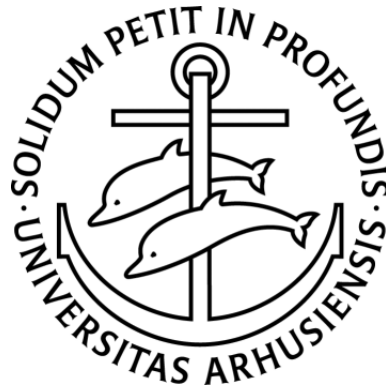


Aarhus Universitet

Speciale - juli 2007

Robusthed af netværk - med fokus på scale-free grafer



Lasse Østerlund Gram

årskortnr. 20000347

vejleder

Gerth Stølting Brodal

Abstract

Networks of various types and sizes occur everywhere in our daily life. Communication systems, the infrastructure, and the Internet are examples of networks playing a significant role in the society today. Breakdown of a given node or link in a network (for example an accident on the freeway or a terrorist attack) is likely to have great social and economic consequences; it is thus of crucial importance that networks function efficiently and are robust to accidents and attacks. This master thesis is a survey that considers several kinds of graphs, which can be used to model real life networks. The traditional way is to use the classic random graph, introduced by Erdős and Rényi, but recently the so-called scale-free graphs have been studied extensively; it turns out that these graphs are very suitable for modelling real life networks. This is the case because of the heterogeneous structure of scale-free graphs caused by the preferential attachment of new nodes to the existing ones. The robustness and efficiency of the classic and the scale-free graphs is the focus point of this thesis. We conclude that the scale-free graphs are superior with respect to random failures such as accidents, whereas the classic random graphs perform better with respect to attacks. In addition, we consider possible cascading effects as a result of overload after a breakdown and find that a node-capacity of 30% more than the initial load is sufficient to prevent cascading failures from happening.

Indhold

1	Introduktion	5
1.1	Opbygningen af specialet	6
2	Basal grafteori	7
2.1	Modellering af netværk	11
3	Robusthedsmål	13
3.1	Knude- og kantsammenhæng	13
3.2	Diameter	16
3.3	Kæmpekomponent	17
3.4	Effektivitet	17
4	Optimalt sammenhængende grafer	18
4.1	Graftyper	18
5	Stokastiske grafer	23
5.1	Klassisk stokastisk graf-model	23
5.1.1	Kæmpekomponent	25
5.1.2	Sammenhæng	26
5.2	Watts-Strogatz modellen	27
5.3	Scale-free modeller	30
5.3.1	Barabási og Alberts model	31
5.3.2	LCD-modellen	34
5.3.3	n -parring	36
5.3.4	Klemm-Equíluz modellen	37
6	Robusthed af scale-free grafer	42
6.1	Effektiviteten af et netværk	42
6.2	Robusthed over for fejl	47
6.2.1	Fejl i BA-graf	47
6.2.2	Fejl i KE-graf	51

6.3	Robusthed over for angreb	55
6.3.1	Angreb i BA-graf	56
6.3.2	Angreb i KE-graf	58
6.3.3	Konklusion på analysen	59
7	Kaskadeeffekter efter nedbrud	60
8	Konklusion	66
A	Stikordsregister	68
B	Notation	72

Kapitel 1

Introduktion

Netværk optræder overalt i vores hverdag om end i mange forskellige former. Infrastrukturen, telefonnettet og internettet er blot nogle få eksempler på netværk, som vi allesammen dagligt benytter. Det har stor betydning for samfundet, at disse netværk fungerer effektivt; f.eks. kan et uheld på motorvejen medføre kødannelser og dermed have stor indflydelse på den videre trafik. Det kan betyde, at folk kommer for sent på arbejde, hvilket igen kan betyde en omkostning for de berørte virksomheder. Et nedbrud i et netværk kan således få store konsekvenser for samfundet, så det er af afgørende betydning, at et netværk er så robust som muligt.

Robusthed af netværk er et emne, som på det seneste har fået meget opmærksomhed. En af grundene hertil er den voksende trussel fra terrorisme og den usikkerhed, det medfører. Jo mere robust et netværk er, jo mindre er risikoen for, at et evt. uheld eller terrorangreb vil få store følger for resten af netværket.

Det kan imidlertid være meget kostbart at designe et optimalt robust netværk (hvis ikke umuligt). Hvis f.eks. den internationale flytrafik fra Danmark skulle fungere optimalt med henblik på robusthed, ville det kræve, at alle lufthavne i Danmark skulle have et identisk udbud af ruter til udlandet. I så fald ville man ikke være afhængig af én bestemt lufthavn, da man i tilfælde af problemer stadig kunne komme frem til den ønskede destination via en anden lufthavn. Af både praktiske og økonomiske årsager er dette scenario åbenlyst ikke realistisk, så derfor fungerer Københavns Lufthavn som det naturlige knudepunkt for flytrafikken fra Danmark.

Situationen er typisk den, at man er stillet over for et netværk, som ikke umiddelbart kan gøres optimalt robust. Det er derfor interessant at undersøge,

hvor robust et eksisterende netværk er, da man så er bedre forberedt på virkningerne af et evt. uheld/angreb. Hvis man opdager svagheder i et givet netværk, har man muligheden for at kunne udbedre disse i tide og dermed forhindre et totalt sammenbrud af netværket.

For at kunne analysere robustheden af forskellige netværk udnytter vi, at et netværk kan modelleres ved hjælp af en *graf*. Vi kan herefter benytte resultater fra grafteori til at beskrive robustheden.

Dette speciale har til formål at give et overblik over forskellige graftyper anvendt til modellering af ”virkelige” netværk. Der fokuseres på robustheden af disse grafer over for såvel hændelige uheld som terrorangreb. Vi kommer i den forbindelse specielt ind på de såkaldte *scale-free* grafer, som er særligt interessante, idet mange virkelige netværk har samme struktur som scale-free grafer. Det er derfor særdeles relevant at studere robustheden af netop denne graftype.

1.1 Opbygningen af specialet

Da vi, som nævnt, benytter grafer til at beskrive forskellige netværk, er det nødvendigt med et basalt kendskab til forskellige begreber inden for grafteori, hvilket er formålet med kapitel 2. Begreberne og notationen tager primært udgangspunkt i [7].

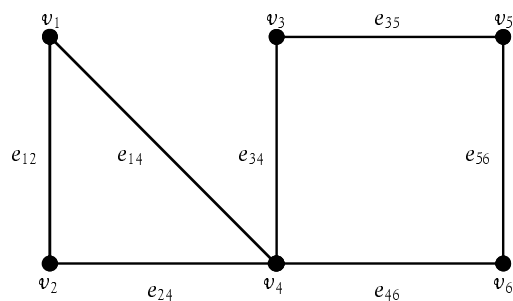
Kapitel 3 er en gennemgang af forskellige grafteoretiske robusthedsmål. Vi ser bl.a. på *sammenhæng*, *diameter* og *effektivitet*. I kapitel 4 undersøges hvilke graftyper, der er *optimalt sammenhængende*, og hvilke betingelser der skal være opfyldt, for at en graf har denne egenskab. Begrebet *stokastisk graf* introduceres i denne forbindelse. Kapitlet tager udgangspunkt i Dekker & Colbert [1].

