

På tværs af 'verdensvævet': om linkstrukturer og 'small-world'- fænomener på WWW

Af Lennart Björneborn

Hvordan ser WWW ud med milliarder af links, der væver websider sammen i komplekse strukturer? Hvor mange link-'hop' skal der til for at komme fra en 'ende' af WWW til en anden? Hvad kan linkstrukturer bruges til? Det er nogle af de spørgsmål, der behandles i artiklen.

Indledning

"Verdens basale vev, som hele tiden vever sig selv, fornyer sig selv, og er sammenknyttet i alle sine dele."

(Auerbach 1946: citeret hos Gundersen 1999)

Artiklen omhandler ideer fra forfatterens Ph.D.-projekt om linkstrukturer på World Wide Web, WWW. Linkstrukturer er et nyt forskningsfelt for biblioteks- og informationsvidenskab. Artiklen giver derfor også et bredere omrids af dette felt med nogle relevante forskningstilgange.

Med en fordanskning af det norske ord for WWW kan man tale om 'verdensvævet', med nærliggende poetiske og mytologiske konnotationer til oldnordiske norner, der sidder ved asketræet Yggdrasil og spinder menneskenes skæbnetråde. Det indledende citat er en norsk oversættelse af den tyske litteraturhistoriker Erich Auerbachs beskrivelse i *Mimesis* fra 1946 af, hvordan hele det shakespeariske skæbneunivers er sammenvævet i et kosmisk fortløbende skuespil. Gundersen (1999) bruger

citater som en allegori for WWW, og det indleder nærværende artikel, fordi det netop er 'selvvævende', 'selvfornyende' og 'sammenknyttende' fænomener ved nutidens verdensvæv, der skal behandles her.

I forlængelse af allegorien om 'verdensvævet' er der en smuk sproglig pointe i, at ordet 'tekst' etymologisk stammer fra det latinske 'textus', der betyder 'vævning' (jf. 'tekstil'). Verdensvævet, WWW, består jo netop af *tekster* (bredt forstået) vævet sammen i større *kontekster*, sammenhænge, ved hjælp af *hypertekst*, dvs. spring mellem forskellige tekstelementer.

Med en bredt dækkende definition kan man sige, at biblioteks- og informationsvidenskab beskæftiger sig med, hvordan informationsressourcer og informationsstrukturer (af indbyrdes forbundne informationsressourcer) opbygges, formidles og bruges af forskellige aktører. Som et helt centralt genstandsfelt står dokumenter (i bred forstand: med indhold af tekst, grafik, lyd, video mm.) og relationer mellem dokumenter (fx citationer, bibliografiske koblinger). WWW udgør derfor et oplagt forskningsfelt for informationsvidenskab, da WWW er et dokumentnetværk, hvor dokumenter i form af websider er forbundet med hinanden ved hjælp af links i komplekse strukturer. Ofte bruges begreberne Internet og WWW i flæng, men her er det vigtigt at skelne, da Internet

er et netværk af computere, hvor WWW netop er et netværk af dokumenter.

Til forskel fra bibliotekers dokumentsystemer, så er WWW uden central kontrol i form af koordineret akquisition, indeksering og katalogisering. I stedet kan WWW opfattes som et økologisk system (se fx Pitkow 1997, Bøgh Andersen 1998, Huberman & Adamic 1999): som et komplekst selvorganiseret og multi-agent ('mange aktører') system, hvor websider og links oprettes, tilpasses og fjernes decentralt og dynamisk af millioner af aktører, som privatpersoner, forskere, institutioner, firmaer, mm. WWW er altså et globalt kollektivt informationssystem bestående af lokale individuelle input. I Ph.D.-projektet undersøges bl.a., hvordan webaktørers lokale handlinger utilsigtet kan få globale konsekvenser på WWW ved såkaldte 'small-world'-fænomener i form af forbavsende korte afstande mellem websider.

WWW er et informationssystem, der i stadig højere grad integrerer og præger al slags videnproduktion, -distribution og -konsumtion. For at forstå, hvad der særlig kendetegner WWW som informationssystem, kan man med en allitteration sige, at WWW er kendetegnet ved at være '3D': *distribueret, diversivt og dynamisk*. 'Distribueret' betyder, at materiale i form af tekst, grafik, lyd, video mm., ligger *fordelt* på millioner af Internet-servere. 'Diversivt' indebærer, at indholdet på websider afspejler menneskets hele *mangfoldige* register af interesseområder: fra reklame, porno og nazipropaganda til videnskabeligt materiale. Det sidste 'D' for 'dynamisk' står for den hurtige *udskiftning* og gensidige *tilpasning* af websideindhold og links på tværs af WWW.

Samtidig stiger antallet websider og links på WWW eksponentielt. Med en forsigtig fremskrivning af skøn fra Lawrence & Giles (1999) og Broder et al. (2000) kan man (her i april 2001) estimere det såkaldt indekserbare WWW - 'indexable web' - til godt 2 milliarder websider og godt 15 milliarder links. 'Ned under' denne 'overflade'-web ligger desuden hele den 'dybe' web bestående af måske over 500 milliarder websider (Bergman 2000). Den 'dybe' web består af alt det materiale, der indtil nu ikke er blevet indekseret i søgemaskiner, fordi det fx kun genereres dyna-

misk ved brugerforespørgsler på websteder med materiale lagret i database-form, fx artikelarkiver og online-kataloger.

Hele den enorme dynamik og kompleksitet, der er beskrevet ovenfor, gør at vores billede af WWW nok altid vil være ufuldstændigt. Men netop derfor indebærer WWW en spændende udfordring for forskere inden for informationsvidenskab og andre discipliner for alligevel at søge at udforske, analysere og forstå delaspekter af dette ekspanderende informationsrum bestående af forskellige aktører, computerprogrammer, dokumentindhold og dokumentsammenkoblinger.

Muligheden for at lave - og følge - links er selve essensen i et hypertextsystem, og er den faktor, der mest af alt kendetegner WWW som informationssystem. I Ph.D.-projektet er der fokus på linkstrukturer på WWW. Linkstrukturer opstår ved at links tilføjes, tilpasses og fjernes (eller glemmes som 'døde links') mellem websider i takt med, at webaktører får ny viden og ideer, samt får kontakt med nye aktører. Dette komplekst sammenflettede og dynamisk adapterende linkvæv afspejler dermed både kognitive og sociale strukturer hos webaktørerne. I forlængelse af dette kunne man måske, med en omskrivning af Descartes' berømte sentens '*cogito ergo sum*' ('jeg tænker, altså er jeg til'), fremsætte følgende motto for det frembrydende netværkssamfund, som WWW jo er en fremtrædende eksponent for: '*connecto ergo sum*' - 'jeg forbinder, altså er jeg til' (Bjørneborn 1998, 2000a).

Der er en tæt sammenhæng mellem struktur og funktionalitet i et informationssystem. Måden hvorpå et informationssystem er opbygget påvirker jo de forskellige muligheder, som det kan udnyttes på. På WWW kan linkstrukturer for eksempel udnyttes af brugere, der navigerer målrettet efter bestemte oplysninger, men også af brugere, der 'browser' eller 'surfer' mere umålrettet, nysgerrigt og intuitivt. I Ph.D.-projektet tales der i denne forbindelse om komplementære sammenhænge mellem såkaldt konvergente og divergente linkstrukturer på den ene side - og konvergent og divergent brugeradfærd på den anden - men mere herom senere.

Omrids af Ph.D.-projektet.

I Ph.D.-projektet søges udarbejdet en teoretisk referenceramme og modeller med mulige implikationer for, hvordan linkstrukturer på WWW kan udnyttes med henblik på såkaldt 'knowledge discovery', hvor man opdager nye sammenhænge mellem allerede eksisterende viden. Ph.d.-projektet har den engelsksprogede, foreløbige arbejdstitel: *'Small-world' linkage and co-linkage: Outlining a conceptual framework for link structures, 'small-world' phenomena and knowledge discovery on the Web.*

I Ph.D.-projektet inddrages ideer fra webometri (bibliometri omsat til WWW), hypertextteori, 'Knowledge Discovery in Databases' (herunder 'data mining' og 'text mining'), grafteori og social netværksanalyse med særligt fokus på såkaldte 'small-world'-fænomener, der som tidligere nævnt handler om forbavsende korte afstande mellem objekter i mange typer af netværk. Fælles berøringspunkter mellem disse forskellige forskningstilgange er bl.a. deres fokus på konstruktion, kortlægning, analyse og visualisering af netværksrelationer. I projektet underbygges og illustreres den teoretiske referenceramme ved hjælp af case-studier om linkstrukturer i forbindelse med forskeres personlige websider. Især er der fokus på såkaldte 'bookmark'-lister, som mange forskere lægger ud offentligt tilgængelige på WWW, så andre kan se, hvilke websider, som forskeren har besøgt og fundet værd at lave et 'bogmærke' for. Disse 'bookmark'-lister giver scientometrikere nye muligheder for at afdække fx frembrydende 'research fronts' eller 'invisible colleges' (Crane 1972) i form af uformelle forskernetværk.

Bag Ph.D.-projektet ligger flere års interesse i spørgsmål som:

- hvordan ser WWW ud med milliarder af links, der væver websider sammen i komplekse strukturer?
- hvor mange link-'hop' skal der til for at komme fra en 'ende' af WWW til en anden?
- hvordan påvirker forskellige typer af linkstrukturer menneskelige brugere og digitale agents muligheder for at bevæge sig rundt på WWW - og fx finde interessefællesskaber mellem ligesindede eller opdage uventet

men brugbar information på tværs af emneområder?

Disse spørgsmål indgår i det helt overordnede spørgsmål, der er drivkraften i Ph.D.-projektet:

- hvordan påvirker den diversive og selvorganiserede opbygning af et informationssystem som WWW muligheder for, at man kan opdage uventede relationer og sammenhænge mellem viden, der er nedfældet i informationssystemet?

Dette er en anderledes tilgang til biblioteks- og informationsvidenskab, der traditionelt har beskæftiget sig med, hvordan informationssystemer skal være opbygget for at understøtte informationsgenfindning, 'recovery', og ikke så meget, hvordan informationssystemer kan understøtte informationsopdagelse, 'discovery'. Denne opdeling i 'recovery' og 'discovery' er hentet hos Garfield (1986). Disse to tilgange udelukker ikke hinanden, men bør ses som komplementære. Websider fundet efter målrettede (i Ph.D.-projektets sprogbrug: 'konvergente') søgninger kan fx bruges som udgangspunkt for mere intuitive og kreative ('divergente') udforskninger og 'surf'-ture i WWW's linkstrukturer, hvor man ved uforudsigelige 'smuthuller' måske finder uventet men brugbar information, der igen kan give anledning til nye målrettede søgninger. Brugere kan skifte mellem konvergent og divergent adfærd, og blande dem i forskellige styrkeforhold, afhængig af behov og muligheder, der gør sig gældende undervejs i brugerens kontakt med 'informationsrummet'.

Her kan det være på sin plads at fremhæve, at informationsvidenskab netop betoner - til forskel fra teorier om massekommunikation - brugerens muligheder for individuelt at søge og få adgang til information, inkl. alternative kilder. Med Hjørland (1995:129-130) kan man sige, at brugeren bør understøttes i at kunne agere som forsker. I Ph.D.-projektets optik betyder det, at informationsvidenskab handler om, at brugeren bør hjælpes med aktivt at kunne udforske og udnytte muligheder indlejret i informationssystemer. Disse muligheder drejer sig bl.a. om, hvordan man kan bevæge sig både konvergent og divergent gennem et informationssystem og påtræffe både forventelig og uventet information undervejs.

