

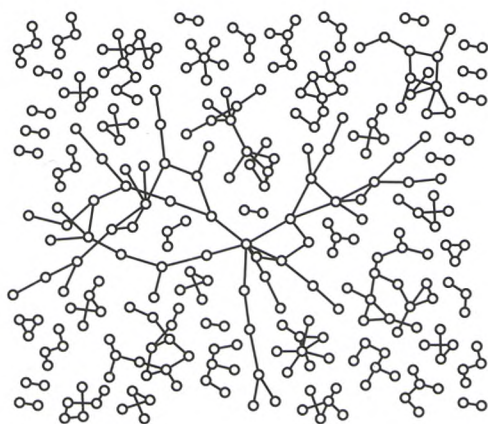
Komplekse netværk

Af Vedran Sekara, DTU Compute, Danmarks Tekniske Universitet og Sune Lehmann, Niels Bohr Institutet, Københavns Universitet og DTU Compute, Danmarks Tekniske Universitet

Vi er omgivet af netværk. Vores sociale forbindelse danner forskellige netværk: netværk af venner, kolleger, familier, eller seksuelle partnere. Vi kommunikerer via et væld af teknologiske netværk: vi kan ringe til hinanden, sende emails, skrive til hinanden via *Facebook*, eller sende *Snaps*. Infrastruktur-netværk leverer el, vand og varme til vores hjem, mens vi bevæger os på transportnetværk som fx motorveje, busser, toge, eller fly. Vores biologi er også reguleret af netværk – fx regulerer vores gener hinanden i et komplekst netværk – så når et menneske får kræft, kan det skyldes en ubalance i gen-regulations netværket. Givet den centrale rolle som netværk spiller henover hele vores verden, er forståelsen og en matematisk beskrivelse af netværk en af det 21. århundredes vigtigste intellektuelle og videnskabelige udfordringer.

Introduktion

Bag hvert komplekst system kan vi tegne et diagram, der definerer samspillet mellem de forskellige komponenter. Vi kalder dette diagram for et netværk. Netværk består af to grundelementer, *knudepunkter* (som vi nogle gange kalder *noder*) og links (som også kaldes *forbindelser*, *interaktioner* eller *kanter* afhængig af den videnskabelige disciplin). Knudepunkterne bruger vi til at repræsentere systemets elementer: fx hjemmesider, personer eller byer. Links repræsenterer systemets vekselvirkninger, så vi forbinder to knudepunkter via et link, hvis de deler en relation. Så to hjemmesider er forbundne, hvis den ene side indeholder et hyperlink til den anden, to personer er forbundne hvis personerne er venner, og vi kan tegne et link mellem to byer, hvis fx man kan flyve fra den ene til den anden (se figur 1). Links kan være retningsbestemte, afhængigt af relationen mellem to knudepunkter; fx hvis person A sender en besked til person B, siger vi, at forbindelsen mellem A og B har en retning (linket peger fra A til B), men hvis A og B derimod er venner på Facebook er retningen på linket uvæsentlig – eftersom Facebookvenskaber er gensidige per definition.



Figur 1. Eksempel på et netværk. Hver knudepunkt (cirkel) repræsenterer en person, mens linjerne mellem personerne kaldes for links. Netværket viser ca. 300 DTU-studerendes indbyrdes telefonopkald over to uger.

Styrken ved netværk er, at de afspejler et systems struktur. De er et matematisk begrebsapparat, der lader

os bruge samme type matematik til at beskrive vidt forskellige fænomener. I princippet betyder dette, at vi kan anvende de samme værktøjer, uanset om vi studerer sociale relationer mellem mennesker eller kemiske forbindelser (fx signalstoffer) mellem gener i kroppen. Sagt anderledes er et netværk et udtryk for et systems rå struktur, der giver os et matematisk sprog, hvor vi kan sammenligne vidt forskellige systemer.