I kapitel 5 beskrives stokastiske grafer nærmere, og vi kommer kort ind på nogle teoretiske resultater om de klassiske stokastiske grafer. Kilderne er her primært [10], [12], [13] og [17]. En scale-free graf er en bestemt type stokastisk graf, som i kapitel 5.3 bliver beskrevet og i kapitel 6 analyseret med henblik på robusthed og anvendelighed efter et eventuelt nedbrud som følge af enten en fejl eller et angreb. Crucitti et al. danner med deres simulationer i [16] grundlaget for kapitlet. I kapitel 7 diskuterer vi med baggrund i [20] mulige konsekvenser af et nedbrud i netværket. Endelig konkluderes på resultaterne i kapitel 8.

Kapitel 2

Basal grafteori

En graf kan beskrives som en samling af knuder og kanter. Vi bruger standardnotationen $G = (V, E)$, hvor G er grafen, V er mængden af knuder, og E er mængden af kanter. En graf består af n knuder og m kanter, medmindre andet er nævnt. En kant i E kan skrives som $e_l = (v_i, v_j)$; $v_i, v_j \in V$ eller blot e_{ij} for at illustrere, at kanten går mellem knuderne v_i og v_j . Til enhver kant hører der to knuder; disse knuder kaldes for kantens *endeknuder*, og man siger, at kanten er *incident* på dens endeknuder. Hvis $e_l = (v_i, v_j)$ er v_i og v_j således e_l 's endeknuder, og e_l er incident på v_i og v_j .

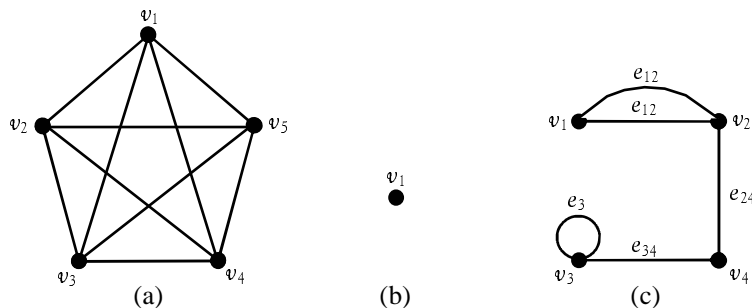


Figur 2.1 En vilkårlig graf med 6 knuder og 7 kanter.

To knuder er *naboer* (eller naboknuder), hvis de er endeknuder af en bestemt kant. Det vil med andre ord sige, at man kan komme fra den ene knude til den anden via én kant. Mængden af alle v 's naboer i en graf betegnes $\Gamma(v)$.

De kanter, som har samme par af endeknuder, siges at være *parallelle*. En *løkke* er en kant med identiske endeknuder. Hvis en graf hverken indeholder parallelle kanter eller løkker, er grafen *simpel*. Det antages ofte, at en graf

er simpel, men som vi skal se senere i specialet, er ikke-simple grafer også anvendelige til modellering af visse netværk. Andre specielle grafter er den *trivielle* graf og den *komplette* graf. En triviel graf er en graf, som blot indeholder én knude og ingen kanter. En komplet graf er en simpel graf, hvori ethvert par af knuder er naboer. En komplet graf med n knuder betegnes ofte K_n ; antallet af kanter i K_n er lig med $n(n - 1)/2$.



Figur 2.2 (a) den komplette graf K_5 , (b) den trivielle graf, (c) en ikke-simpel graf (v_3 har en løkke, og v_1 og v_2 har to parallelle kanter).

Begrebet *valensen* af en knude v er antallet af kanter incidente på v ; vi bruger betegnelsen $d(v)$ for valensen af knude v . Betragtes valensen af en knude v i en bestemt graf G , bruges også betegnelsen $d_G(v)$. Den mindste valens i en graf G betegnes $\delta(G)$, og den største valens betegnes $\Delta(G)$. Da enhver kant er incident på to knuder, bidrager den med 2 til summen af valenserne i G . Heraf følger, at summen af valenserne i en graf G er lig med $2m$, hvor m er antallet af kanter i G . En graf, hvori alle knuder har samme valens, kaldes en *regulær* graf. Det følger umiddelbart, at K_n også er regulær, da $d(K_n) = \Delta(K_n) = n - 1$.

Grafen $G' = (V', E')$ er en *delgraf* af $G = (V, E)$, hvis V' og E' er delmængder af henholdsvis V og E , sådan at en kant (v_i, v_j) kun er i E' , hvis v_i og v_j er i V' . $G' = (V', E')$ er en *ægte* delgraf af $G = (V, E)$,¹ hvis enten E' er en ægte delmængde af E , eller V' er en ægte delmængde af V . E' siges at være en ægte delmængde af E , hvis alle kanter i E' er indeholdt i E , men der *mindst* eksisterer én kant i E , som ikke er i E' . Tilsvarende gælder for V' og V . $G' \subseteq G$ siges at være en *maksimal* delgraf af G med hensyn til en egenskab \mathcal{Q} , hvis G' opfylder \mathcal{Q} og ikke er en ægte delgraf af en anden delgraf af G , som også opfylder \mathcal{Q} . Tilsvarende siges en delgraf G' at være *minimal* mht. \mathcal{Q} , hvis G'

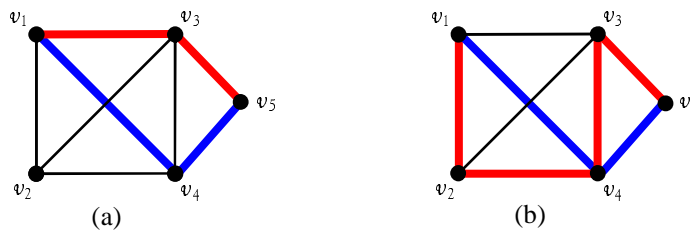
¹Hvis G' er en delgraf af G skrives også $G' \subseteq G$, og hvis G' er en ægte delgraf af G skrives $G' \subset G$.

opfylder \mathcal{Q} , og ingen delgraf af G med egenskaben \mathcal{Q} er en ægte delgraf af G' .

En *tur* i en graf G er en endelig sekvens af alternerende knuder og kanter:

$$v_0, e_1, v_1, e_2, \dots, v_{k-1}, e_k, v_k,$$

hvor knuderne kan besøges mere end en gang, men kanterne er forskellige. Hvis endeknuderne er forskellige, er turen *åben* - ellers er den *lukket*. En *kreds* er en lukket tur, hvor alle knuder med undtagelse af endeknuderne er forskellige. Antallet af knuder i en kreds er lig med antallet af kanter i kredsen. En *sti* er en åben tur, hvor alle knuder er forskellige. Antallet af knuder i en sti er lig med $k + 1$, hvor k betegner antallet af kanter i stien. To stier P_1 og P_2 siges at være *knudedisjunkte*, hvis P_1 og P_2 højst har endeknuderne tilfælles; stierne er *kantdisjunkte*, hvis P_1 og P_2 ikke har nogen kanter tilfælles. Se figur 2.3 for illustration.