I Ph.D.-projektet er der især fokus på divergente ('spredende') linkstrukturer i form af såkaldt transversale links, der går på tværs af forskelligartede emneområder, der hver især består af konvergente, ('samlende') linkstrukturer i form af webklynger, som fx interessefællesskaber og emnespecifikke portaler. Transversale links kan påvirke muligheder for såkaldt serendipitet ('serendipity'), dvs. uventede informationsopdagelser, der sker, når man snubler over brugbar information, som man ikke bevidst har ledt efter - og som man måske ikke engang ved, at man har brug for, før man støder på den. Serendipitet kan opstå i forbindelse med skæve fund i en søgemaskine, eller avisartikler om nye fænomener, noget interessant på modsatte side af opslag i leksika, en spændende bog ved græsning i bogreoler. Serendipitet sker hyppigere, end mange nok forestiller sig. Man kan hævde, at det er en forudsigelig konsekvens af, at mennesker færdes i en uforudsigelig verden (Björneborn 2000a).

Artiklen er disponeret på følgende vis. Indledningsvis skitseres Ph.D.-projektets placering i en historisk kontekst, hvor WWW er et nyt led i rækken af menneskets redskaber, der er udviklet til at formidle viden på tværs af rum og tid. I forlængelse heraf berøres kortfattet Vannevar Bush's vision fra 1945 om et associationsbaseret informationssystem 'Memex', der har været en stor inspirationskilde for såvel forfatteren som for udviklerne af hypertext, Internet og WWW. I Ph.D.-projektet bruges begreber og forskningstilgange, der nok er ret nye og uvante for de fleste af læserne. Derfor forklares bl.a. det basale begrebsapparat omkring linkstrukturer. Der gives en metafor om WWW som sammenkrøllet papir, der kan gøre det nemmere for læseren at visualisere WWW og forstå ræsonnementer omkring ideer i Ph.D.-projektet. Derefter gives et bredere omrids af nogle forskningstilgange, der er relevante for det forskningsfelt om linkstrukturer, som Ph.D.-projektet indgår i. Det drejer sig især om webometri, hypertextteori, social netværksanalyse, grafteori, 'knowledge discovery' og social informatik. I dette afsnit forklares også Ph.D.-projektets centrale begreber: transversale links, 'small-world'-fænomener og co-linknings-kæder. Sidstnævnte begreb bruges afslutningsvis i et casestudie, der indgår i Ph.D.-projektet.

Rids af historisk kontekst

"Nye teknologier forandrer strukturen i vore interesser: de ting, vi tænker over. De forandrer karakteren af vore symboler: de ting, vi tænker med."

(Postman 1993:38)

Mennesker har op gennem historien udviklet informations- og kommunikationsteknologier som tale, skrift, bogtryk, telekommunikation, Internet, og nu senest WWW, for bl.a. bedre at kunne tackle væksten i den akkumulerede videnmængde og for at skabe bedre muligheder for videndeling på tværs af tid og rum. Skriften opstod fx for at kunne håndtere og gemme oplysninger om bl.a. stigende høstudbytte, lagerbeholdninger og handelstransaktioner. Men selv om hver ny informations- og kommunikationsteknologi altså bl.a. har vundet indpas for at gøre det nemmere at håndtere den voksende mængde viden, der er skabt af foregående teknologier, så medfører den ny teknologi endnu mere vækst i videnmængden...! WWW blev eksempelvis udviklet for at globalt spredte forskere knyttet til det europæiske center for partikelfysik, CERN, skulle få bedre overblik og muligheder for informationsadgang og -udveksling (Berners-Lee 1989/1990). Idag kan vi så se, hvordan den eksponentielt voksende mængde materiale på WWW er med til at give 'information overload'.

Informations- og kommunikationsteknologier kan altså opfattes som forandringsagenter, 'change agents', på godt og ondt, der op gennem historien har ændret menneskers handlemuligheder kognitivt, socialt, økonomisk, politisk, mm. (jf. Harasim 1993). Det er meget svært at forudse konsekvenserne af en ny informations- og kommunikationsteknologi. Bogtrykkerkunstens betydning for udviklingen af videnskab, nationalsprog, nationalstat, skolevæsen og en borgerlig offentlighed var fx ikke noget som Gutenberg kunne forudse. Eller som Gotved (1997) udtrykker det i forhold til Internet:

"Teknologien er ikke færdigopfundet, når den forlader samlebåndet; den er både mål og middel i en stadig proces af gen-opfindelse, hvor dens potentialer for medierede handlinger hele

tiden udforskes og strækkes længere ud, end nogen fra starten kunne forudsige.”

Denne stadige genopfindelse gør sig altså bl.a. gældende for udviklingen i brugen af computere og computernetværk. Inspireret af forskerne bag det såkaldte Interspace-projekt (1998), kan man se en udvikling fra, at formålet med de første computere og netværk var *'data transmission'* - til at der kom fokus på *'information retrieval'* - og nu videre til at computere og netværk også kan bruges til *'knowledge discovery'*, hvor enorme data-mængder, fx om brugeradfærd på WWW eller om linkstrukturer, kan udnyttes og analyseres for at opdage nye mønstre og sammenhænge.

Disse udviklingstendenser præger WWW, hvor millioner af aktørers ideer og handlinger hele tiden skaber nye potentialer for, hvordan WWW som en ny informations- og kommunikationsteknologi kan bruges til videndeling og videnudvikling. I denne kontekst søges i Ph.D.-projektet afdækket nogle forholdsvis upåagtede sammenhænge mellem nogle af disse tendenser og potentialer i form af forskeres 'bookmark'-lister, transversale links, 'small-world'-fænomener, serendipitet, 'knowledge discovery', mm. Denne vinkel på Ph.D.-projektet er, som tidligere nævnt, inspireret af ideer, som amerikaneren Vannevar Bush publicerede allerede i 1945. Bush's ideer er også vigtige for at forstå både baggrunden og potentialerne for WWW.

Vannevar Bush's 'Memex'-vision

Ideer om kreativitetsstimulering og videnudvikling ved hjælp af informationssystemer optog Vannevar Bush (1890-1974), der under 2. verdenskrig var chef for amerikanernes krigsforskningsprogram (atombombe, radar, mm.). Her oplevede han den enorme synergi, der opstod, da tusindvis af forskere samarbejdede på tværs af videnskabelige skel. Bush (1945) mente, at måden, hvorpå forskningsbiblioteker organiserede publiceret viden i stive klassifikationshierarkier, hæmmede videnskabelig innovation og udvikling, fordi traditionelle indekseringsmetoder ikke støttede udforskning af sammenhænge på tværs af klassifikationshierarkierne. I stedet var der brug for mere associative og idéskabende sammenkoblinger, ligesom i menneskets hjerne, mente Bush. I artiklen 'As we may

think', publiceret i *The Atlantic Monthly* i juli 1945, beskrev han en skrivebordslignende maskine 'Memex' (for 'memory extender', dvs. 'hukommelsesudvider'), hvor forskere skulle kunne lave såkaldte 'trails' (med nutidens sprogbrug: links) mellem forskellige tekstafsnit lagret på mikrofilm. Disse associative 'trails' skulle kunne udveksles mellem forskere og hjælpe dem med at udforske sammenhænge på tværs af de gængse emnehierarkier. Nogle korte centrale citater fra artiklen kan illustrere Bush's ideer:

“any item can be joined into numerous trails”
... “gathered together to form a new book” ...
“builds a trail of his interest through the maze of materials available to him” ... “new forms of encyclopedias will appear, ready-made with a mesh of associative trails” ... “new profession of trail blazers” (Bush 1945).

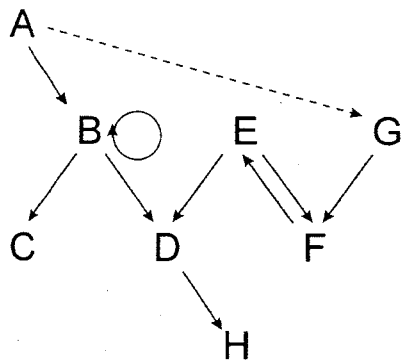
Selv om Bush aldrig fik realiseret sit 'Memex'-projekt, blev hans ideer en stor inspirationskilde for udviklerne af hypertext, Internet og WWW (se fx Simpson et al. 1996).

Basalt begrebsapparat mm.

Før vi kommer nærmere ind på relevante forskningstilgange omkring linkstrukturer, skal her gennemgås nogle basale begreber, en 'butterfly'-model og en metafor, der alle kan hjælpe læseren med at forstå linkstrukturer på WWW.

Linktopologi

Topologi er læren om sammenhængsforhold i geometriske figurer, som fx kurver, flader og netværk. Man kan fx tale om topologien i et landskab. Man kan også tale om *linktopologi*, når man vil beskrive, hvordan WWW bogstavelig hænger sammen i linkstrukturer (fx Gibson et al. 1998). I Ph.D.-projektet opereres med et linktopologisk begrebsapparat, der kortfattet skal beskrives her ved hjælp af omstående figur.



- B har et *udlink* til C: svarer til en reference
- B har et *indlink* fra A: citation
- B har et *selvlink*: selvcitation
- A er *ulinket*: mangler indlinks
- C er *ulinkende*: mangler udlinks
- E og F er *gensidigt* linkende
- A er *transitivt* linket med H via B og D
- A har et *transversalt link* til G: genvej 'på tværs'
- C og D er *co-linkede* fra B: svarer til co-citation
- B og E er *co-linkende* til D: bibliografisk kobling

Fig. 1. Simple linktopologi.

'Butterfly'-modellen

WWW udgør ikke et fuldstændigt sammenhængende netværk. Man kan ikke bevæge sig fuldstændig frit via mellemlid af links fra enhver webside til enhver anden webside, fordi der nogle steder mangler udlinks og/eller indlinks. WWW kan beskrives som en 'butterfly' - 'bow tie' på engelsk - der er tegnet i nedenstående figur baseret på Broder et al. (2000). 'Butterfly'-modellen kan bruges til at forstå, hvordan WWW's linktopologi hænger sammen og dermed påvirker muligheder for at mennesker, søgerobotter mm. kan bevæge sig rundt på WWW.

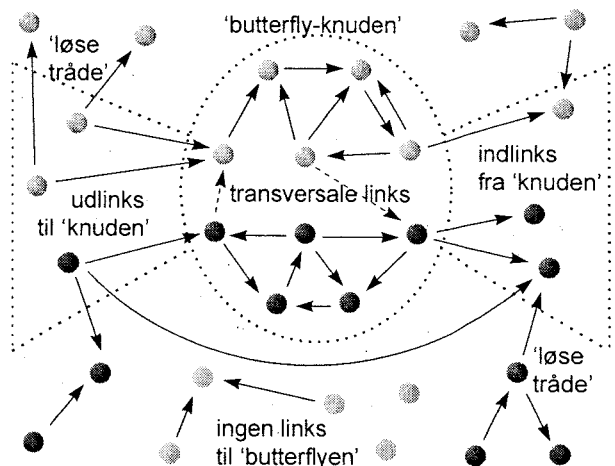


Fig. 2. 'Butterfly'-modellen for WWW baseret på Broder et al. (2000).