Netværk som matematisk begreb har eksisteret siden 1700-tallet, hvor Leonhard Euler ifølge legenden opfandt graf-teorien til at løse en berømt gåde om gåture på Königsbergs syv broer. Siden den gang har netværk slået dybe rødder indenfor traditionelle videnskabelige felter som matematik, biologi, datalogi og sociologi. Men det er først i det 21. århundrede, at feltet har udviklet sig til en selvstændig videnskab. Denne udvikling faldt sammen med en øget interesse for netværk fra en højst uventet kant, nemlig fra statistisk fysik. Da dette også er vores perspektiv, vil vi fortrinsvis fokusere på fysikernes rolle indenfor netværksvidenskaben i denne artikel.

Indenfor fysikken er historien om netværk drevet af en søgen efter forståelse af struktur og dynamik. Derfor er et naturligt sted at starte for en fysiker en meget simpel model. Den mest forsimplede model af et empirisk ('*real-world*') netværk vi kan forestille os, er et *random* (tilfældigt) netværk bestående af n tilfældigt forbundne knudepunkter. Med '*random*' mener vi, at hvert par af knudepunkter er forbundet med en sandsynlighed p , hvor p er et tal mellem 0 og 1. Allerede ud fra disse simple definitioner, kan vi begynde at beskrive strukturen af dette netværk – systemet for, hvordan de forskellige knuder er forbundne. Et meget simpelt mål for strukturen er antallet af naboer, k , også kaldt knudepunktets *degree*. Givet vores definition ovenfor kan vi se, at den gennemsnitlige degree i et *random* netværk er $\langle k \rangle = np$.

Men alle knudepunkter har selvfølgelig ikke præcis np forbindelser. På grund af stokastiske fluktuationer er der en sandsynlighedsfordeling $P(k)$ over for, hvor mange forbindelser hvert knudepunkt har, hvor $P(k)$ betegner sandsynligheden for, at et tilfældigt udvalgt knudepunkt har præcis k forbindelser. Vi kalder $P(k)$ for en *degree-fordeling*. I et stort, *random* netværk kan vi regne os frem til, at fordelingen af antallet af naboer

følger en Poisson-fordeling

$$P(k) = \frac{\langle k \rangle^k}{k!} e^{-\langle k \rangle}. \quad (1)$$

Når $\langle k \rangle$ er stor, kan (1) approksimeres med en normal-fordeling. *Degree*-fordelingen er et mål for struktur i netværk, og har vist sig at give interessante indsigter i mange empiriske netværk. I sociale netværk ved vi, at mennesker kan have vidt forskellige antal venner – hypersociale personer kan have hundredevis af aktive bekendtskaber, mens andre nøjes med nogle få tætte venskaber.