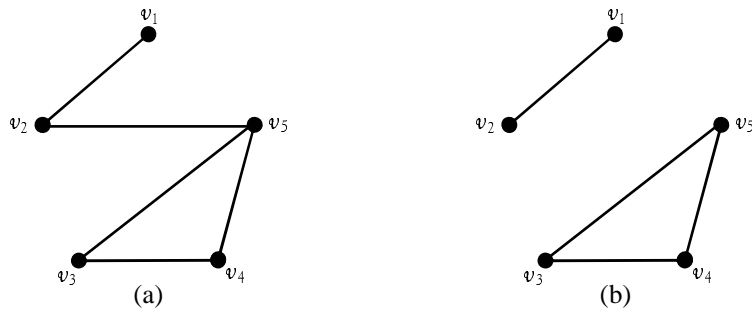


Figur 2.3 (a) to knudedisjunkte stier, (b) to kantdisjunkte stier.

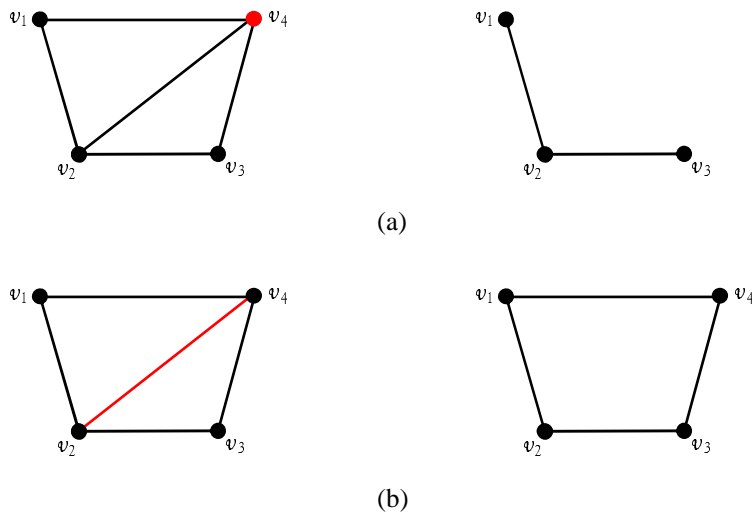
To knuder v_i og v_j i en graf G er *sammenhængende*, hvis der eksisterer en sti mellem de to knuder. Grafen G er sammenhængende, hvis der eksisterer en sti mellem *ethvert* par af knuder i grafen; i modsat fald er G usammenhængende. En *sammenhængskomponent* af en graf G er en maksimalt sammenhængende delgraf af G . Ved en maksimalt sammenhængende delgraf af G forstås, at sammenhængskomponenten ikke er en ægte delgraf af andre sammenhængende delgrafer af G . Hvis G er sammenhængende, er sammenhængskomponenten af G således hele G .

Hvis en knude v_i fjernes i en graf G , resulterer det i en graf $G - v_i$, hvor både v_i og alle kanter incidente på knuden er slettet. Hvis en kant e_i fjernes, resulterer det i en graf $G - e_i$, hvor det *kun* er kanten, der er væk; endeknuderne til e_i fjernes således ikke. I figur 2.5(a) og 2.5(b) illustreres henholdsvis fjernelse af en knude og en kant.

En *vægtet* graf er en graf, hvor hver knude og/eller kant er tilknyttet en



Figur 2.4 (a) en sammenhængende graf, (b) en graf bestående af to sammenhængskomponenter.



Figur 2.5 (a) fjernelse af den røde knude, (b) fjernelse af den røde kant.

bestemt værdi. Værdien på kanterne kan betragtes som den *strøm*, der flyder igennem grafen, og værdien på knuderne kan betragtes som den *kapacitet*, der er til rådighed.

En graf $G = (V, E)$ med n knuder og m kanter kan beskrives ved en $(n \times n)$ -matrix med indgange $[e_{ij}]$, hvor

$$e_{ij} = \begin{cases} 1, & \text{hvis } (v_i, v_j) \in E; \\ 0, & \text{ellers.} \end{cases}$$

Matricen illustrerer således direkte imellem hvilke knuder, der findes en kant. Grafen i figur 2.4(a) kan beskrives ved følgende matrix:²

$$\begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 1 & 0 \end{bmatrix}$$

Hvis grafen er vægtet, kan matricen blot modificeres, så vægten for hver kant står i stedet for ettallet i definitionen for e_{ij} .

2.1 Modellering af netværk

En graf er særdeles velegnet til at beskrive adskillige former for netværk.³ Hvis vi f.eks. skal modellere et trafiknetværk, kan kanterne repræsentere veje, og knuderne kan repræsentere trafikryds. På tilsvarende vis kan en lang række andre netværk modelleres vha. grafer - f.eks. telefonnettet, tognettet, flynettet, elnettet og internettet.

Det vil ofte være mest naturligt at modellere virkelige netværk ved hjælp af *orienterede* grafer. I en orienteret graf har alle kanter tilknyttet en retning, så en kant $e_l = (v_i, v_j)$ går fra v_i til v_j . I trafikeksemplet svarer det til en ensrettet vej. Ønsker man, at trafikken skal gå i begge retninger, kan dette også modelleres i en orienteret graf ved blot at tilføje en kant $e'_l = (v_j, v_i)$, som går fra v_j til v_i . En ikke-orienteret graf G kan derfor repræsenteres ved en orienteret graf G' , hvor enhver kant i G er erstattet af to nye orienterede kanter (orienteret i hver sin retning).

En analyse af orienterede grafer er typisk mere vanskelig og mindre overskuelig end en tilsvarende analyse for ikke-orienterede grafer. Det er en af grundene til, at man i litteraturen som regel vælger at simplificere problemstillingen ved at koncentrere sig om den underliggende graf (dvs. grafen uden retninger på kanterne). Når vi f.eks. i kapitel 5.3 skal beskrive processen for at generere en scale-free graf, får vi bl.a. brug for at kende valensen af en knude. Hvis vi i stedet havde at gøre med en orienteret graf, skulle vi bl.a. tage højde for både ind- og ud-valensen, hvilket besværliggør processen (se

²Matricen er pr. definition altid symmetrisk for ikke-orienterede grafer.

³Forskellen på "graf" og "netværk" ligger i, at "graf" er det teoretiske og "netværk" det praktiske begreb. Forskellige former for netværk beskrives i specialet, men med udgangspunkt i grafteori.

[6] for en beskrivelse af orienterede scale-free grafer). Som før nævnt kan man ud fra resultaterne om ikke-orienterede grafer konstruere en orienteret graf, som har samme egenskaber. I dette speciale vil der således også kun blive fokuseret på ikke-orienterede grafer.

I kapitel 3, 4, 5 og 6 vil alle grafer være ikke-vægtede, da det er selve konstruktionen, der er vigtig i disse kapitler. De vægtede grafer bliver interessante i kapitel 7, hvor vi ser på strømmen i netværket.

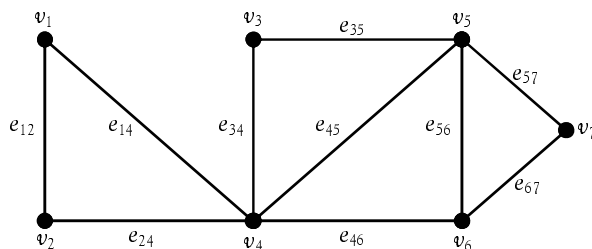
Kapitel 3

Robusthedsmål

Der findes forskellige metoder til at angive robustheden af en graf. Dette kapitel beskæftiger sig med *knude-* og *kantsammenhæng*, *diameteren* af en graf, begrebet *kæmpekomponent* samt *effektiviteten* af en graf.