'Butterfly-knuden' udgøres af den største såkaldt 'strongly connected component' (grafteoretisk udtryk) af websider, hvor man kan bevæge sig frit mellem vilkårligt udvalgte par af websider via 'stier' af mellemliggende ind- og udlinks (jf. pilene i figurens 'knude'). Ifølge Broder et al. (2000) udgør denne 'knude' ca. 28% af WWW. Den venstre del af 'butterflyen' består af de ca. 21% websider med udlinks til - men ingen indlinks fra - websider inde i 'knuden'. Det drejer sig fx om nye websider, der endnu ikke har fået indlinks. Den højre del af 'butterflyen' udgøres af de ligeledes ca. 21% websider med indlinks fra - men ingen udlinks til - websider i 'knuden'. Her kan det fx være kommercielle websider, der fungerer som 'blindgyder', fordi man vil holde på besøgere uden at give dem mulighed for at surfe videre til andre websteder... Uden for 'butterflyen' ligger de ca. 22% websider, der ikke har direkte kontakt med 'knuden', men er forbundet som 'løse tråde' ('tendrils') med yderdelene via udlinks eller indlinks. WWW indeholder desuden ca. 8% websider, der ligger som isolerede øer uden ind- eller udlinks til og fra 'butterflyen'. Bemærk, at figuren ikke afspejler de faktiske procentuelle fordelinger mellem de forskellige typer websider. I figuren er indføjret to transversale links, der skaber bevægelsesmuligheder på tværs af to webklynger, markeret med forskellige grånuancer.

WWW som sammenkrøllet papir

For nemmere at kunne forestille sig WWW's link-topologi og forstå ræsonnementer omkring ideer i Ph.D.-projektet, kan man bruge en metafor, hvor WWW lignes ved et kompakt sammenkrøllet papir. Som tankeeksperiment kan man forestille sig, at den største sammenhængende del af WWW, 'butterflyknuden', bestående af for nærværende godt 500 millioner websider og godt 3,5 milliarder links (baseret på Broder et al. 2000), er spredt ud over et fladt papir. I tankeeksperimentet har hver webside indledningsvis kun ind- og udlinks fra og til andre websider, der ligger emnemæssigt tæt på og derfor i tankeeksperimentet ligger i nærheden på det flade papir. Hvis man nu vil bevæge sig fra et sted på det flade papir til et emnemæssigt fjernt sted på papiret, er man nødt til møjsommeligt at bevæge sig fra webside til webside, hvor man kun gradvist kan skifte emne undervejs. Den korteste rute gennem det flade papirs netværk ville her kunne tage tusindvis af link-'hop'.

Hvad sker der nu, hvis man i tankeeksperimentet finder et punkt i den ene ende af papiret og trækker dette punkt over mod et punkt i modsat ende af papiret? Det kunne fx være en webside for en forsker i informationsvidenskab, der har lavet et transversalt link til et andet af hans interesseområder, nemlig kreativitetsforskning. Papiret får nu et krøl, når den valgte webside om informationsvidenskab (plus alle de tilstødende 'naboskaber' af websider og websteder) bliver trukket hen til websiden om kreativitetsforskning. I tankeeksperimentet kan man nu forestille sig andre områder på papiret, der også bliver trukket sammen og får berøring med hinanden. Resultatet bliver en kompakt sammenkrøllet struktur, hvor der er korte afstande mellem punkter. Dette kan sammenlignes med, at der ifølge Broder et al. (2000) i snit kun er ca. 16 link-'hop' mellem to vilkårlige websider inde i 'butterflyknuden'.

Et punkt på det sammenkrøllede papir, hvor to forskellige områder får berøring med hinanden, svarer til et transversalt link på WWW. Hvor optræder så disse 'sammenkrøllende' emne-spring på WWW? Et forskningsspørgsmål i Ph.D.-projektet handler derfor om, hvordan man kan identificere og lokalisere sådanne transversale links, der går på tværs af emneområder og skaber

korte afstande på WWW - men mere herom senere.

Ligesom andre metaforer har også denne metafor sine begrænsninger. WWW er ikke tredimensionelt, men snarere milliard-dimensionelt, da hvert nyt link på WWW bogstavelig kan gå på tværs af allerede eksisterende linkstrukturer. Det svarer til, at et punkt på ydersiden af 'kuglen' af hårdt sammenkrøllet papir skulle kunne få direkte berøring med et punkt et sted midt inde i 'kuglen'. Denne manøvre kan ikke lade sig gøre i vores begrænsede tredimensionelle verden, men kan godt udføres i et mangedimensionelt matematisk-abstrakt rum. Da vores hjerner har svært ved at visualisere mere end tre dimensioner, er det nemmere her at bruge det kompakt sammenkrøllede papir som anskueliggørende metafor for WWW.

Omrids af relevante forskningstilgange og begreber

I dette afsnit gives et bredere omrids af nogle forskningstilgange, der er relevante for det forskningsfelt om linkstrukturer, som Ph.D.-projektet indgår i. Det drejer sig især om webometri, hypertextteori, social netværksanalyse, grafteori, 'knowledge discovery' og social informatik. I afsnittet forklares også Ph.D.-projektets centrale begreber: transversale links, 'small-world'-fænomener og co-linknings-kæder. Sidstnævnte begreb bruges i et efterfølgende afsnit om et casestudie, der indgår i Ph.D.-projektet.

Webometri og hypertextteori

Det er nærliggende at overføre traditionelle bibliometriske og informetriske undersøgelsesmetoder som publikationsanalyse, citationsanalyse og klyngeanalyse til WWW, hvor udlinks kan opfattes som analoge med referencer og indlinks med citationer. Almind & Ingwersen (1997) danner betegnelsen 'webometri' om denne forskningstilgang. Webometri defineres i Ph.D.-projektet som *studiet af kvantitative aspekter ved konstruktion og brug af informationsressourcer og -strukturer på WWW, med brug af bibliometrisk/informetriske baserede metoder*. Der kan udskilles tre overordnede undersøgelsesfelter for webometri: (1) websiders indhold, (2) linkstrukturer, (3) bruges informationsadfærd (via søgemaskiner og browsing).

I Ph.D.-projektets fokus på linkstrukturer bruges metoder fra citationsanalyse, hvor begreber som co-citation og bibliografisk kobling overføres til WWW. Denne webometriske tilgang uddybes i afsnittet om 'co-linknings-kæder'.

Foruden tilknytningen til traditionel bibliometri, har webometri også berøringsflader med hypertextteori. Det drejer sig dels om mere litterære tilgange i hypertextteori, som fx hos Bolter (1991), Landow (1997) og Bernstein (1998), dels om mere teknologiske tilgange som fx hos Conklin (1987), Botafogo & Shneiderman (1991) og Agosti & Melucci (2000).

Ted Nelson skabte begrebet 'hypertext' i 1960'erne for at beskrive et dokumentsystem med elektronisk understøttede spring mellem forskellige tekstelementer. Op gennem 1970'erne og 1980'erne udvikledes forskellige eksperimentelle og kommercielle hypertextsystemer (se fx Rada 1991), der gik forud for WWW. Der forskes stadig i udvikling af hypertextsystemer, der ikke nødvendigvis er webbaserede, men fx kan bruges i cd-rom-baserede opslagsværker. Fra dette store forskningsfelt, der indbefatter både den litterært og teknologisk orienterede hypertextteori kan hentes værdifuld viden, der kan berige webometri i almindelighed og Ph.D.-projektet i særdeleshed i forhold til at forstå forskellige typer af linkstrukturer.

Der er selvfølgelig forskel på, hvorfor akademiske forfattere laver referencer og hvorfor en mangfoldighed af forskellige typer webforfattere laver links. Man kan sige, at nogle links repræsenterer forskellige sociale attituder, interesser og præferencer hos webforfattere, hvor forskellige motiver ligger til grund for oprettelsen af et link til et givet webdokument. Indlinks og udlinks kan her repræsentere mere eller mindre bevidste fællesskabsorienteringer på WWW. Andre links fungerer i stedet som rene navigationsredskaber internt på et websted.

Ligesom en reference i en akademisk tekst kan et udlink udgøre en implicit anbefaling (af den linkede webside). Men ligesom for akademiske referencer er dette ikke altid sandt. Man kan fx forestille sig negativ omtale af nazipropaganda og bør-

neporno, underbygget med links til de omtalte websteder. Der mangler stadig forskning om webforfatters linkmotiver - hvorfor man laver links - men gode tilløb kan findes hos fx Haas & Grams (1998) og Kim (2000).

En vigtig forskel mellem papirverden og WWW er, at referencer til trykte dokumenter akkumuleres over tid, hvorimod links til webdokumenter skifter dynamisk over tid: links oprettes, tilpasses og slettes (jf. Koehler 1999).

Ligesom nogle artikler, forfattere og institutioner citeres oftere end andre i papirverdenens bøger, tidsskrifter mm., 'klumper' links sig omkring bestemte websider og websteder, og der dannes webklynger af sammenkoblede websider og websteder svarende til papirverdenens 'citation clusters'. Webklynger repræsenterer typisk beslægtede emneområder og interessefællesskaber (se fx Chakrabarti et al. 1999). Som tidligere nævnt, opfattes derfor webklynger i Ph.D.-projektet som konvergente ('samlende') linkstrukturer. På grund af at links tilføjes, tilpasses og fjernes, så ændrer webklynger sig dynamisk over tid, hvor center/periferi-konstellationer også kan ændres. Man kan også forestille sig, at der på WWW findes 'meta'-klynger inden for beslægtede emneområder bestående af mange 'under'-klynger af websider og websteder.

Linkstrukturer kan ses som implicite menneskelige annotationer, hvor den dynamiske tilpasning og udskiftning af links afspejler udvikling af menneskers viden og ideer. Denne "*annotative power*" (Chakrabarti et al. 1999), der ligger i millioner af webaktørers linktildelninger, kan søges indfanget i forskellige former for 'link structure analysis'. Det drejer sig fx om at forbedre søgemaskinens ranking-algoritmer (Brin & Page 1998), at kortlægge nærhed mellem emner ('topical proximity') (Bharat & Henzinger 1998), at afdække interessefællesskaber ('web communities') (Gibson, Kleinberg, Raghavan 1998) og at identificere autoritative websider (Kleinberg 1998).

Tim Berners-Lee, der var med til at udvikle World Wide Web som en slags hypertextbaseret intranet hos CERN omkring 1990, forudså muligheden for at analysere klynge- og linkstrukturer over tid

for at afdække komplekse dynamiske relationer mellem projekter, personer, dokumenter, ideer mm. i en organisation (Berners-Lee 1997).

Idag er der udviklet forskellige webometriske metoder til analyse af linkstrukturer og webklynger, hvorved 'kort' over fx relationer inden for og mellem faglige domæner søges tegnet, som fx co-link-analyse brugt hos Larson (1996) og konnektivitetsanalyse hos Chen (1998). Bar-Ilan (2001) giver et godt overblik over udviklingen af forskellige metoder inden for webometri. Se også Dodge & Kitchin (2001) for suggestive 'kartografier' over udsnit af Internet og WWW.

Analyse af linkstrukturer bruges som før nævnt hos søgemaskiner som Google, der indarbejder såkaldt linkpopularitet i deres rankingalgoritmer (Brin & Page 1998). Linkpopularitet indebærer, at en webside med mange indlinks vægtes højere end en webside med færre indlinks. Desuden vægtes en webside højere, hvis den får indlinks fra en webside, der har fået mange indlinks.

Et metodeproblem inden for webometri er, hvordan datamateriale indsamles til brug for empiriske undersøgelser. Det er svært at indsamle repræsentative data, som der kan generaliseres ud fra. Dataindsamling er vanskelig dels pga. WWW's særlige distribuerede og dynamiske natur, dels manglende metadata og mangelfulde søgemaskiner. Sekundært datamateriale indsamlet via søgemaskiners databaser skal behandles med omtanke pga. søgemaskinernes uigennemsigtige dækningsgrad, opdateringshyppighed (inkl. 'kassation' af forældet data), indekseringsdybde, rankingalgoritmer og reproducerbarhed af søgeresultater (Snyder & Rosenbaum 1998, Bjørneborn & Ingwersen 2001).