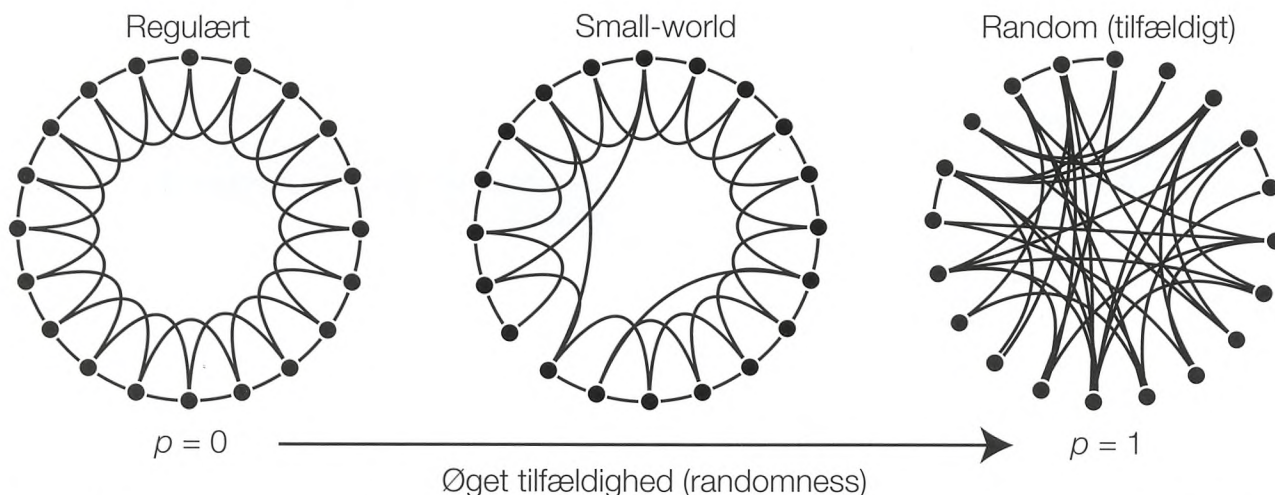
Moderne netværksvidenskab fødes

Som sagt handler historien om moderne netværksvidenskab på mange måder om, hvordan videnskabsfolk gradvist har afdækket, hvor ikke-tilfældigt de netværk, som vi observerer ude i verden, er struktureret. En begyndende erkendelse af, at *random* netværk er en alt for simpel model til at beskrive empiriske netværk, opstod i en artikel af Watts og Strogatz, der blev udgivet i det prestigefyldte tidsskrift *Nature* i 1998 [1]. Watts og Strogatz forsøger i deres banebrydende artikel at skabe en model, der forener to fundamentale observationer om netværk.

Den første observation er, at mange empiriske netværk har høj *clustering* (eller klyngedannelse). *Clustering* er let at forklare ud fra sociale netværk. Begrebet

udtrykker, at hvis jeg har to venner, så er der en høj sandsynlighed for, at de to personer også er venner med hinanden. Mere generelt handler *clustering* om, hvordan et knudepunkts naboer er indbyrdes forbundne – dvs. om brøkdelen af trekanter i netværket. Empirisk set kan *clustering* observeres i en bred vifte af netværk, ikke kun de sociale. De *random* netværk som vi diskuterede ovenfor udviser dog typisk meget lav *clustering*, da deres definition ikke fremmer dannelsen af trekanter.

Den anden observation stammer fra et fascinerende eksperiment, som socialpsykologen Stanley Milgram udførte i 1967 [2]. I sit udspekulerede eksperiment viste Milgram, at den gennemsnitlige afstand mellem to individer i det menneskelige sociale netværk er overraskende kort. Milgram kom til dette resultat ved at uddele en række breve til en gruppe personer i ‘verdens ende’ (*Omaha, Nebraska*, ifølge Milgram) adresseret til en af Milgrams bekendte. Milgrams ven var advokat og boede i *Boston, Massachusetts*, som Milgram (der på det tidspunkt var ansat på Harvard) mente var ‘verdens centrum’. Tricket var, at hver deltager skulle guide brevet gennem det sociale netværk ved at give det videre til en person som hun/han kendte på fornavnsbasis. Ved at registrere alle skridt på brevenes vej viste Milgram, at afstanden fra Nebraska til Massachusetts kun var 6 hop i netværket, hvilket ledte til den populære betegnelse: “*six degrees of separation*”. Vi kalder netværk som har denne egenskab (korte stier mellem alle knudepunkter) for “small-world” netværk (se figur 2).



Figur 2. Watts og Strogatz’ small-world netværk model. Kan det være, at komplekse netværk ligger på grænsen mellem det fuldstændigt ordnede netværk – eksemplificeret ved det regulære 1-dimensionale gitter til venstre – og det helt tilfældigt forbundne netværk til højre? Det viser sig, at et gitter med ganske få tilfældige forbindelser har både høj *clustering* og *small-world* egenskaber.