3.1 Knude- og kantsammenhæng

Jævnfør kapitel 2 er en graf $G = (V, E)$ sammenhængende, hvis der eksisterer en sti mellem ethvert par af knuder i V . I et toget vil det betyde, at man skal kunne komme fra en vilkårlig station til enhver anden på *mindst* en måde. Robusthed af en graf G hænger nøje sammen med sammenhængen af G . Det er derfor interessant at undersøge, hvor ”godt” G er sammenhængende. Vi vil med andre ord finde ud af, hvor mange knuder eller kanter der skal fjernes, før G ikke længere er sammenhængende. Det vil sige, hvor mange togstationer eller strækninger der skal være fejl på, før toget bryder sammen.



Figur 3.1 En vilkårlig graf med 7 knuder og 10 kanter.

Et *knude-snit* i en sammenhængende graf G er en delmængde af knuderne, hvis fjernelse gør den resterende del af G usammenhængende. Det *mindste*

antal af knuder, hvis fjernelse gør G usammenhængende (eller resulterer i den trivielle graf med kun en knude), kaldes *knudesammenhængen* af G og betegnes $\kappa(G)$. For den komplette graf K_n , kan vi fjerne op til $n - 2$ knuder og stadig have en sammenhængende graf; derfor er $\kappa(K_n) = n - 1$. Grafen i figur 3.1 bliver usammenhængende, hvis knude v_4 fjernes; knudesammenhængen i denne graf er således 1. Når $\delta(G)$ betegner den mindste valens i en simpel, sammenhængende graf G , gælder der følgende øvre grænse for knudesammenhængen:

Lemma 3.1.1 *I en simpel graf G gælder:*

$$\kappa(G) \leq \delta(G)$$

Bevis:

Betragt en vilkårlig knude v i en simpel, sammenhængende graf $G = (V, E)$ og lad $\Gamma(v)$ betegne mængden af naboer til v . Fjernelse af $\Gamma(v)$ vil enten resultere i den trivielle graf kun bestående af knude v eller i en usammenhængende graf, hvor v ikke er forbundet til andre knuder. Det betyder med andre ord, at antallet af knuder, der skal fjernes for at få en usammenhængende graf, højst er lig med antallet af knuder i $\Gamma(v)$, det vil sige $|\Gamma(v)|$. Vi har således:

$$\kappa(G) \leq |\Gamma(v)|, \quad \forall v \in V .$$

Da G er en simpel graf, er antallet af knuder i $\Gamma(v)$ lig med valensen af v ; det vil sige:

$$|\Gamma(v)| = d(v) .$$

Heraf følger, at:

$$\kappa(G) \leq \min_{v \in V} d(v) = \delta(G) . \quad \square$$

Et *kant-snit* i en sammenhængende graf G er på tilsvarende vis en delmængde af kanterne, hvis fjernelse resulterer i en usammenhængende graf. Antallet af kanter i det *mindst* mulige kant-snit kaldes *kantsammenhængen* og betegnes $\kappa'(G)$. Da man kan gøre enhver graf usammenhængende ved at fjerne alle kanter incidente på en vilkårlig knude, har vi åbenlyst følgende ulighed:

$$\kappa'(G) \leq \delta(G) .$$

I figur 3.1 er kantsammenhængen af grafen lig med 2, idet fjernelse af kanterne incidente på v_1, v_2, v_3 eller v_7 vil resultere i en usammenhængende graf. Følgende sætning¹ angiver sammenhængen mellem knude- og kantsammenhængen samt den mindste valens i en graf G :

¹Theorem 8.5 i [7].

Sætning 3.1.2 *I en simpel graf G gælder:*

$$\kappa(G) \leq \kappa'(G) \leq \delta(G)$$

Bevis:

Anden del af uligheden er allerede vist, så vi mangler kun at vise, at $\kappa(G) \leq \kappa'(G)$. Hvis G er usammenhængende, er $\kappa(G) = \kappa'(G) = 0$, hvormed uligheden er opfyldt.

Lad nu G være sammenhængende. Hvis $\kappa'(G) = 1$, da har G en bro e , hvilket vil sige en kant i en graf G , så $G - e$ har flere sammenhængskomponenter end G . Broen er incident på to knuder; hvis en af disse knuder fjernes, resulterer det i en usammenhængende graf (eller den trivielle graf, hvis e er den eneste kant i G). Derfor er uligheden $\kappa(G) \leq \kappa'(G)$ også opfyldt i dette tilfælde.

Antag at $\kappa'(G) \geq 2$. Der eksisterer da $\kappa'(G)$ kanter i G , hvis fjernelse resulterer i en usammenhængende graf. Hvis $\kappa'(G) - 1$ af disse kanter fjernes, eksisterer der en bro $e = (v_1, v_2)$ i den resulterende graf. For hver af disse $\kappa'(G) - 1$ kanter vælges en endeknude forskellig fra v_1 og v_2 . Hvis disse knuder fjernes, svarer det til *mindst* at fjerne de $\kappa'(G) - 1$ kanter fra G . Hvis den resulterende graf er usammenhængende, har vi:

$$\kappa(G) \leq \kappa'(G) - 1 .$$

Hvis grafen stadig er sammenhængende, vil $e = (v_1, v_2)$ være bro, og en fjernelse af enten v_1 eller v_2 vil resultere i en usammenhængende graf (eller den trivielle graf). Det betyder, at $\kappa(G) \leq \kappa'(G)$. Vi har således i alle tilfælde, at:

$$\kappa(G) \leq \kappa'(G) \leq \delta(G), \quad \square$$

Det ses let, at $\kappa(G) = \kappa'(G) = \delta(G)$, hvis G er komplet. Hvis G er en generel graf, er det ikke helt så let at finde $\kappa(G)$ og $\kappa'(G)$. Esfahanian skriver i [8] om en række algoritmer, der kan anvendes til udregning af hhv. knude- og kantsammenhængen for en generel graf; for den historisk interesserede er der endvidere angivet en kronologisk liste med tidspunktet for de enkelte algoritmer, hvilket fint illustrerer udviklingen igennem årene.

Et fundamentalt resultat i grafteori om sammenhængen af en graf G er Mengers sætning, som her angives uden bevis:²

²Sætning og bevis kan bl.a. læses i kap 8.4 og 12.10 i [7].

Sætning 3.1.3 (*Mengers sætning*)

- i) Knudesammenhængen $\kappa(G)$ er lig med det mindste antal af knudedisjunkte stier mellem ethvert par af knuder i G .*
- ii) Kantsammenhængen $\kappa'(G)$ er lig med det mindste antal af kantdisjunkte stier mellem ethvert par af knuder i G .*

Disse resultater har bl.a. dannet grundlaget for forskellige algoritmer til udregning af henholdsvis knude- og kantsammenhængen af en graf; algoritmerne vil dog ikke blive studeret nærmere i dette speciale. I stedet henvises til f.eks. [8].

Som et eksempel på anvendelsen af Mengers sætning, betragter vi grafen i figur 3.1. For at finde knudesammenhængen skal vi iflg. punkt 1 i Mengers sætning undersøge antallet af knudedisjunkte stier mellem ethvert par af knuder i grafen. Det ses umiddelbart, at stierne fra hhv. v_1 og v_2 til enten v_3, v_5, v_6 eller v_7 alle går igennem v_4 . Det mindste antal af knudedisjunkte stier er således lig med 1, hvilket betyder, at knudesammenhængen i grafen er 1. Antallet af kantdisjunkte stier mellem ethvert par af knuder, hvor den ene knude er v_1, v_2, v_3 eller v_7 , er lig med 2. For kombinationer af v_4, v_5 og v_6 er antallet af kantdisjunkte stier lig med 3. Ved anvendelse af punkt 2 i Mengers sætning kan vi derfor konkludere, at kantsammenhængen af grafen er lig med 2.

3.2 Diameter

Længden af den korteste sti mellem to knuder u og v i en graf G kaldes også for *afstanden* mellem u og v og betegnes $d(u, v)$. Hvis der ikke eksisterer en sådan sti, defineres $d(u, v) = \infty$. *Diameteren* af G betegnes $\text{diam}(G)$ og er defineret ved:

$$\text{diam}(G) = \max\{d(u, v) \mid u, v \in V\} .$$

Diameteren i en graf er således lig med afstanden mellem de to knuder, der er længst fra hinanden. Grafen i figur 3.1 har en diameter på 3, hvilket svarer til afstanden mellem knuderne v_1 og v_7 . Den komplette graf har en diameter på 1, da alle knuder er forbundet direkte med hinanden. Så snart vi fjerner en kant i den komplette graf, vokser diameteren, og grafen bliver mindre robust, da der bliver længere mellem knuderne. Intuitivt er en graf med en lille diameter mere robust end en graf med en stor diameter, hvilket underbygges i kapitel 5.2 og 6. For eksempel vil et toget med en lille diameter være at

foretrække, da man undgår for mange mellemstationer undervejs og derfor er mindre påvirket af eventuelle problemer i resten af tognettet. Diameteren er særligt vigtig ved lokalisering af faciliteter, som skal nå ud til husstande eller virksomheder inden for en bestemt afstand - det kan f.eks. være mobilsendere og redningskorps.

Som et alternativ til diameteren anvendes også den *gennemsnitlige afstand*, der tager højde for alle afstande i stedet for kun den maksimale. En nærmere beskrivelse af den gennemsnitlige afstand sker i kapitel 5.2.

3.3 Kæmpekomponent

I robusthedssammenhænge er det interessant at undersøge, hvor stor den største sammenhængskomponent er efter et eventuelt nedbrud i en knude. Hvis denne sammenhængskomponent er lille, betyder det, at grafen ikke har været særlig robust. Hvis sammenhængskomponenten i stedet er stor, har nedbruddet ikke haft den store betydning for det samlede netværk, og grafen er derfor mere robust. En sammenhængskomponent siges iflg. Bollobás [10] at være stor, hvis den indeholder mere end $n^{2/3}$ knuder og lille, hvis den indeholder under $\frac{n^{2/3}}{2}$ knuder, hvor n er antallet af knuder i G .

Begrebet *kæmpekomponent* er betegnelsen for den i G største sammenhængskomponent med flere end $n^{2/3}$ knuder. Hvis den største sammenhængskomponent består af færre end $n^{2/3}$ knuder, har grafen ikke en kæmpekomponent. I kapitel 5.1.1 kommer vi ind på, hvornår en stokastisk graf indeholder en kæmpekomponent.

3.4 Effektivitet

Et robusthedsmål, vi kommer til at beskæftige os med i kapitel 6 og 7, er *effektiviteten* af en graf. Det er ikke et traditionelt robusthedsmål som f.eks. kant- og knudesammenhæng, men særdeles velegnet i praksis til at undersøge, hvordan et netværk fungerer. Begrebet blev introduceret af Latora og Marchiori i [15] som et alternativ til den gennemsnitlige afstand og er intuitivt et udtryk for, hvor ”godt” et netværk fungerer i forhold til den optimale situation, hvor grafen er komplet. Kapitel 6 indeholder en mere præcis beskrivelse af effektiviteten samt en anvendelse af begrebet på forskellige netværk med henblik på at finde den netværkstype, som efter et nedbrud fungerer mest effektivt.

Kapitel 4

Optimalt sammenhængende grafer

Indtil nu har vi introduceret forskellige basale begreber og robusthedsmål inden for grafteori. Hvis vi frit kan designe et netværk (uden hensyntagen til omkostninger) og ønsker, at sammenhængen skal være optimal, hvordan skal netværket da se ud? Vi ser i dette kapitel på nogle bestemte grafter, som er kendetegnet ved en maksimal knude- og kantsammenhæng, og som dermed er optimalt sammenhængende (jvf. definition af Dekker i [1]). Det betyder med andre ord, at der for en graf G skal gælde:

$$\kappa(G) = \kappa'(G) = \delta(G) ,$$

da vi af sætning 3.1.2 har $\kappa(G) \leq \kappa'(G) \leq \delta(G)$. Den komplette graf opfylder som nævnt i kapitel 3.1 denne optimalitetsbetingelse, men som vi skal se i det følgende, er der også andre grafter, som kan siges at være optimalt sammenhængende.

4.1 Grafter

For at kunne beskrive forskellige grafter, som er optimalt sammenhængende, er det nødvendigt at kende til følgende begreber inden for grafteorien.¹

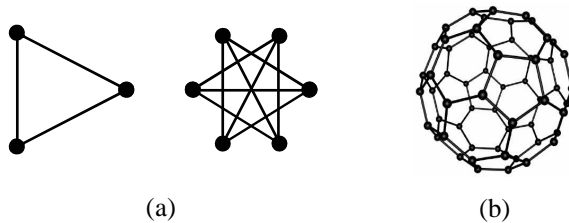
Definition 4.1.1

i) En graf er regulær, hvis alle knuder har samme valens.

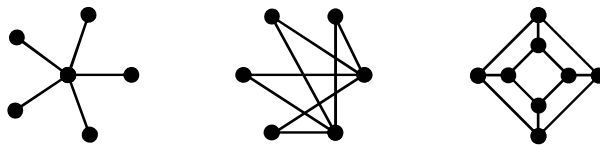
¹Vi har tidligere nævnt *regulære grafer*, men for fuldstændighedens skyld gentages definitionen her.

- ii) En automorfi π i en graf er en permutation af knuderne, som bevarer kanterne. Det vil med andre ord sige, at to knuder v_i og v_j i en graf er forbundet med en kant, hvis og kun hvis $\pi(v_i)$ og $\pi(v_j)$ også er forbundet med en kant.²
- iii) En graf er knude-transitiv, hvis der for ethvert par af knuder v_i og v_j eksisterer en automorfi π , så $\pi(v_i) = v_j$.
- iv) En graf er kant-transitiv, hvis der for ethvert par af kanter e_i og e_j eksisterer en automorfi π , så $\pi(e_i) = e_j$.
- v) En graf er symmetrisk, hvis den både er knude- og kant-transitiv.

Lidt mindre formelt siger man, at alle knuderne i en knude-transitiv graf ser ens ud, ligesom det er tilfældet med alle kanterne i en kant-transitiv graf. En symmetrisk graf er således en graf, hvor både knuderne og kanterne ser ens ud. Figur 4.1 og 4.2 illustrerer forskellige knude- og kant-transitive grafer.



Figur 4.1 (a) to eksempler på knude-transitive grafer med hhv. 3 og 6 knuder, (b) den såkaldte fodboldgraf, som også er knude-transitiv.