I Ph.D.-projektet er afprøvet et antal pilottests i forbindelse med udvikling af metoder til lokalisering af transversale links. Her er bl.a. brugt primært datamateriale nedhentet direkte fra WWW, fx ved såkaldte 'random walks' (grafteoretisk udtryk), hvor links bliver fulgt på randomiseret vis fra webside til webside. For nærværende opereres med casestudier, hvor udlinks bliver fulgt på udvalgte forskeres websider, og hvor indlinks bliver fundet ved hjælp af AltaVista's specielle søge-

koder, hvor man kan søge efter websider, der har lavet links til en given webside. Dette beskrives nærmere i afsnittet om casestudiet.

Social netværksanalyse og grafteori

I Ph.D.-projektet sammentænkes ideer fra webometri og hypertextteori med ideer fra social netværksanalyse og grafteori. Netværksanalyse fokuserer på *relationer* mellem entiteter, såkaldte *noder*, typisk i form af sociale aktører. Disse relationer kan tegnes som linier (såkaldte '*edges*', dvs. kanter), mellem punkter (*noder*, eller '*vertices*', dvs. hjørner) i et netværk (også kaldet en graf). Social netværksanalyse (fx Wasserman & Faust 1994, Borell & Johansson 1996, Scott 2000) og dens matematiske 'hjælpedisciplin', grafteori (fx Bollobás 1998, Gross & Yellen 1999), opererer idag med et omfattende begrebsapparat, der gør den brugbar inden for en bred vifte af forskningsområder, fx socialpsykologi, sociologi, socialantropologi, kulturgeografi, økonomi og statskundskab. Problemfelter, der kan tackles netværksanalytisk, er fx opbygning af interessefællesskaber, det globale politiske og økonomiske system, social support, innovationsspredning, problemløsning i grupper og magtstrategier (se fx Wasserman & Faust 1994).

Det netværksanalytiske og grafteoretiske begrebsapparat er også velegnet at bruge på et dokumentnetværk som WWW, hvor *noder* i form af websider, websteder eller 'top level'-domæner (fx edu, com, dk) er forbundne med *relationer* i form af links i netværksstrukturer. I grafteoretisk forstand er WWW en såkaldt '*directed graph*', da links er retningsbestemte og derfor kan tegnes som pile. I en '*directed graph*' kan man foretage en '*directed walk*' langs en '*directed path*' ved at følge links i pilenes retning, fx når man vil finde den korteste vej mellem to punkter i netværket, grafen.

WWW kan altså betragtes som en link-graf. Forskere, der analyserer linkstrukturer på WWW, bruger i høj grad denne grafteoretiske tilgang (fx 'butterfly-modellen' hos Broder et al. 2000). En af de første forskere inden for bibliometri, der brugte denne tilgang, var Eugene Garfields bror Ralph Garner, der brugte grafteori til at analysere citationsmønstre (Garner 1967).

Inden for social netværksanalyse opereres med begreber som bl.a. 'weak ties' og 'small-world'-fænomener, der hænger sammen med transversale links, der beskrives i næste afsnit.

Transversale links

Begrebet 'transversale links' stammer fra Bjørneborn (2000b). 'Transversal' betyder 'på tværs' på latin og bruges i matematikken til at beskrive linier der går på tværs af geometriske figurer. Der er desuden en fin pointe i, at det latinske 'de transverso' betyder 'uventet', da Ph.D.-projektet netop handler om uventede konsekvenser af disse transversale links.

Som tidligere nævnt går transversale links på tværs af uligeartede webklynger, på tværs af videndømnener, emneområder og interessefællesskaber (se også Bjørneborn & Ingwersen 2001). En bruger, der surfer rundt på WWW ved at følge links fra webside til webside, kan via et enkelt transversalt link komme fra én webklynge til en anden emnemæssigt 'fjerntliggende' webklynge. Transversale links skaber uventede berøringsflader på tværs af forventede, 'normale' kontaktflader. Tidligere i artiklen, i afsnittet om WWW som sammenkrølet papir, blev der brugt et eksempel med en webside for en forsker i informationsvidenskab, der har lavet et transversalt link til et andet af hans interesseområder, nemlig kreativitetsforskning. Denne lokalt og uforudsigeligt konstruerede 'på-tværs-kobling' er med til at krølle WWW sammen og skabe korte afstande. I nedenstående figur er eksemplet illustreret i en forsimplet graf.

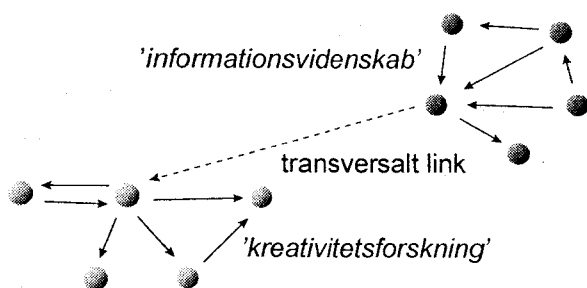


Fig. 3. Eksempel på transversalt link.

Man kan se lighedspunkter mellem transversale links og netværksanalysens 'weak ties'. Dette begreb, der stammer fra Granovetter (1973), handler om, at det ofte er via perifert, 'svagt' tilknyttede personer, såkaldt 'weak ties', at fx ideer hurtigt spredes mellem forskellige sociale grupper. Hvis man vil have anderledes oplysninger, skal man ikke spørge sine nærmeste venner og bekendte, såkaldte 'strong ties', da disse typisk får oplysninger fra samme kilder som en selv. Begrebet 'weak ties' bruges inden for diffusionsteori, der er en gren af social netværksanalyse, bl.a. til at analysere, hvordan innovation og ideer spredes i et samfund (se fx Rogers 1995).

Granovetter (1973) taler desuden om "*the strength of weak ties*", da han mener, at det er de svage personrelationer, der skaber sammenhængskraft i et samfund, der ellers ville være endnu mere opsplittet i isolerede grupperinger uden berøringsflader med hinanden.

I Ph.D.-projektet hævdes, at transversale links på lignende vis skaber sammenhængskraft til WWW ved at skabe berøringsflader mellem webklynger, der ellers ville ligge isoleret fra hinanden. Man kan sige, at 'weak ties' og transversale links handler om '*connecting the unconnected*'. Spørgsmålet er så, om transversale links - ligesom 'weak ties' - har vidensspredende og kreative potentialer, så man måske kunne tale om "*the strength of weak ties crossing the Web*" ...

Det begreb inden for informationsvidenskab, der måske kommer tættest på transversale links er 'boundary crossings' fra videnskabssociologisk forskning om interdisciplinaritet (Klein 1996, Pierce 1999). Ifølge Pierce (1999) citerer de fleste akademiske forfattere mindst nogle få kilder uden for deres egne videnskabelige discipliner, indimellem fra discipliner "*even further afield*". 'Boundary crossings' mellem discipliner '*further afield*' svarer til Ph.D.-projektets brug af begrebet 'transversale links'.

Hvordan afgører man, hvorvidt et link er transversalt? Hvordan måler man graden af uligeartethed? Og hvordan afgøres, hvilket emne en webside repræsenterer? Dette er slet ikke ukompliceret. Tværtimod er disse spørgsmål informationsvi-

denskabelige 'minefelter' med paralleller til problematikker omkring begreber som 'lighed' og 'emne' inden for klassifikations- og indekseringsteorier (se fx Hjørland 1997). Ligesom opfattelsen af et 'emne' kan skifte, så kan også opfattelsen af 'uligeartethed' skifte mellem forskellige brugere til forskellige tider i forskellige kulturer, der kan opfatte forskellige emner som værende indbyrdes uligeartede, 'transversale'. Uligeartethed af idag kan takket være tværvideenskabelig og tværkulturel udvikling og paradigmeskift blive til morgendagens normalitet ...!

I Ph.D.-projektet søges problematikken løst med mere objektive kriterier. I en metode, der afprøves for nærværende i projektet, får forskeres websider tilføjet ekstra 'metadata' hentet fra citationsdatabaser. 'Subject Category'-feltet for de tidsskrifter, som forskerne publicerer i, kan give et fingerpeg om, hvilke overordnede emner forskeren beskæftiger sig med. Lav co-citationsstyrke mellem forskere sammenholdt med oplysninger om 'subject category' kan måske bruges som indikator for uligeartethed.

Transversale links er tæt forbundet med såkaldte 'small-world'-fænomener på WWW, der behandles i næste afsnit.

'Small-world'-fænomener

Når mennesker på rejse langt hjemmefra møder personer, som de kender, eller hvor de har fælles bekendte, udbryder de typisk - hvis de taler dansk: "Hvor er verden lille!" - eller på engelsk: "My, it's a small world!". Begrebet 'small-world' bruges inden for social netværksanalyse til at beskrive korte afstande mellem vilkårlige mennesker i sociale netværk. Man kan spore begrebet tilbage til 1960'erne hos Manfred Kochen, kendt fra informationsvidenskab. Han skrev dog først om fænomenet i 1989 (Kochen 1989). Derimod publicerede den amerikanske socialpsykolog Stanley Milgram i 1967 en undersøgelse af, hvor mange mellemlid af bekendtes bekendte, der skulle til for at forbinde to vilkårlige personer med hinanden. Hans resultater (baserede på amerikanske forhold) pegede på, at der i snit skulle bruges 5-6 mellemlid (Milgram 1967). Senerehen er dette tematiseret i skuespillet 'Six degrees of separation' af John Guare fra 1990, filmatiseret i 1993

(Collins & Chow 1999). Samme idé gør sig gældende i computer-'quizen' 'Six degrees of Kevin Bacon' (www.cs.virginia.edu/oracle/), hvor man kan få oplyst den kortest mulige kæde af medspilleres medspillere mellem to vilkårlige filmskuespillere fra hele filmhistorien (databasen indeholder ca. 120.000 filmskuespillere).

'Small-world'-begrebet har i de sidste par år fået en genopblomstring takket være to forskere i anvendt matematik, Duncan Watts og Steven Strogatz. I en skelsættende artikel starter Watts & Strogatz (1998) med at analysere egenskaber hos to idealtyper af netværk. Den ene idealtipe består af det fuldstændigt *regulære* netværk, der har en høj såkaldt klyngekoefficient (dvs. de fleste punkter tæt på hinanden er forbundet med hinanden), men til gengæld er der i denne type netværk lang afstand mellem to vilkårlige punkter. I det fuldstændigt *randomiserede* netværk derimod, forholder det sig lige omvendt: lav klyngekoefficient (dvs. punkter tæt på hinanden er ofte ikke forbundet med hinanden) og korte afstande mellem to vilkårlige punkter (dvs. der skal ikke så mange link-'hop' til for at komme fra en ende af netværket til en anden). De to forskere fandt så, at hvis de tog et regulært, velordnet netværk og lavede ganske få forbindelser (ca. 1 % var tilstrækkeligt) om til 'long-range connections' (læs: transversale links) til tilfældige punkter i netværket, så fik dette netværk egenskaber fra begge idealtyperne: høj klyngekoefficient og korte afstande.

Watts & Strogatz bruger betegnelsen 'small-world'-netværk om denne type netværk, som de fandt repræsenteret i bl.a. biologiske netværk (fx hjernevæv), tekniske netværk (fx el-forsyninger) og sociale netværk (fx spredning af innovation og epidemiske sygdomme). Dette uddybes hos Watts (1999). I stedet for at man skal forbi en lang række af noder i det regulære netværk for at komme fra et sted til et andet, kan man i 'small-world'-netværket skyde genvej og springe noder over. 'Small-world'-fænomener handler altså om korte afstande via genveje ('short cuts') mellem noder i netværk.

Watts (1999) og Adamic (1999) beskriver WWW som et muligt 'small-world'-netværk. Den korteste afstand mellem to websider beregnes som det mindste antal nødvendige link-'hop' via mellem-

liggende websider. Da WWW som tidligere nævnt indeholder godt 2 milliarder websider og godt 15 milliarder links kræves det selvsagt enorm computerkraft at udregne disse afstande.

Som tidligere nævnt fandt Broder et al. (2000), at der inde i 'butterfly-knuden' var en gennemsnitlig afstand på ca. 16 link-'hop', hvorimod der kan være flere hundrede 'hop' mellem to vilkårlige websider på hele WWW.

Det kunne her være interessant at tage to websider fra forskellige videnskabelige områder og søge at finde den korteste vej mellem dem via links på WWW. En sådan fremgangsmåde, der som sagt kræver enorme men gennemførlige computerberegninger (jf. Broder et al. 2000), ville også afsløre, hvor der optræder transversale links undervejs. I Ph.D.-projektet søges derfor etableret adgang til en database med oplysninger om linkstrukturer. *Internet Archive* (www.archive.org) i Californien kunne være en sådan mulighed, da de jævnligt laver 'snapshots' af hele WWW og stiller materiale til rådighed for forskere. En anden mulighed kunne være *Connectivity Server*, hvis linkstruktur-data blev brugt til at lave den før nævnte 'butterfly'-model (Broder et al. 2000).

Inden for biblioteks- og informationsvidenskab mangler endnu forskning om, hvad 'small-world'-fænomener, i form af korte afstande, kan bruges til i informationsbaserede netværk: WWW, citationsdatabaser, semantiske netværk, tesauri, ...

Mulige forskningsområder kan aflæses i nedenstående figur, hvor forskellige typer *noder* i informationsbaserede netværk er listet op over for mulige afstandsforkortende, 'small-world'-skabende *relationer*. I figuren er relationerne ikke tegnet som pile, da ikke alle de oplyste relationer er retningsbestemte.

Hvis eksempelvis noderne i figuren repræsenterer dokumenter, så kan relationerne repræsentere bibliografiske koblinger mellem disse dokumenter. Der eksisterer en del litteratur om et andet eksempel fra figuren, nemlig hvor noderne svarer til forfattere og relationerne er medforfatterskaber. Der er lavet grafer over medforfatterskaber, der knytter sig til den ungarnske matematiker Paul Erdős. Denne var uhyre produktiv i løbet af sit lange liv (1913-1996), og skrev næsten 1500 artikler sammen med 472 forskellige medforfattere (Bar-Ilan 1998). Sådanne Erdős-numre beregnes som antallet af mellemliggende medforfattere

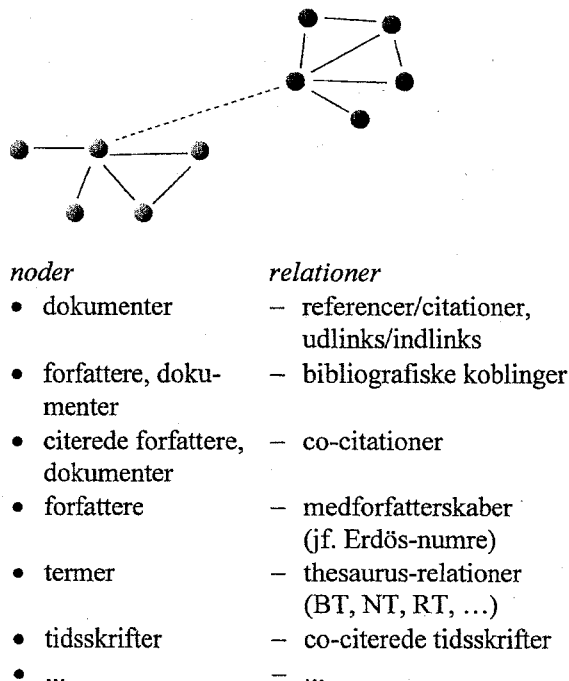


Fig. 4. Mulige forskningsområder for 'small-world'-fænomener i informationsbaserede netværk

mellem en given forsker og Erdös. En forsker har fx Erdös-numret 2, hvis han/hun har været medforfatter med nogen, der har været medforfatter med Erdös. Erdös selv har Erdös-nummer 0. Lister over Erdös-numre findes hos fx Grossman (2000). Newman (2001) viser, at medforfatterskaber giver 'small-world'-fænomener i form af korte afstande i forskernetværk af dataloger, medicinere og fysikere.

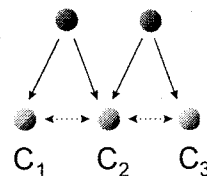
I Ph.D.-projektet fokuseres på de tre øverste mulige afstandsforkortende, 'small-world'-skabende relationer i figuren: links, bibliografiske koblinger og co-citationer. I næste afsnit skitseres en metode, der for nærværende afprøves som casestudie i projektet, med såkaldte 'co-linknings-kæder' bestående af bibliografiske koblinger og co-citationer mellem websider.

'Knowledge discovery' og co-linknings-kæder

I Ph.D.-projektet bruges casestudier til at underbygge og illustrere den teoretiske referenceramme. Casestudierne omfatter linkstrukturer og såkaldte 'co-linknings-kæder' ('co-linkage chains') i forbindelse med forskeres personlige websider. For at forstå begrebet 'co-linknings-kæder' skal her først forklares begrebet 'co-citations-kæder', der kan relateres til forskning inden for 'knowledge discovery' i databaser, der omfatter såkaldt 'data mining' og 'text mining'. Dette er et forskningsfelt tæt knyttet til datalogi, hvor man søger at udvikle metoder til at udtrække brugbare mønstre og sammenhænge fra enorme datamængder lagret i databaser (se fx Frawley et al. 1991, Fayyad et al. 1996, Qin & Norton 1999). Et omrids af dette felt med perspektiver for webometri er skitseret hos Björneborn & Ingwersen (2001).

Inden for såkaldt 'literature-based knowledge discovery', hvor man bl.a. bruger bibliografiske databaser og citationsdatabaser, er Don Swanson en central forsker. I den første af en lang række artikler om såkaldt 'undiscovered public knowledge' anfører Swanson (1986), at offentliggjort viden kan have uopdagede implicite relationer med anden offentliggjort viden. Det kan fx være litteratur fra to videnskabelige discipliner, der ikke er direkte co-citerede med hinanden, dvs. de figurerer ikke på samme litteraturlister, men hvor en tredje litteratur fungerer som uopdaget transitiv

(jf. fig. 1) relation mellem dem. Nedenfor er Swansons (1986) eget eksempel tegnet om til en lille graf i form af, hvad der i Ph.D.-projektets terminologi betegnes som en 'co-citations-kæde'.



C₁ = litteratur om fiskeolie
 C₂ = litteratur om blodplader
 C₃ = litteratur om Raynauds syndrom

Pilene står for citationer. De stiplede dobbeltrettede pile er co-citationer.

Fig. 5. Kort co-citations-kæde som illustration af Swanson (1986).

Her er C₁ videnskabelig litteratur om fiskeolie, C₂ litteratur om blodplader og C₃ litteratur om kredsløbssygdommen Raynauds syndrom. Som det fremgår af figuren, er C₁ co-citeret med C₂ (dvs. referencer til de to videnskabelige forskningsfelter figurerer på samme litteraturlister). C₂ er desuden co-citeret med C₃, hvorimod C₁ ikke er direkte co-citeret med C₃.

Swanson brugte en systematisk 'trial-and-error'-metode for at opdage denne litteraturbaserede transitive forbindelse mellem fiskeolie og Raynauds syndrom. Efterfølgende medicinsk forskning har vist, at fiskeolie faktisk har en gavnlig virkning over for 'døde tæer og fingre', der skyldes Raynauds syndrom. Dvs. Swansons litteraturbaserede 'knowledge discovery'-metode kunne bruges til at opdage "interesting but previously unknown implicit information" i videnskabelig litteratur (Swanson & Smalheiser 1999), hvor der afsløres relationer mellem ideer og begreber, der ikke er har været kendte før.

Small (1999) har udviklet Swansons idé til at omfatte flere led af transitive relationer i stedet for det ene transitive led, som Swanson har opereret med. I artiklen 'A passage through science: crossing disciplinary boundaries' kalder Small (1999) dette for "indirect multi-step co-citation", og i figuren nedenfor er hans eksempel tegnet som

en lang co-citations-kæde, der forbinder økonomi med astrofysik.

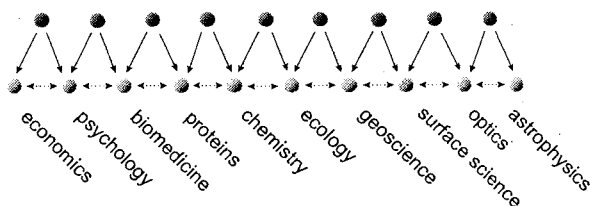


Fig. 6. Lang co-citations-kæde (stiplede dobbeltrettede pile) som illustration af Small (1999).

Ifølge Small (1999) er det muligt at bevæge sig via co-citationer fra enhver videnskabelig disciplin til en anden, takket være "the interconnected fabric of scientific disciplines". Til forskel fra Ph.D.-projektets fokus på 'weak ties' består Smalls lange co-citations-kæde af 'strong ties', hvor hvert led udgøres af stærke co-citationer, dvs. referencerne figurerer sammen på mange litteraturlister.

Her kunne det også være interessant at se på 'small-world'-fænomener i form af korte afstande via co-citationer mellem videnskabelige discipliner, fx mellem økonomi og biologi. Man kunne forestille sig, at der også eksisterede svage, transversale co-citationer mellem forskelligartede discipliner, der kunne 'klippe' nogle af leddene over i en ellers længere co-citations-kæde som vist i figuren ovenfor. Sådanne transversale 'boundary crossings' kunne have implikationer for 'knowledge discovery' på tværs af videnskabelige discipliner, fx i form af 'undiscovered public knowledge', som hos Swanson, hvor man søger at identificere nye frugtbare udforskningsområder.

I Ph.D.-projektet indgår disse ideer om 'knowledge discovery' i en kontekst af teorier om, hvordan informationssystemer kan bruges til at stimulere 'scientific discovery', der også omfatter videnskabelig kreativitet og innovation. Foruden Swanson (1986ff) kan her nævnes forskere som Bawden (1986), O'Connor (1988), Davies (1989) og Ford (1999). I denne kontekst indgår også teorier om serendipitet, uventede informationsopdagelser. Et beslægtet begreb er 'incidental information encountering' (Erdelez 2000, Toms 1998). I videnskabssociologi bruges udtrykket 'serendipitet' til at beskrive, hvordan mange videnskabelige opdagelser sker ved heldige slumpræf kombineret

med stor årvågenhed (se fx van Andel 1994, Björneborn 2000a, 2000b). Serendipitet afhænger af manglende 'perfektionisme' i et informationssystem, da inkonsistens og uforudsigelighed kan skabe et mulighedsrum med plads til 'smuthuller', uventede vinkler og anderledes 'access points' (jf. 'polyrepræsentation' hos Ingwersen 1996) og dermed større mulighed for, at man snubler over noget uventet. Serendipitet sker typisk ved en eksplorativ, browsende 'bevægelig' brugeradfærd (se fx Bates 1986, Chang & Rice 1993), der kan hjælpe forskere m.fl. til at opdage ting, som de ikke på forhånd ved, at de har behov for - og dermed heller ikke kan lave en forespørgsel til et informationssystem om. Transversale linkstrukturer kan ifølge en hypotese i Ph.D.-projektet understøtte sådan 'divergent' brugeradfærd og muligheder for serendipitet.

Man kan opfatte 'knowledge discovery' som computer-støttet 'systemiseret serendipitet' - med et udtryk fra Garfield (1986) - hvor man kombinerer menneskelig ekspertise, årvågenhed, nysgerrighed og intuition med computers råstyrke ved bearbejdning af store datamængder. Swansons 'trial-and-error'-metode er et eksempel på 'systemiseret serendipitet' med brug af både menneske og computer.

I Swansons (1986) og Smalls (1999) forskning om litteratur-baseret 'knowledge discovery' bruges co-citations-kæder baseret på materiale fra citationsdatabaser. I nogle casestudier, der skal indgå i Ph.D.-projektet, bruges i stedet co-linkningskæder, hvor noderne i kæden består af websider. En co-linknings-kæde, som den kaldes i Ph.D.-projektets terminologi (eller 'co-linkage chain') består af sammenkoblede co-linkede og co-linkende noder, analogt med co-citationer og bibliografiske koblinger, hvilket illustreres i nedenstående figur.

Noderne i figuren repræsenterer websider. I figuren er noden C_1 co-linket (co-citeret) med C_2 , da de begge får indlinks fra samme kildenode B_1 . C_2 er desuden co-linket med C_3 fra B_2 , og så videre. B_1 er co-linkende (bibliografisk koblet) med B_2 , da de begge giver udlinks til samme destinationsnode C_2 . B_2 er desuden co-linkende med B_3 , og fremdeles. For eksempel kunne B_1 være en 'bookmark'-

liste på en webside med udlinks til to forskeres personlige hjemmesider C_1 og C_2 . C_2 er også co-linket med en tredje forsker C_3 fra 'bookmark'-listen B_2 , etc.

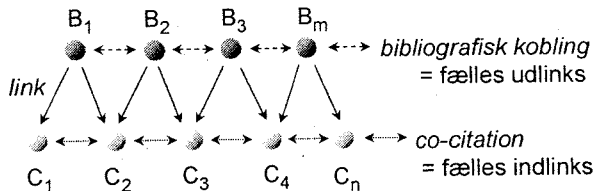


Fig. 7. Co-linknings-kæde mellem websider.

Som det fremgår af figuren er der en tæt sammenhæng mellem de to rækker af bibliografiske koblinger resp. co-citationer. De to rækker genererer så at sige hinanden. En række forbundne co-citationer genererer en række bibliografiske koblinger - og vice versa. En hypotese i Ph.D.-projektet postulerer, at det at finde den korteste co-citations-sti (i grafteoretisk forstand) mellem C_1 and C_n er analogt med at finde den korteste bibliografisk-koblings-sti mellem B_1 og B_m , forudsat at B_1 har et udlink til C_1 samt B_m har et udlink til C_n .

Figuren viser også tydeligt forskellen mellem bibliografisk kobling og co-citation. Bibliografisk kobling (Kessler 1963) indebærer, at to dokumenter har fælles udlinks (referencer) til samme destinationsdokument. Co-citation (Small 1973) indebærer, at to dokumenter har fælles indlinks (citationer) fra samme kildedokument. Den underliggende antagelse ved brug af bibliografisk kobling og co-citation er, at de afspejler semantiske relationer.

Co-linknings-kæder indgår i casestudiet af forskeres personlige websider, der beskrives i næste afsnit. I casestudiet søges identificeret og lokaliseret transversale forbindelser i form af bibliografiske koblinger og co-citationer på tværs af videnskabelige discipliner.

Et casestudie

"It's probably written in stone somewhere: 'On thine homepage thou shalt keep a list. That list shalt contain thine most precious links.'"

(Fra den svenske computerprogrammør Mattias Borells hjemmeside, www.lub.lu.se/~mattias/)

I et casestudie, der indgår i Ph.D.-projektet bruges to typer af forskeres personlige websider. Den ene type udgøres af forskeres hjemmesider, dvs. de indgangs- eller startside, hvor de præsenterer sig selv som forskere. Typisk indeholder sådanne hjemmesider elementer i form af cv, publikationsliste samt beskrivelse af forskningsinteresser og -projekter. Hos nogle forskere fylder disse elementer så meget, at de ligger opdelt på flere underordnede websider. Den anden type websider i casestudiet er forskeres 'bookmark'-lister (og lignende typer af lange link-lister), hvor der typisk er brugt et software-program til at konvertere personlige 'bookmarks' fra personens browser til en link-liste på en webside. Også ikke-forskere gør deres 'bookmark'-lister offentligt tilgængelige. Man skal dog være opmærksom på den mulige 'bias' ved denne type af link-lister. Måske er det typisk mere computer-mindede personer, der laver dem.

'Bookmark'-lister afspejler websider, som personen selv har besøgt på WWW, og som personen nu giver mulighed for, at også andre personer kan besøge. Forskere kan have lavet 'bookmarks' til relaterede forskere, institutioner, projekter, artikler mm. Ofte er der også lavet 'bookmarks' til mere perifere faglige eller hobbyprægede interesser og emneområder. Nogle 'bookmarks' afspejler måske frembrydende 'research fronts' eller 'invisible colleges' i form af uformelle forskernetværk.

'Bookmark'-lister kan give besøgende web-brugere adgang til en persons *mangfoldighed* af interesser, præferencer og handlinger på WWW samlet *ét sted*. Forskeres 'bookmark'-lister er en ny og forholdsvis upåagtet genre inden for videnskabelig formidling, der udgør et oplagt forskningsobjekt for scientometri og webometri. De lidt perifere, 'skæve' links, som disse lister kan indeholde, gør dem også til oplagte steder at lokalisere transversale links, der forkorter afstande på WWW.

I det eksempel på en case, der er gengivet i nedenstående figur, er udgangspunktet en forskers hjemmeside, nærmere bestemt en forsker inden for 'computer science', Jon Kleinberg, ved Cornell University i USA. Han er bl.a. valgt, fordi han er interesseret i linkstrukturer og 'small-world'-fænomener på WWW. Han er også en velesti-

meret forsker, der derfor får mange indlinks fra andre forskere. I casen bruges AltaVista til at finde indlinks fra forskeres 'bookmark'-lister o.l. til Kleinbergs hjemmeside. Dette gøres ved at skrive følgende søgestreng i AltaVista Advanced Search, der tillader komplekse kombinationer af booleske operatoren og søgekoder:

(title:bookmark OR title:link*)
AND anchor:"jon kleinberg"*

Herved søges efter websider, hvor ordene 'bookmark', 'bookmarks', 'link', 'links' mm. indgår i websidens titelfelt og 'Jon Kleinberg' står som klikbar tekst ('anchor') et sted på websiden. I stedet for 'anchor:' kunne man også bruge søgekoden 'link:', hvor man skriver URL'en (web-adressen) til Kleinbergs hjemmeside. Men fordelene ved 'anchor:' er, at man undgår at få støj i form af websider, der peger på forskerens (måske ikke relevante) undersider, da AltaVista laver automatisk højretrunkering af URL'er. Et andet problem med 'link:' er, at mange forskere har forskellige URL'er, fordi deres hjemmesider findes i identiske udgaver på flere forskellige servere.

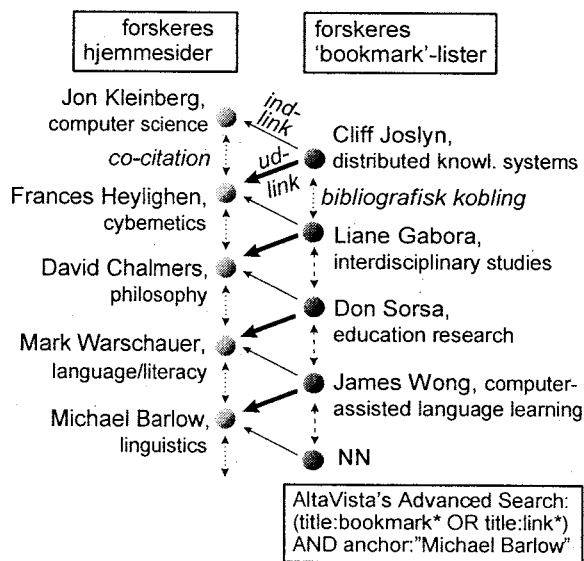


Fig. 8. Eksempel på case i form af en co-linknings-kæde.

I figuren kan man se, at Kleinberg har fået et indlink fra Cliff Joslyns 'bookmark'-liste (www.c3.lanl.gov/~joslyn/people.html). I udvælgelsen af

Joslyn blandt de godt 20 poster som ovenstående søgestreng resulterede i, er følgende kriterier brugt:

1. den valgte 'bookmark'-liste skal være lavet af en forsker med videnskabelige publikationer,
2. den valgte 'bookmark'-liste skal indeholde udlinks til andre forskeres hjemmesider, ellers er co-citation mellem hjemmesider ikke mulig i kæden,
3. den valgte 'bookmark'-liste skal indeholde et udlink til en hjemmeside, der er lavet af en forsker, der ikke er for emnemæssigt tæt på den co-linkede (co-citerede) forsker (in casu: Kleinberg)

Det sidste punkt er svært at afgøre. Som tidligere nævnt søges derfor udviklet mere objektivt operationaliserbare kriterier her, fx med ekstra 'metadata' i form af tidsskriftsnavne, emnefelter og co-citationsfrekvens hentet fra citationsdatabaser. I det brugte eksempel er spørgsmålet afgjort ved subjektiv bedømmelse af forskere i emner og links.

Næste trin i konstruktionen af co-linknings-kæden i figuren ligger i direkte forlængelse af punkt 3 ovenfor:

4. den valgte hjemmeside skal være lavet af en forsker med videnskabelige publikationer,
5. den valgte hjemmeside skal have fået indlinks også fra andre forskeres 'bookmark'-lister (afprøvet i AltaVista med en søgestreng i stil med ovenstående), ellers er bibliografisk kobling mellem 'bookmark'-lister ikke mulig i kæden,
6. den valgte hjemmeside skal have fået et indlink fra en 'bookmark'-liste, der er lavet af en forsker, der ikke er for emnemæssigt tæt på den co-linkende (co-citerende) forsker (in casu: Joslyn)

På Joslyns 'bookmark'-liste blev Heylighen valgt, igen ved en subjektiv bedømmelse.

Ovenstående trin af skiftevis søgninger i AltaVista efter indlinks til hjemmesider fra 'bookmark'-lister og udvælgelse af udlinks (fede pile i figuren) direkte fra 'bookmark'-lister til hjemmesider gentages derefter. Denne fremgangsmåde,

der er udviklet i Ph.D.-projektet, resulterer i en co-linknings-kæde bestående af co-linkede (co-citerede) hjemmesider og co-linkende (bibliografisk koblete) 'bookmark'-lister. Fremgangsmåden kan beskrives som en 'trial-and-error'-metode, da typisk det andet og femte kriterie (om ud- og indlinks til og fra andre forskere) ikke altid kan opfyldes. I givet fald må man gå et eller flere trin tilbage i kæden og vælge en anden forskers hjemmeside eller 'bookmark'-liste i stedet, og så fortsætte kædekonstruktionen derfra. Kæden kan selvfølgelig blive vilkårligt lang. I eksemplet vist i figuren er der standset ved lingvisten Barlow, der altså er transitivt relateret til datalogen Kleinberg.

Hvad kan man så bruge dette til? I Ph.D.-projektet søges i første omgang udviklet metoder, understøttet ved computerhjælp, til at konstruere disse co-linknings-kæder. Den bagvedliggende idé er, som tidligere nævnt, at svage, transversale co-citationer og bibliografiske koblinger mellem forskelligartede forskningsfelter kan have implikationer for 'knowledge discovery' på tværs af videnskabelige discipliner, fx i form af 'undiscovered public knowledge' som hos Swanson (1986).

Social informatik og websociologi

Hvordan opstår linkstrukturer, hvad afspejler de og hvad kan de bruges til? I Ph.D.-projektet behandles disse spørgsmål i tilknytning til forskeres personlige hjemmesider og 'bookmark'-lister. Her er det vigtigt at undersøge og forstå den sociale og kognitive kontekst, som linkstrukturer optræder i. I Ph.D.-projektet inddrages bl.a. et begrebsapparat baseret på Björneborn (2000b) til at behandle disse spørgsmål. Det drejer sig om fem centrale netværksprocesser:

- *synliggørelse* - skabelse af synlighed for netværksaktører
- *sammenkobling* - skabelse af 'links' mellem netværksaktører
- *fællesgørelse* - skabelse af interessefællesskaber mellem netværksaktører
- *'på-tværs'-kobling* - skabelse af 'links' på tværs af uligeartede netværksaktører og derigennem skabelse af muligheder for at sprede og møde anderledes idéer
- *idéskabelse* - skabelse af kreativitet og innovation hos netværksaktører

De fem punkter ovenfor kan opfattes som en progression, hvor hvert nyt punkt forudsætter alle de foregående, men hvor et punkt ikke automatisk foranlediger næste punkt. Eksempelvis forudsætter skabelse af interessefællesskaber på WWW selvfølgelig, at nogle aktører synliggør sig (fx ved at lave websider) og kobler sig sammen (fx ved at lave links) med andre aktører, men synliggørelse og sammenkobling skaber ikke automatisk interessefællesskaber. Hos Björneborn (2000b) er punkterne lavet for at forklare videndeling og videnudvikling på WWW - men kan også overføres på forhold i netværk af aktører i en organisation.

Punkterne hænger tæt sammen med spørgsmål om, hvordan sociale faktorer og informations- og kommunikationsteknologier gensidigt påvirker hinanden. Dette behandles i informationsvidenskab i det subdomæne, der af nogle kaldes social informatik (se fx Kling 2000). Her er der store overlap med det kraftigt ekspanderende sociologiske felt, der benævnes cybersociologi, hvis hele Internet er omfattet (dvs. WWW, email, nyhedsgrupper, chat, mm.) eller websociologi, hvis kun WWW er omfattet. Her kan nævnes forskere som fx Castells (1996: om netværkssamfundet), de Beer (1996: om cyberantropologi), Erickson (1996: om social hypertext), Dieberger (1997: om social navigation), Kautz et al. (1997: om 'collaborative filtering'), Jackson (1997: om webkommunikation), Chandler (1998: om selvpresentation på WWW), de Kerckhove (1998: om 'konnektiv kollektiv intelligens' på WWW), Wellman & Gulia (1999: om interessefællesskaber) og Gotved (2000: om cybersociologi).

Afsluttende bemærkninger

I de foregående afsnit er skitseret forskningstilgange og begreber, der er relevante for det forskningsfelt om linkstrukturer, som Ph.D.-projektet indgår i. Ved at udvikle en teoretisk referenceramme og empiriske undersøgelsesmetoder, søges i Ph.D.-projektet etableret et 'brohoved' i udforskningen af dette ny forskningsfelt i informationsvidenskab, hvor et centralt spørgsmål kan formuleres:

Hvordan opstår 'small-world'-fænomener i informationsbaserede netværksstrukturer, in casu

WWW - og hvordan kan disse fænomener bruges (fx til 'knowledge discovery')?

Omdrejningspunktet i dette udforskningsprojekt er begrebet 'transversale links'. I artiklen er dette begreb blevet brugt til at beskrive nogle centrale sammenhænge og hypoteser, der kan sammenfattes således:

- transversale links konstrueres lokalt og uforudsigeligt af webforfattere, fx forskere, der laver 'bookmark'-lister
- transversale links skaber 'small-world'-fænomener i form af korte afstande på WWW
- transversale links gør dermed WWW mere sammenhængende og sammenkrøllet
- transversale links påvirker menneskelige og digitale aktørers muligheder for at bevæge sig rundt i og eksplorere WWW's dokumentnetværk
- transversale links samt transversale co-citationer og bibliografiske koblinger påvirker muligheder for serendipitet og computerstøttet 'knowledge discovery', når uventet men brugbar information påtræffes og udnyttes

Rationalet i det sidste punkt er, at jo kortere afstande, jo færre link-'hop', der er mellem uligeartede websider, dvs. jo mere 'small-world'-agtigt og 'sammenkrøllet' verdensvævet er, takket være transversale links, desto større mulighed er der for, at man opdager noget uventet.

I artiklen er eventuelle ulemper ved transversale links ikke blevet særskilt omtalt. Her kan man pege på, at transversale links forstærker uforudsigeligheden på WWW. Dette kan være en fordel, hvis man vil bruge WWW som et mulighedsrum med uventede bevægelses- og opdagelsesmuligheder på tværs af emneområder. Men uforudsigeligheden i det sammenkrøllede, sammenfiltrede verdensvæv medfører også, at brugere kun kan få overblik over begrænsede udsnit af de komplekse linkstrukturer. Det er derfor nemt at fare vild og blive overmæt af linkvævet - at blive 'lost in cyberspace'. Mulighedsrummets bagside er altså desorientering og 'information overload'.

Hvordan lærer vi så at håndtere dette mulighedsrum? Heldigvis er der mange lokale og globale

struktureringsiltag i gang på WWW, fx ressourceguides, metadata-standarder mm., der kan gøre det nemmere at bruge WWW mere målrettet og 'konvergent'. Konvergente, 'samlende' strukturer udgør samtidig den nødvendige kontrasterende 'baggrund', der gør det muligt at få øje på mere divergente, 'spredende' strukturer på WWW, der kan bruges mere umålrettet og intuitivt til uventede fund. Der eksisterer altså en komplementaritet mellem konvergente og divergente strukturer. Samtidig må vi nok ydmygt erkende, at verdensvævet, i lighed med andre komplekse selvorganiserende systemer, som fx biosfæren, ikke kan - eller bør - kontrolleres i alle detaljer, da man fx risikerer at ødelægge gode 'smutveje'. Men øvelse kan alligevel gøre os bedre til at navigere, gøre opdagelser og sortere i en overflod af muligheder. Det er nok især vigtigt at opøve nogle unikt menneskelige kompetencer såsom nysgerrighed, tværfaglig viden, skelneevne og intuition. Browsere, søgemaskiner, mm. får forhåbentlig også bedre visualiserings- og navigationsmuligheder og mere brugervenlige måder at holde styr på benyttede søge-/ 'snuble'-veje.

Samtidig er det vigtigt at huske på, at WWW kun har godt 10 år på bagen. WWW's opbygning og brugermæssige potentialer er altså langt fra udviklet endnu. WWW er således stadig i sin vorden som et mulighedsrum, hvis potentialer for 'information recovery' og 'discovery' - for menneskelig videndeling og videnudvikling - venter på udforskning og udfoldelse. Her har informationsvidenskab en vigtig forskningsrolle, hvortil Ph.D.-projektet forhåbentlig vil kunne bidrage.

Litteraturliste

[Alle URL'er er validerede 30.4.2001.]

Adamic, Lada (1999). 'The small world web'. In: Abitotoul, Serge & Vercoustre, Anne-Marie (eds.). *Proceedings of the 3rd European Conference on Digital Libraries*. Berlin: Springer. pp. 443-452. (Lecture Notes in Computer Science; 1696).
Tilgængelig: <http://www.parc.xerox.com/istl/groups/iea/www/smallworldpaper.html>

- Agosti, Maristella (1996). 'An overview of hypertext'. In: Agosti, Maristella & Smeaton, Alan F. (eds.). *Information retrieval and hypertext*. Boston: Kluwer Academic Publishers. pp. 27-47.
- Almind, Tomas C. & Ingwersen, Peter (1997). 'Informetric analyses on the World Wide Web: methodological approaches to 'webometrics''. *Journal of Documentation*, 53(4): 404-426.
- Bar-Ilan, Judit (1998). 'The mathematician, Paul Erdos (1913-1996) in the eyes of the Internet'. *Scientometrics*, 43(2): 257-267.
- Bar-Ilan, Judit (2001). 'Data collection methods on the Web for informetric purposes : a review and analysis'. *Scientometrics*, 50(1): 7-32.
- Bates, Marcia J. (1986). 'An exploratory paradigm for online information retrieval'. In: Brookes, B.C. (ed.). *Proceedings of the 6th International Research Forum in Information Science (IRFIS 6)*. Amsterdam: North-Holland. pp. 91-99.
- Bawden, David (1986). 'Information systems and the stimulation of creativity'. *Journal of Information Science*, 12: 203-216.
- Bergman, Michael K. (2000). 'The Deep Web: Surfacing hidden value. White paper'. Bright Planet. Tilgængelig: <http://128.121.227.57/download/deepwebwhitepaper.pdf>
- Berners-Lee, Tim (1989/1990). 'Information management: a proposal'. CERN. Tilgængelig: <http://www.w3.org/History/1989/Proposal.html>
- Berners-Lee, Tim (1997). *Realising the full potential of the Web*. Tilgængelig: <http://www.w3.org/1998/02/Potential.html>
- Bernstein, Mark (1998). 'Patterns of hypertext'. *Proceedings of ACM Hypertext '98*. Tilgængelig: <http://www.eastgate.com/patterns/Print.html>
- Bharat, Krishna & Henzinger, Monika R. (1998). 'Improved algorithms for topic distillation in a hyperlinked environment'. In: Croft, W. Bruce et al. (eds.). *Proceedings of the 21st annual International ACM SIGIR Conference on Research and Development in Information Retrieval*. ACM Press, 1998. pp. 104-111.
- Björneborn, Lennart (1998). *Connecto ergo sum : synliggørelse - sammenkobling - synergi - webside-fællesskaber : fænomener omkring forskeres brug af personlige websider og links*. Hovedopgave. København: Danmarks Biblioteksskole.
- Björneborn, Lennart (2000a). 'Connecto ergo sum'. *Kronik, Dagbladet Information*, 24.3.2000.
- Björneborn, Lennart (2000b). *Verdensvævet som 'small-world'-netværk og mulighedsrum : omridset af en forståelsesmodel for transversale links på World Wide Web*. Speciale. København: Danmarks Biblioteksskole.
- Björneborn, Lennart & Ingwersen, Peter (2001). 'Perspectives of webometrics'. *Scientometrics*, 50(1): 65-82.
- Bollobás, Béla (1998). *Modern graph theory*. New York: Springer.
- Bolter, Jay David (1991). *Writing space : the computer, hypertext, and the history of writing*. Hillsdale, N.J.: Lawrence Erlbaum Associates.
- Borell, Klas & Johansson, Roine (1996). *Samhället som nätverk : om nätverksanalys och samhällsteori*. Lund: Studentlitteratur.
- Botafogo, Rodrigo A. & Shneiderman, Ben (1991). 'Identifying aggregates in hypertext'. *Proceedings of the 3rd ACM Conference on hypertext*. pp. 63-74.
- Brin, Sergey & Page, Lawrence (1998). 'The anatomy of a large-scale hypertextual Web search engine'. WWW7 Conference. Tilgængelig: <http://citeseer.nj.nec.com/brin98anatomy.html>
- Broder, Andrei et al. (2000). 'Graph structure in the Web'. WWW9 Conference. Tilgængelig: <http://www.almaden.ibm.com/cs/k53/www9.final>
- Bush, Vannevar (1945). 'As we may think'. In: Nyce, James M. & Kahn, Paul (eds.) (1991). *From Memex to hypertext : Vannevar Bush and the Mind's Machine*. Boston: Academic Press. pp. 85-110.

Note: Oprindeligt publiceret i *The Atlantic Monthly*, juli 1945.

Bøgh Andersen, Peter (1998). 'WWW as self-organizing system'. *Cybernetics & Human Knowing*, 5(2): 5-41.

Castells, Manuel (1996). *The rise of the network society*. Oxford: Blackwell Publishers. (The information age : economy, society and culture; 1)

Chakrabarti, Soumen *et al.* (1999). 'Hypersearching the Web'. *Scientific American*, 280(6): 54-60. Tilgængelig: <http://www.sciam.com/1999/0699issue/0699raghavan.html>

Chandler, Daniel (1998). 'Personal home pages and the construction of identities on the Web'. Tilgængelig: <http://www.aber.ac.uk/media/Documents/short/webident.html>

Chang, Shan-Ju & Rice, Ronald E. (1993). 'Browsing : a multidimensional framework'. *Annual Review of Information Science and Technology*, 28: 231-276.

Chen, Chaomei *et al.* (1998). 'How did university departments interweave the Web: a study of connectivity and underlying factors'. *Interacting with Computers*, 10: 353-373.

Collins, James J. & Chow, Carson C. (1999). 'It's a small world'. *Nature*, 393(June 4): 409-410.

Conklin, Jeff (1987). 'Hypertext : an introduction and survey'. *IEEE Computer*, 20(9): 17-41.

Crane, Diana (1972). *Invisible colleges : diffusion of knowledge in scientific communities*. Chicago: University of Chicago Press.

Davies, Roy (1989). 'The creation of new knowledge by information retrieval and classification'. *Journal of Documentation*, 45(4): 273-301.

de Beer, Fanie (1996). 'The anthropology of Cyberspace'. In: Ingwersen, Peter & Pors, Niels Ole (eds.). *Integration in perspective : Proceedings: CoLIS 2 : Second International Conference on Conceptions of Library and Information Science*.

Copenhagen: Royal School of Librarianship. pp. 117-133.

de Kerckhove, Derrick (1998). *Connected intelligence : the arrival of the web society*. London: Kogan Page.

Dieberger, Andreas (1997). 'Supporting social navigation on the World Wide Web'. *International Journal of Human-Computer Studies*, 46(6): 805-825. Tilgængelig: <http://citeseer.nj.nec.com/andreas97supporting.html>

Dodge, Martin & Kitchin, Rob (2001). *Mapping cyberspace*. London: Routledge.
Se også: <http://www.mappingcyberspace.com>

Erdelez, Sanda (2000). 'Towards understanding information encountering on the Web'. In: *Proceedings of the 63rd ASIS Annual Meeting*, 37. Medford, NJ.: Information Today. pp. 363-371.

Erickson, Thomas (1996). 'The World Wide Web as social hypertext'. *Communications of the ACM*, 39(1): 15-17. Tilgængelig: http://www.pliant.org/personal/Tom_Erickson/SocialHypertext.html

Fayyad, Usama; Piatetsky-Shapiro, Gregory; Smyth, Padhraic & Uthurusamy, Ramasamy (eds.) (1996). *Advances in knowledge discovery and data mining*. Menlo Park, Cal.: AAAI Press / The MIT Press.

Ford, Nigel (1999). 'Information retrieval and creativity : towards support for the original thinker'. *Journal of Documentation*, 55(5): 528-542.

Frawley, W.J.; Piatetsky-Shapiro, G. & Matheus, C.J. (1991). 'Knowledge discovery in databases: an overview'. In: G. Piatetsky-Shapiro, W.J. Frawley (eds.). *Knowledge discovery in databases*. Menlo Park, Cal.: AAAI Press. pp. 1-27.

Garfield, Eugene (1986). 'The metaphor-science connection'. *Current Comments*, 42(October 20): 3-10.

Garner, Ralph (1967). 'A computer-oriented graph theoretic analysis of citation index structures'. In: Flood, Barbara (ed.) (1967). *Three Drexel*

information science research studies. Drexel Press. pp. 3-46. Tilgængelig: <http://www.garfield.library.upenn.edu/rgarner.pdf>

Gibson, David, Kleinberg, Jon & Raghavan, Prabhakar (1998). 'Inferring web communities from link topology'. *Proceedings of the 9th ACM Conference on Hypertext and Hypermedia*, 1998. Tilgængelig: <http://www.cs.cornell.edu/home/kleinber/ht98.pdf>

Gotved, Stine (1997). *Cybersociologi : om fællesskab på Internet*. København: Sociologisk Institut. Note: Upubliceret oplæg i Ph.D.-projekt.

Gotved, Stine (2000). *Cybersociologi : det samme på en anden måde*. Ph.D.-afhandling. København: Sociologisk Institut, Københavns Universitet.

Granovetter, Mark S. (1973). 'The strength of weak ties'. *American Journal of Sociology*, 78(6): 1360-1380.

Gross, Jonathan & Yellen, Jay (1999). *Graph theory and its applications*. Boca Raton, Flor.: CRC Press.

Grossman, Jerrold W. (2000). 'The Erdős number project'. Tilgængelig: <http://www.acs.oakland.edu/~grossman/erdoshp.html>

Gundersen, Karin (1999). *Allegorier : innganger til litteraturens rom*. Oslo: Aschehoug.

Haas, Stephanie W. & Grams, Erika S. (1998). 'A link taxonomy for web pages'. In: *Proceedings of the 61st ASIS Annual Meeting*, 35. Medford, NJ: Information Today. pp. 485-495.

Harasim, Linda M. (1993). 'Global networks: an introduction'. In: Linda M. Harasim (ed.). *Global networks : computers and international communication*. Cambridge, Mass.: MIT Press. pp. 3-14.

Hjørland, Birger (1995). *Informationsvidenskabelige grundbegreber*. 2. udg. København: Danmarks Biblioteksskole.

Hjørland, Birger (1997). *Information seeking and subject representation : an activity-theoretical approach to information science*. Westport, Conn.: Greenwood Press.

Huberman, Bernardo A. & Adamic, Lada (1999). 'Growth dynamics of the World-Wide Web'. *Nature*, 401 (September 9): 131.

Ingwersen, Peter (1996). 'Cognitive perspectives of information retrieval interaction: elements of a cognitive IR theory'. *Journal of Documentation*, 52(1): 3-50.

Interspace (1998). 'The third wave of net evolution'. University of Illinois at Urbana-Champaign. Tilgængelig: <http://www.canis.uiuc.edu/projects/interspace/thirdwave.html>

Jackson, Michele H. (1997). 'Assessing the structure of communication on the World Wide Web'. *Journal of Computer Mediated Communication*, 3(1). Tilgængelig: <http://www.ascusc.org/jcmc/vol3/issue1/jackson.html>

Kautz, Henry; Selman, Bart & Shah, Mehul (1997). 'Referral Web: combining social networks and collaborative filtering'. *Communications of the ACM*, 40(3): 63-65.

Kessler, M.M. (1963). 'Bibliographic coupling between scientific papers'. *American Documentation*, 14(1): 10-25.

Kim, Hak Joon (2000). 'Motivations for hyperlinking in scholarly electronic articles: a qualitative study'. *Journal of the American Society for Information Science*, 51(10): 887-899.

Klein, Julie Thompson (1996). *Crossing boundaries : knowledge, disciplinarity, and interdisciplinarity*. Charlottesville, Virg.: University Press of Virginia.

Kleinberg, Jon M. (1998). 'Authoritative sources in a hyperlinked environment'. *Proceedings of the 9th annual ACM-SIAM Symposium on Discrete Algorithms*. pp. 668-677.

Kling, Rob (2000). 'Learning about information technologies and social change: the contribution of social informatics'. *The Information Society*, 16: 217-232.

Kochen, Manfred (ed.) (1989). *The small world*. Norwood, N.J.: Ablex Publishing Corporation.

Koehler, Wallace (1999). 'An analysis of web page and web site constancy and permanence'. *Journal of the American Society for Information Science*, 50(2): 162-180.

Landow, George P. (1997). *Hypertext 2.0 : the convergence of contemporary critical theory and technology*. Revised, amplified ed. Baltimore: The Johns Hopkins University Press.

Larson, Ray R. (1996). 'Bibliometrics of the World Wide Web: an exploratory analysis of the intellectual structure of Cyberspace'. In: Hardin, Steve (ed.). *Global complexity : information, chaos, and control. Proceedings of the 59th ASIS Annual Meeting*. Medford, NJ: Learned Information Inc./ASIS. Tilgængelig: <http://sherlock.berkeley.edu/asis96/asis96.html>

Lawrence, Steve & Giles, C. Lee (1999). 'Accessibility of information on the Web'. *Nature*, 400, July 8: 107-109. Tilgængelig: http://www.mri.mq.edu.au/~einat/web_ir/LandG_Nature99.pdf

Milgram, Stanley (1967). 'The small-world problem'. *Psychology Today*, 1(1): 60-67.

Newman, M.E.J. (2001). 'The structure of scientific collaboration networks', *Proceedings of the National Academy of Sciences*, 16 Jan. 2001, 98(2): 404-409.

O'Connor, Brian (1988). 'Fostering creativity : enhancing the browsing environment'. *International Journal of Information Management*, 8: 203-210.

Pierce, Sydney J. (1999). 'Boundary crossing in research literatures as a means of interdisciplinary information transfer.' *Journal of the American Society for Information Science*, 50(3): 271-279.

Pitkow, James E. (1997). *Characterizing World Wide Web ecologies*. Doctoral dissertation. Georgia Institute of Technology. Tilgængelig: <ftp://ftp.gvu.gatech.edu/pub/gvu/tr/1997/97-16.pdf>

Postman, Neil (1993). *Teknopolis : kulturens knæfald for teknologien*. København: Hekla.

Rada, Roy (1991). *Hypertext: from text to Expertext*. London: McGraw-Hill.

Rogers, Everett M. (1995). *Diffusion of innovations*. 4.ed. New York: The Free Press.

Scott, John (2000). *Social network analysis : a handbook*. 2. ed. Thousand Oaks, Cal.: SAGE Publications.

Simpson, Rosemary et al. (1996). '50 years after 'As we may think': The Brown/MIT Vannevar Bush Symposium'. Tilgængelig: http://www.cs.brown.edu/memex/Bush_Symposium_Interact_2.html

Small, Henry (1973). 'Co-citation in the scientific literature: a new measure of the relationship between two documents'. *Journal of the American Society for Information Science*, 24(4): 265-269.

Small, Henry (1999). 'A passage through science: crossing disciplinary boundaries'. *Library Trends*, 48(1): 72-108.

Snyder, Herbert & Rosenbaum, Howard (1999). 'Can search engines be used as tools for web-link analysis? A critical view'. *Journal of Documentation*, 55(4): 375-384.

Swanson, Don R. (1986). 'Undiscovered public knowledge'. *Library Quarterly*, 56(2): 103-118.

Swanson, Don R. & Smalheiser, Neil R. (1999). 'Implicit text linkages between Medline records: using Arrowsmith as an aid to scientific discovery'. *Library Trends*, 48(1): 48-59.

Toms, Elaine G. (1998). 'Information exploration of the third kind: the concept of chance encounters'. A position paper for the CHI98 Workshop

on Information Exploration. Tilgængelig: <http://www.fxpal.com/CHI98IE/submissions/long/toms/>

Van Andel, Pek (1994). 'Anatomy of the unsought finding : serendipity: origin, history, domains, traditions, appearances, patterns and programmability. *British Journal for the Philosophy of Science*, 45(2): 631-648.

Wasserman, Stanley & Faust, Katherine (1994). *Social network analysis : methods and applications*. Cambridge: Cambridge University Press.

Watts, Duncan J. (1999). *Small worlds : the dynamics of networks between order and randomness*. Princeton, N.J.: Princeton University Press. (Princeton Studies in Complexity)

Watts, Duncan J. & Strogatz, Steven H. (1998). 'Collective dynamics of 'small-world' networks'. *Nature*, 393(June 4): 440-442.

Wellman, Barry & Gulia, Milena (1999). 'Virtual communities as communities : net surfers don't ride alone.' In: Smith, Marc A. & Kollock, Peter (eds.). *Communities in Cyberspace*. London: Routledge. pp. 167-194.