Ved første øjekast kunne man godt forestille sig, at *random* netværk havde *small-world* egenskaber. Lad os forestille os, at jeg er et knudepunkt i et *random* netværk med 100 venner. Hvis nu hver af mine venner også har 100 venner er vi allerede oppe på $100^2 = 10.000$ personer efter to hop i mit netværk. Går vi tre skridt i netværket finder vi $100^3 = 1.000.000$ knudepunkter – og efter seks hop i et *random* netværk med en gennemsnitlig *degree* $\langle k \rangle = 100$ er vi oppe på 10^{12} , ca. 150 gange mere end Jordens befolkning. Men der er selvfølgelig et hul i dette argument, og problemet

er *clustering*. Mange af mine venner er jo venner med hinanden, så i eksemplet ovenfor er der ikke 10.000 nye personer efter to hop i netværket. Dette betyder, at *clustering* forklarer, hvorfor *random* netværk ikke kan bruges som en model for mange virkelige systemer.

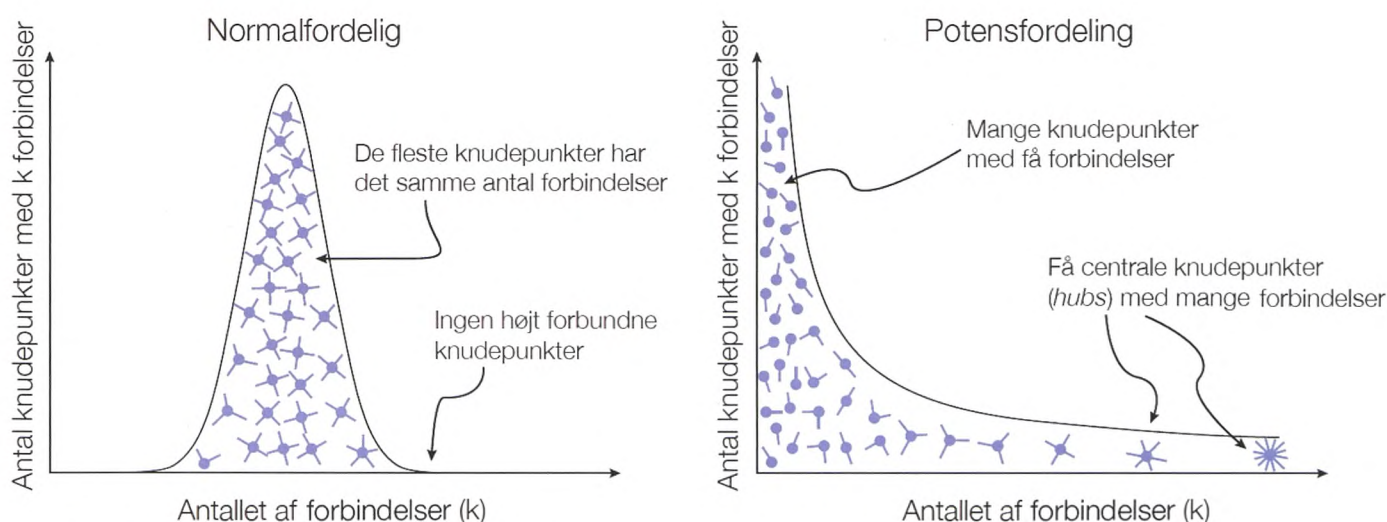
Watts og Stogatz foreslog i stedet en ny model, der sigter efter at forene høj *clustering* med *small-world* egenskaben. Deres model udspringer af et meget simpelt netværk, et cirkulært gitter, hvor alle knuder har præcis k nærmeste naboer (se figur 2 helt til venstre). Dette gitter har høj *clustering* på grund af de mange

forbindelser til naboerne. Men med mindre k er meget stor, i forhold til antallet af knudepunkter i netværket, kræver det mange skridt for at nå fra den ene side af netværket til den anden, dvs. det er ikke et *small-world* netværk.