Figur 4.2 Tre eksempler på kant-transitive grafer.

Da en knude-transitiv graf består af "ens" knuder, medfører det naturligvis, at valensen af alle knuder er identisk. Det betyder dermed, at alle knude-transitive grafer også er regulære. Det omvendte er imidlertid ikke altid tilfældet; grafen i figur 4.3 er således regulær, men ikke knude-transitiv, da

² $\pi(v_i)$ betegner her knude v_i efter permutationen.

knuderne i midten ikke har samme struktur som knuderne i enderne. Alle symmetriske grafer er pr. definition knude-transitive, men fodbold-grafen illustreret i figur 4.1 er et eksempel på en knude-transitiv graf, som ikke er symmetrisk. I fodbold-grafen er enhver knude lig med skæringen af en femkant og to sekskanter, hvilket medfører knude-transitivitet; til gengæld forbinder nogle kanter en femkant og en sekskant, mens resten forbinder to sekskanter. Det betyder, at grafen ikke er kant-transitiv og dermed heller ikke symmetrisk.



Figur 4.3 Regulær graf som ikke er knude-transitiv.

En knude-transitiv graf er velegnet til konstruktion af et netværk, hvor alle knuder er lige vigtige. På den måde undgås endvidere, at bestemte knuder er særligt udsatte for eksempelvis terrorangreb. Følgende sætning beskriver sammenhængen mellem knude- og kantsammenhængen i en knude-transitiv graf.

Sætning 4.1.1 For enhver sammenhængende knude-transitiv graf G gælder der følgende:

- i)* $\kappa'(G) = \delta(G)$
- ii)* $\kappa(G) \geq \frac{2}{3}(\delta(G) + 1)$
- iii)* Hvis $\delta(G) \leq 4$, da er $\kappa(G) = \delta(G)$.
- iv)* Hvis G er symmetrisk, da er $\kappa(G) = \delta(G)$.

Bevis:

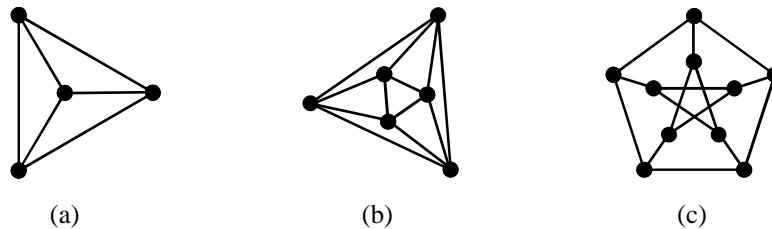
(*i*), (*ii*) og (*iv*): Kapitel 3 i Godsil & Royle [21] eller sætning 3.7 i Babai [22].

(*iii*): Af sætning 3.1.2 har vi for enhver simpel graf følgende ulighed: $\kappa(G) \leq \kappa'(G) \leq \delta(G)$. Hvis $\delta(G) = 4$ har vi af (*ii*), at $\kappa(G) \geq \frac{2}{3} \cdot (4 + 1) = \frac{10}{3}$, hvilket betyder $\kappa(G) = 4$, da $\kappa(G) \leq \delta(G)$. På tilsvarende vis fås $\kappa(G) = \delta(G)$ for

$\delta(G) \in \{1, 2, 3\}$. \square

Bemærk at der i sætningen gælder $\delta(G) = d_G(v)$, for alle $v \in V$, da G er knude-transitiv og dermed også regulær. Hvis en graf er knude-transitiv, følger det således af første punkt i sætningen, at grafen er optimalt robust med hensyn til nedbrud på kanterne. Af (iii) har vi tillige, at grafen er optimalt sammenhængende, hvis valensen er højst 4. Denne vigtige egenskab gør det bl.a. let at konkludere, at fodbold-grafen i figur 4.1 er optimalt sammenhængende med $\kappa(G) = \kappa'(G) = \delta(G) = 3$. På tilsvarende måde kan optimalitet let sluttet af en række andre knude-transitive grafer.

Hvis valensen af knuderne er større end 4, er en knude-transitiv graf ikke nødvendigvis optimalt sammenhængende. Det sidste punkt i sætning 4.1.1 betyder imidlertid, at alle symmetriske grafer - uanset valens - er optimalt sammenhængende. Figur 4.4 viser forskellige eksempler på grafer, som opfylder symmetri-kriteriet; heriblandt et tetraeder, et oktaeder og den velkendte Petersen graf.



Figur 4.4 (a) tetraeder, (b) oktaeder, (c) Petersen graf.

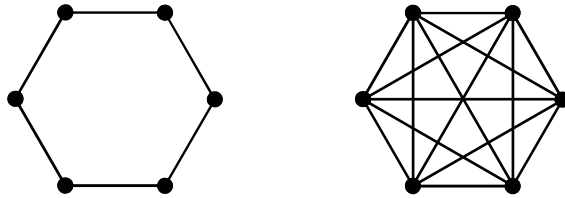
En *stokastisk* graf er en speciel graftype, som bliver beskrevet nærmere i kapitel 5. Kendetegnet ved denne graftype er groft sagt, at knuderne forbindes tilfældigt. Denne forbindelsesmetode medfører lidt overraskende, at en stokastisk graf hører til blandt mængden af optimalt sammenhængende grafer, når antallet af knuder er stort nok. Vi angiver følgende resultat uden bevis og henviser i stedet til Dekker [1] og Bollobás [10] for detaljerne:

Sætning 4.1.2 *Sandsynligheden for at en stokastisk graf med n knuder er optimalt sammenhængende går mod 1, for $n \rightarrow \infty$.*

Vi kan af dette kapitel konkludere, at et netværk kan designes optimalt med hensyn til knude- og kantsammenhæng, hvis man benytter en af nedenstående muligheder:

- En knude-transitiv graf hvor valensen af knuderne er højst 4.
- En symmetrisk graf.
- En stokastisk graf med tilstrækkeligt mange knuder.³

Selve robustheden af netværket hænger selvfølgelig nøje sammen med antallet af kanter. Figur 4.5 afbilder to symmetriske grafer, som begge er optimalt sammenhængende, men figuren til højre er åbenlyst den mest robuste med hensyn til nedbrud.



Figur 4.5 *To symmetriske grafer.*

Den stokastiske graf er særlig interessant at studere nærmere. Foruden egenskaben at være optimalt sammenhængende har denne graftype dannet grundlag for en lang række modeller til modellering af virkelige netværk. Stokastiske grafer er derfor emnet for det kommende kapitel.

³Dekker skriver i [1], at en stokastisk graf med 30 knuder med sandsynlighed 99,98% er optimalt sammenhængende. Med 100 knuder er sandsynligheden ca. 99,9999999%.

Kapitel 5

Stokastiske grafer

En stokastisk graf er betegnelsen for en graf, hvor en samling knuder forbindes tilfældigt med et antal kanter. Der er en lang række grafter, som hører ind under begrebet stokastiske grafer; vi beskæftiger os i dette kapitel med den klassiske stokastiske graf-model, Watts-Strogatz modellen samt tre scale-free modeller¹ og diskuterer modellernes egenskaber samt relevans med henblik på virkelighedens netværk.

5.1 Klassisk stokastisk graf-model

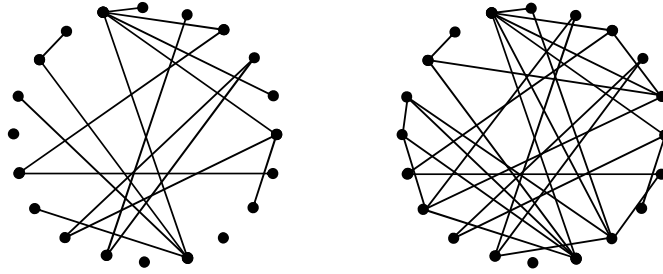
Stokastiske grafer optrådte første gang i slutningen af 1950'erne, da Erdős og Rényi forskede i udseendet af en "typisk" graf med n knuder og m kanter. Det resulterede i en sandsynlighedsteoretisk model, hvor udfaldsrummet udgøres af alle de mulige grafer, og hvor hver graf betragtes som en stokastisk variabel.² Erdős og Rényi studerede udfaldsrummet $\mathcal{G}(n, m)$, som består af alle grafer med knudemængde $V = \{1, \dots, n\}$ og kantmængde $E = \{1, \dots, m\}$, hvor hver graf har samme sandsynlighed.³ Antallet af mulige kanter mellem knuderne i V er $N = \binom{n}{2}$, så antallet af mulige grafer i $\mathcal{G}(n, m)$ er $\binom{N}{m}$. Sandsynligheden for udfaldet af en bestemt graf med n knuder og m kanter er således $\binom{N}{m}^{-1}$.⁴ Vi bruger betegnelsen $G_{n,m}$ eller blot G_m for en stokastisk graf i $\mathcal{G}(n, m)$. I det følgende antages det, at antallet af kanter er en funktion af antallet af knuder; det vil sige m er en funktion af n .

¹BA-, LCD- og KE-modellen.

²Artiklen af Erdős og Rényi, hvor stokastiske grafer første gang optræder, er fra 1959 og har titlen: "On Random Graphs" i Publ. Math. Debrecen 6, s. 290-297.

³Bollobás [10], side 34.

⁴En sådan klassisk stokastisk graf betegnes ofte ER-graf (efter Erdős og Rényi).



Figur 5.1 To stokastiske grafer med 20 knuder og hhv. 16 og 28 kanter.

Vi er interesseret i at undersøge, hvilke egenskaber en stokastisk graf har, når antallet af knuder er stort. Til det formål bruger vi begrebet ”med stor sandsynlighed”. Jævnfør Bollobás [11] siger vi, at en graf med n knuder har en egenskab \mathcal{Q} med stor sandsynlighed, hvis:

$$P(\mathcal{Q}) \rightarrow 1, \text{ for } n \rightarrow \infty .$$

En egenskab \mathcal{Q} er *monoton*,⁵ hvis

$$G \in \mathcal{Q} \text{ og } G \subset H \Rightarrow H \in \mathcal{Q} .$$

Ovenstående betyder, at hvis en graf G har egenskaben \mathcal{Q} , og G er en delgraf af H (hvor G og H begge har n knuder), da har H også egenskaben \mathcal{Q} . Sammenhængen af en graf er således et eksempel på en monoton egenskab. De fleste monotone egenskaber opstår ret pludseligt; forstået på den måde, at for et bestemt m har en graf G_m med stor sandsynlighed *ikke* egenskaben \mathcal{Q} , mens grafen G_{m^*} for et $m^* > m$ med stor sandsynlighed har \mathcal{Q} .⁶

Det interessante er at undersøge, på hvilket tidspunkt en graf har en given egenskab. Til det formål ser vi på en *graf-proces*, som er en slags organisme, der udvikler sig ved, at flere og flere kanter lægges til ”begyndelsesgrafen” (en tom graf med n knuder). Mere formelt er en grafproces en Markov kæde⁷ $\tilde{G} = (G_t)_0^\infty$, hvor tilstandene er grafer på knudemængden $V = \{1, \dots, n\}$. Processen starter med den tomme graf på V . For $1 \leq t \leq \binom{n}{2}$ fremkommer grafen (tilstanden) G_t ved at lægge en ny kant til grafen G_{t-1} , hvor alle nye kanter har samme sandsynlighed for at blive valgt. Grafen G_t har således præcis t kanter, så ved at vælge $t = m$ fås grafen G_m . Alle de mulige grafer, som

⁵Bollobás [10], side 36.

⁶Bollobás [10], side 40.

⁷En stokastisk proces er en Markov kæde, hvis sandsynligheden for en bestemt hændelse på tidspunkt t udelukkende afhænger af tilstanden på tidspunkt $t - 1$.

kan fremkomme på tidspunkt $t = m$, udgør præcist udfaldsrummet $\mathcal{G}(n, m)$, så grafprocessen kan betragtes som en fremgangsmåde til konstruktion af en stokastisk graf med n knuder og m kanter.

Tidspunktet, hvor grafen G_t første gang har en given monoton egenskab \mathcal{Q} , kaldes *stoptiden*. Stoptiden τ er defineret ved:

$$\tau = \tau_{\mathcal{Q}} = \min\{t \geq 0 \mid G_t \in \mathcal{Q}\} .$$

Hvis ønsket er, at grafen skal have en bestemt egenskab, er det således ikke nødvendigt at benytte flere kanter end τ .

For at kunne sige noget generelt om, hvad der kræves, for at en graf med stor sandsynlighed har en bestemt egenskab \mathcal{Q} , introduceres begrebet *tærskelfunktion*. Tærskelfunktionen for en monoton egenskab \mathcal{Q} betegnes $m^*(n)$; den giver det antal kanter, som er grænsen for, hvornår en stokastisk graf med stor sandsynlighed enten har \mathcal{Q} eller ikke har \mathcal{Q} . Hvis $m = m^*(n) + \omega(n)$ hvor $\omega(n) \rightarrow \infty$, da er $G_m \in \mathcal{Q}$ med stor sandsynlighed, og hvis der omvendt gælder $m = m^* - \omega(n)$, så er $G_m \notin \mathcal{Q}$ med stor sandsynlighed. Dette gælder, uanset hvor hurtigt $\omega(n) \rightarrow \infty$; eneste krav er blot, at $\omega(n)$ er en voksende funktion.

En vigtig egenskab ved stokastiske grafprocesser er, at graferne omkring tidspunkt $n/2$ begynder at indeholde en kæmpekomponent og kort tid derefter med stor sandsynlighed er sammenhængende. Dette vil blive præciseret i de følgende to afsnit. En anden vigtig egenskab ved stokastiske grafer, som vi ikke vil gå i detaljer med her, er, at diameteren typisk er betydeligt mindre end ved f.eks. regulære grafer.⁸

5.1.1 Kæmpekomponent

Dette afsnit omhandler tidspunktet for, hvornår en stokastisk graf med stor sandsynlighed indeholder en kæmpekomponent. Ud over at være interessant i sig selv, er afsnittet også væsentligt for at kunne forstå relationen mellem kæmpekomponenten og sammenhængen af en graf (kapitel 5.1.2). For at simplificere tingene udelader vi beviserne i dette afsnit og koncentrerer os i stedet om betydningen af sætningerne. Beviserne kan alle læses i Bollobás [10].

Vi starter med et resultat, som giver en grænse for, hvornår en stokastisk graf udelukkende består af store og små sammenhængskomponenter.

⁸Vi kommer nærmere ind på regulære grafer i kapitel 5.2.

Sætning 5.1.1 Lad $s_0 = (2 \log n)^{1/2} n^{2/3}$ og sæt $k_0(s) = \lfloor (3 \log s - \log n - 2 \log \log n) n^2 / s^2 \rfloor$ for $s \neq 0$. Da er en grafproces $\tilde{G} = (G_t)_0^N$ med stor sandsynlighed sådan, at hvis $t = n/2 + s$, $s_0 \leq |s| < n/2$ og G_t har en sammenhængskomponent af størrelse k , da er enten $k < k_0(s)$ eller $k > n^{2/3}$. For $t_0 = \lceil n/2 + s_0 \rceil \leq t \leq n$ gælder der med stor sandsynlighed, at grafen G_t ikke har nogen sammenhængskomponenter af størrelse mellem $\frac{n^{2/3}}{2}$ og $n^{2/3}$.

Bevis: Bollobás [10] - sætning 6.1. \square

Sætningen betyder, at enhver sammenhængskomponent af G_t i en grafproces, hvor $t \geq t_0 = \lceil n/2 + (2 \log n)^{1/2} n^{2/3} \rceil$ med stor sandsynlighed enten er større end $n^{2/3}$ eller mindre end $\frac{n^{2/3}}{2}$. Vi har endvidere, at G_{t+1} højst har ligeså mange store sammenhængskomponenter som G_t , idet to små sammenhængskomponenter ikke kan blive til én stor, når der blot tillægges en enkelt kant. Spørgsmålet er, om G_t indeholder alle store sammenhængskomponenter fra G_{t_0} i én sammenhængskomponent for $t > t_0$. I så fald har G_t en kæmpekomponent, hvilket gælder for alle G_{t^*} hvor $t^* \geq t$.

Bollobás [10] har vist, at for $t \geq t_1 = \lfloor n/2 + 2(\log n)^{1/2} n^{2/3} \rfloor > t_0 = \lceil n/2 + (2 \log n)^{1/2} n^{2/3} \rceil$ gælder der netop, at en grafproces med stor sandsynlighed er sådan, at alle store sammenhængskomponenter af G_{t_0} er indeholdt i en enkelt sammenhængskomponent af G_{t_1} . Dermed indeholder G_t for alle $t \geq t_1$ en kæmpekomponent. Sætningen, som vi angiver uden bevis, lyder som følger:

Sætning 5.1.2 En grafproces $\tilde{G} = (G_t)_0^N$ er sådan, at for ethvert $t \geq t_1 = \lfloor n/2 + 2(\log n)^{1/2} n^{2/3} \rfloor$ indeholder grafen G_t med stor sandsynlighed en unik sammenhængskomponent af størrelse mindst $n^{2/3}$, samtidig med at størrelsen af de andre sammenhængskomponenter højst er $\frac{n^{2/3}}{2}$.

Ønsker man at bestemme den præcise størrelse af kæmpekomponenten, kan man blot udregne forskellen mellem n og antallet af knuder i de små sammenhængskomponenter. Når t vokser i grafprocessen, vokser størrelsen af kæmpekomponenten også, og antallet af knuder, som ikke er i kæmpekomponenten går dermed mod nul.

5.1.2 Sammenhæng

I kapitel 5.1.1 om kæmpekomponenten af en stokastisk graf fandt vi, at $\lfloor n/2 + 2(\log n)^{1/2} n^{2/3} \rfloor$ er det antal kanter, der mindst kræves, for at G_t i en grafproces $(G_t)_0^N$ med stor sandsynlighed indeholder en kæmpekomponent. På et tidspunkt i grafprocessen vil kæmpekomponenten af G_t være blevet

så stor, at den indeholder samtlige knuder i grafen, hvorved G_t er sammenhængende. Følgende resultat giver en sandsynlighed for sammenhængen af en stokastisk graf, givet at antallet af kanter er lig $(n/2)(\log(n) + c + o(1))$, hvor c er en konstant, og $o(1)$ betegner et led gående mod 0 for $n \rightarrow \infty$.

Sætning 5.1.3 *Lad $c \in \mathbb{R}$ være fast og lad $m = (n/2)(\log(n) + c + o(1)) \in \mathbb{N}$. Da er sandsynligheden, for at grafen G_m er sammenhængende, givet ved:*

$$P(G_m \text{ sammenhængende}) \rightarrow e^{-e^{-c}}$$

Bevís: Bollobás [10] - sætning 7.3. \square

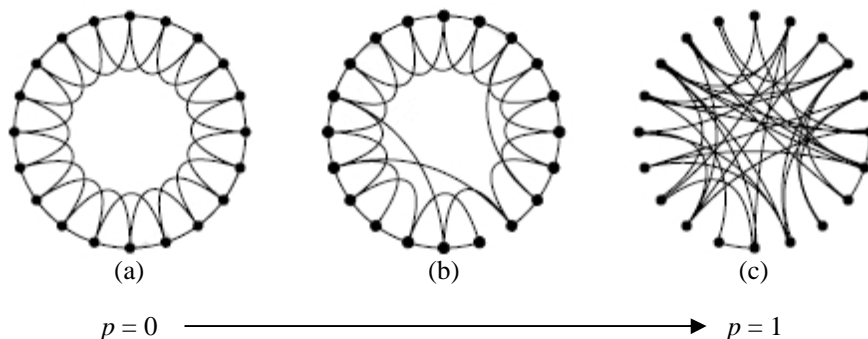
Sandsynligheden, for at G_t i en grafproces er sammenhængende, når $t = m$, går således mod 1 for $c \rightarrow \infty$. For $c = 0$ er sandsynligheden 0,3679, for $c = 1$ er sandsynligheden 0,6922, og for $c = 5$ er sandsynligheden 0,9933. Konvergensten mod 1 sker altså relativt hurtigt, når c vokser. På tilsvarende vis er sandsynligheden for sammenhæng meget lille, så snart c er negativ. En c -værdi på -1 giver således en sandsynlighed på 0,066, og $c = -2$ medfører en sandsynlighed på blot 0,0006. Vi har således en tærskelfunktion for sammenhængen af en stokastisk graf, som er $m^* = (n/2) \log(n)$.

5.2 Watts-Strogatz modellen

I 1998 startede Watts og Strogatz [12] diskussionen om modellering af virkelige netværk ved hjælp af stokastiske grafer. Virkelige netværk har ofte den egenskab, at knuderne er "klumpet sammen"; forstået på den måde, at forskellige delgrafer er forbundet med særligt mange kanter.⁹ Samtidig er diameteren typisk ret lille i virkelige netværk. Tidligere antog man ofte, at et netværk var modelleret enten ved en helt regulær graf, hvor valensen af alle knuder er ens, eller ved en klassisk stokastisk graf. En regulær graf er kendetegnet ved en høj klyngekoefficient og en stor diameter, mens en klassisk stokastisk graf er kendetegnet ved en lav klyngekoefficient og en lille diameter. Ingen af de to grafter er derfor særligt velegnede til modellering af virkelige netværk.

Watts og Strogatz foreslår i [12] en procedure til generering af en graf, som ligger et sted mellem de to ekstreme grafter: den regulære graf og den klassiske stokastiske graf. Udgangspunktet er en regulær kreds, hvor hver knude har det samme antal naboer. Figur 5.2 viser en regulær kreds med 20 knuder, hvor hver knude har fire naboer. