll och det som hittills rapporteras ger inget klart svar på frågan om dessa två nya databaser på allvar kan komma att konkurrera med WoS.

Visualisering av bibliometriska relationer

Artiklarna från WoS kan laddas ned och utsättas för en närmare analys. Programmet Bibexcel, som utarbetats av undertecknad och finns gratis tillgängligt på nätet, möjliggör en bearbetning av nedladdade poster för en rad olika analyser, som t ex bildandet av citeringslänkar, bibliografisk koppling och co-citeringar. Programmet, som är en verktygslåda för bibliometriska analyser, kräver dock en hel del träning för att fungera smidigt. En liten illustration till vad som Bibexcel kan åstadkomma följer här. Frågeställningen är hur de nordiska forskarna citerar varandra. Från de 1015 source items enligt ovan valdes 205 genuina artiklar skrivna av de mest produktiva författarna. Dessa laddades ned från WoS. Med Bibexcel framräknades citatlänkar mellan författarna. I Figur 1 ser vi vilka som är hopkopplade med varandra. Vakkari, Järvelin, Ingwersen och Wormell bildar en tydlig gruppering med starka band. Hjärland och Andersen är ett nära sammankopplat par, liksom Savolainen och Ivonon. De övriga är direkt eller indirekt kopplade till Ingwersen och Wormell. Den som känner fältet inser snart att citatkartan tämligen väl stämmer överens med forskarnas inriktning och interaktioner i övrigt.

Figur 1. Direkta citatlänkar mellan de författare som skrivit minst 10 artiklar.



Visualiseringar av det slag som presenteras i Figur 1 är ett mycket dynamiskt område inom bibliometrin. Kartor över samspelet i form av samförfattarskap, direkta citat, samciteringar mm kan studeras på olika nivåer, mellan forskare, universitet och länder. Kartan har en fördel i och med att den sammanfattar komplexa datastrukturer som oftast går att tolka på ett meningsfullt sätt. Men kartbildningen innebär också en reduktion av informationen till två dimensioner, vilket kan dölja eller snedvrider de faktiska relationerna. Kravet på exakta avbildningar är dock mindre om kartorna skall fungera som en första överblick och vägledning i en komplex informationsmängd.

En annan tydlig trend är att en stor mängd undersökningar görs av webben, dess tillväxt och länkstruktur. Då talar man om webometri, men det påminner starkt om bibliometrin eftersom webben är en struktur av hoplänkade dokument och som i princip låter sig studeras på samma sätt som citeringslänkarna i WoS. Teknikerna för att skapa webometriska data är annorlunda liksom kvaliteten på de sammanlänkade webbdokumentet.

Bibliometrin har verkligen slagit rot både inom och utanför informations- och biblioteksvetenskap. Vi kan därför förvänta oss många intressanta studier och avhandlingar på området, också i vårt eget land den närmaste tiden.