Man kan nu tilføje tilfældighed til dette gitter ved at tilføje lidt støj og forbinde en brøkdel p af netværkets links tilfældigt. Hvis $p = 0$ har vi vores regulære gitter (figur 2, til venstre), og hvis $p = 1$ er netværket helt tilfældigt forbundet; vi genfinder altså det gode gamle *random* netværk (figur 2, til højre). Det spændende ved denne model er, at hvis bare p er en lille smule større end nul, så fungerer de tilfældige links som genveje – og stierne blive dramatisk kortere (figur 2, i midten). Den fysiske fortolkning af denne model er klar: Vi interagerer til dagligt med en meget sammentømret klike, men samtidig har vi også fjerne links med barndomsvenner

eller andre bekendtskaber med personer, som vi fx har mødt på andre tidspunkter i vores liv. De tætte links bidrager til høj *clustering*, mens de fjerne bekendtskaber sænker afstanden i netværket.

Small-world modellen, som Watts og Strogatz foreslog, er simpel og intuitiv, men den gør ikke rede for en vigtig egenskab, som vi finder igen og igen i empiriske netværk. I *small-world* modellen er *degree*-fordelingen nemlig lig med et fast k givet ved definitionen af gitteret (eller en Poisson-fordeling afhængig af, hvordan vi tilføjer støj til netværket). Men denne *degree*-fordeling svarer ikke til, hvad vi ser ude i verden. Dette blev påpeget året efter, i 1999, hvor Albert og Barabási viste at *degree*-fordelingen i en bred vifte af empiriske netværk følger en potens-fordeling $P(k) \propto k^{-\gamma}$, hvor γ typisk har værdier mellem 2 og 3 [3]. Forskellen mellem de to fordelinger er illustreret i figur 3.



Figur 3. Albert og Barabásis opdagelse belyste vigtigheden af *hubs*, højt forbundne knudepunkter i halen af potensfordelingen. De to *degree* fordelinger er fundamentalt forskellige. I potensfordelinger er ekstremt forbundne knuder meget mere sandsynlige.

En potensfordeling over *degrees* har den fortolkning, at de fleste knudepunkter har et relativt lavt antal links, mens nogle få centrale *hubs* kan være forbundet til flere hundrede, tusinder, eller endda millioner af andre knudepunkter. Tilstedeværelsen af *hubs* har en gennemgribende indflydelse på strukturen af netværk. For eksempel var Watts og Strogatz nødt til eksplicit at indbygge *small-world* effekten i deres model, mens denne egenskab optræder spontant i mange netværk, der kan beskrives ved potens-fordelinger – her fungerer *hubs* som langtrækkende forbindelser. Forekomsten af *hubs* betyder også, at vigtige dynamiske processer, der udspiller sig på netværk, fx spredningen af sygdomme, computer-vira, eller rygter og informationer, spreder sig fundamentalt anderledes, end hvad vi ser, når vi modellerer disse via *random* netværk.

Netværk i det 21. århundrede

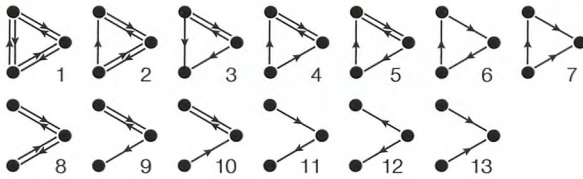
Albert og Barabásis opdagelse markerede begyndelsen på den moderne netværks-videnskab og medførte en hektisk aktivitet. Feltet tiltrak nemlig en stor gruppe af statistiske fysikere på jagt efter nye teoretiske

udfordringer. Dette skyldes potenslovenes dybe forbindelse til kritiske eksponenter, skala-invarians, og universalitets-klasser, som havde åbnet et nyt niveau af forståelse indenfor den statistiske fysik nogle år tidligere. Rejsen har hele tiden været hen imod at forstå strukturen i empiriske netværk, og alle steder blev der nu fundet ikke-tilfældige mønstre.

Efter fundet af den store uligevægt i fordelingen af forbindelser (*degree*) var det et logisk næste skridt at se på *degree*-korrelationer mellem knudepunkter [4]. Er højt forbundne knuder altid forbundet til andre stærkt tilsluttede knuder, eller er det omvendt? Det viser sig, at det afhænger af netværket. I biologiske netværk er der fx en tendens til, at knudepunkter med høj *degree* er forbundet med knudepunkter med lav *degree*, mens det i sociale og teknologiske netværk typisk er anderledes. I sådanne netværk er der en tendens til, at knudepunkter forbinder sig til knuder med tilsvarende *degree*, dvs. vi finder høj-høj og lav-lav forbindelser. Det betyder, at mennesker er mere tilbøjelige til at forme venskaber med personer, der minder om dem selv. *Degree*-korrelationer er interessante fordi to netværk sagtens

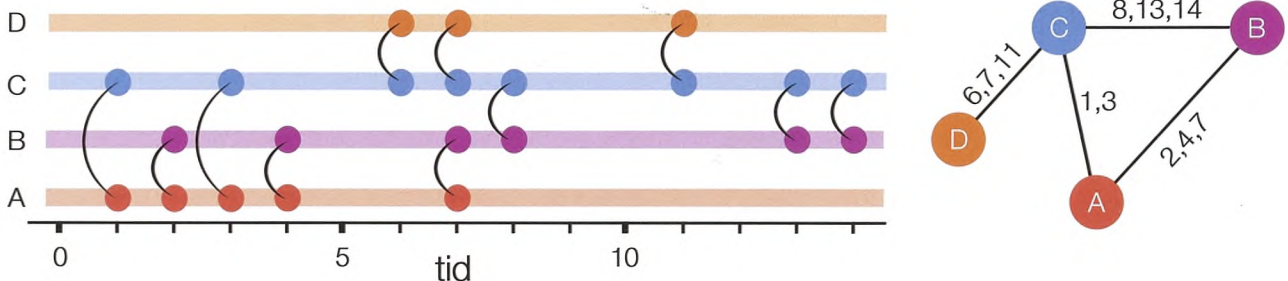
kan have ens *degree*-fordelinger – men, afhængig af deres *degree-degree* korrelationer, kan processerne, der udfolder sig på netværket, have helt forskellige udfald.

Vi kan nu gå videre og undersøge, hvordan netværket i sig selv er bygget op. Her kan vi undersøge en vigtig type struktur, *motiver*, der er små sub-netværk, som består af et lille antal knudepunkter [5]. Vi kan tænke på motiver som netværkets byggesten. Trekanten i *clustering* er et godt eksempel på sådan et motiv, men dette fænomen kan formuleres mere generelt. I netværk, hvor retningen på linket spiller en vigtig rolle, er der 13 typer af 3-knude motiver (se figur 4). Man kan også studere forekomsten af motiver, der indeholder flere end tre knudepunkter. Milo *et al.* viste, at nogle motiver er undertrykte, dvs. de opstår mindre hyppigt end hvad man skulle forvente, hvis netværket var tilfældigt forbundet, mens andre konfigurationer opstår langt hyppigere, end man skulle forvente. Motiverne viser hvilke processer, der er på spil når netværk opstår, og de afspejler netværkets funktion. Ligesom med *degree*-korrelationerne er de vigtige motiver forskellige fra netværk til netværk, men ligheder i motiv-forekomsten kan afsløre interessante forbindelser mellem forskellige netværk.



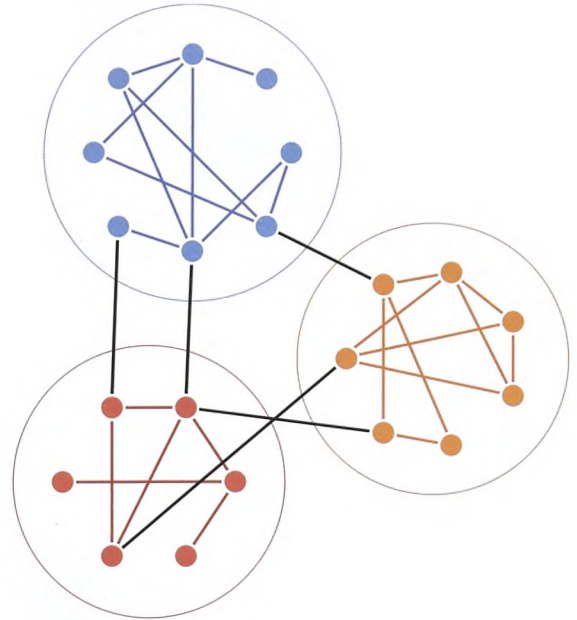
Figur 4. De 13 måder at tre knuder kan være forbundet til hinanden på i et rettet netværk. Der findes 199 typer motiver af størrelse fire, osv. Pilene på hvert link indikerer retningen på interaktionen.

Hvis vi zoomer ud og leder efter netværksstrukturer, som er større end motiverne, finder vi, at knudepunkter ofte har en tendens til at danne klynger, der er forbundne internt men svagt forbundne til resten af netværket [6]. Disse strukturer kalder vi for *communities* eller grupper (se figur 5). Denne tendens er måske ikke overraskende for en person indlejret i et socialt netværk. Vi interagerer alle til dagligt i grupper, som kan være opdelt efter kriterier som fx familie, alder, interesser, beskæftigelse, etnicitet, osv. *Communities* opstår ikke kun i sociale netværk. Nogle eksempler er grupper af hjemmesider, der beskæftiger sig med relaterede emner, eller klynger af proteiner, der udfører lignende funktioner i en celle.



Figur 6. Dynamisk netværk for fire knudepunkter og deres indbyrdes interaktioner. Til venstre ses kontakt-sekvensen blandt de fire knuder, mens netværket er tegnet til højre.

Yderligere så behøver *communities* ikke nødvendigvis at være helt adskilt som på figur 5, men kan have et vist overlap eller være fuldstændig hierarkisk indlejret i hinanden, hvilket kan gøre dem vanskelige at identificere [7].



Figur 5. *Communities* (eller grupper) i komplekse netværk. Hver gruppe har flere interne end eksterne forbindelser.

Hidtil har vi fokuseret på netværk som statiske strukturer, men den nyeste forskning tager højde for, at netværk i høj grad er dynamiske – de ændrer sig henover tid. Vi ved jo godt, at nye venskaber opstår og gamle forgår, gamle hjemmesider lukker ned og nye kommer frem, faktisk er alle netværk under en konstant udvikling. For at forstå disse dynamiske netværk kræver det, at vi udvikler nye statistiske redskaber, samt genovervejer mange af de tidligere fundamentale opdagelser. Links forsvinder og opstår, dette påvirker fx *degree*-fordelinger, som får tilført en tidsafhængig komponent. *Clustering* bliver ligeledes påvirket. Yderligere, betyder dynamiske netværk, at vi skal gentænke fundamentale begreber som *stier* på netværk. For at der kan findes en sti mellem knude *A* og *D* (se figur 6), kræves der en ordnet sekvens af fortløbende aktiveringer af links således, at link $A \rightarrow C$ opstår før $C \rightarrow D$. Dette betyder også, at dynamiske netværk ikke nødvendigvis er transitive, dvs. der kan godt gå en sti fra *A* til *D* men ikke fra *D* til *A*.

Nye tendenser indenfor netværksvidenskaben

Alle vores vekselvirkninger med computere bliver indsamlet og gemt i store centrale databaser. Hvert opkald, email, kreditkort-transaktion, Snap, Tweet, 'Like' på Facebook, klik på Netflix, og geografiske check-ins på diverse tjenester bliver registreret. Desuden indsamler offentlige instanser omfattende statistik vedrørende sundhedspleje, offentlige udgifter og kriminalitet. Vores liv bliver med andre ord registreret i historisk set uhørt omfattende grad, og mængden af information, der indsamles ventes kun at stige i fremtiden. Disse datamængder rejser helt nye problemer for privatlivets fred og talefrihed, men samtidig giver de forskere nogle enestående muligheder for at forstå vores samfund, menneskelig adfærd samt det underliggende sociale netværk [8].

Lige nu finder et af verdens største netværksforsøg sted på DTUs campus. Her har 1000 studerende fået udleveret gratis smartphones mod, at forskere får indsigt i deres daglige færden. Projektet, *The Copenhagen Network Study*, går ud på at kortlægge menneskelig adfærd på tværs af de mange netværk vi deltager i – fra ansigt-til-ansigt til telekommunikation og Facebook, kombineret med geografiske spor og personlighedsprofiler. Vi ønsker at forstå hvordan disse *multiplekse* netværk kan beskrives matematisk, hvordan de udvikler sig henover tid, og hvordan deres forandringer påvirker dynamiske processer (fx epidemier, indflydelse, eller rygtespredning), som udfolder sig henover de mange netværk. Samtidig handler projektet om at belyse og forstå de fundamentale problematikker vedrørende privatlivets fred.

Netværk er kommet for at blive. Når man først får øje på dem, er netværk alle vegne – og komplekse netværk har eksisteret siden livets begyndelse. Forskellen er, at henover de sidste 10-15 år, har vi som noget helt nyt fået adgang til data, der beskriver netværkenes komplekse og ikke-trivielle strukturer og dynamikker. Det betyder, at selvom det snart er 300 år siden, at Euler lagde fundamentet for netværksvidenskaben, er netværksvidenskaben langt fra modnet og støvet. Der er stadig masser af saftige, uløste problemer at tage fat på – og det er præcis dét, som gør vores arbejde fantastisk spændende.

Litteratur

- [1] Watts, D.J. and Strogatz, S.H. (1998), Collective dynamics of "small-world" networks. *Nature*, **393**:440.
- [2] Milgram, S. (1967), The small world problem. *Psychology Today*, **2**:60.
- [3] Barabási, A.-L. and Albert, R. (1999), Emergence of scaling in random networks. *Science*, **286**:509.
- [4] Maslov, S. and Sneppen, K. (2002). Specificity and stability in topology of protein networks. *Science*, **296**:910.
- [5] Milo, R. et al. (2002), Network motifs: Simple building blocks of complex networks. *Science*, **298**:824.
- [6] Girvan, M. and Newman, M. E. (2002). Community structure in social and biological networks. *Proceedings of the National Academy of Sciences of the United States of America*, **99**:12.

- [7] Ahn, Y., Bagrow, J., and Lehmann, S. (2010). Link communities reveal multiscale complexity in networks. *Nature*, **466**(7307).
- [8] Lazer, D. et al. (2009). Computational social science. *Science*, **323**:5915.



Sune Lehmann er lektor ved DTU Compute og adjungeret lektor ved Niels Bohr Institutet.



Vedran Sekara er ph.d. fra DTU Compute, Danmarks Tekniske Universitet og nu Data Scientist hos Sony.

Kif årsmøde den 13. november 2015

Netværk for Kvinder i fysik (Kif) holder årsmøde den 13. november 2015 på Aarhus Universitet.

Programmet fokuserer på fastholdelse af kvinder indenfor forskning, med foredrag af professor Tine Jess, dr.med., fra *Taskorcen for Flere Kvinder i Forskning*, professor Meg Urry, Yale University og American Astronomical Society samt repræsentanter fra forskningsfonde.

Mødet varer fra kl. 10:30 til 17:30 og efterfølges af generalforsamling, hvor der vil være et forslag til afstemning om at åbne bestyrelsen for mandlige medlemmer samt kvinder der ikke har hovedfag i fysikrelaterede fag.

Til mødet vil årets Kif-prisvinder blive annonceret. Opslaget ses på Kifs hjemmeside: www.kif.nbi.dk.

Registrering til Kifs årsmøde foregår via link på Kifs hjemmeside (www.kif.nbi.dk). Sørg for at tage kolleger og studerende eller medstuderende med!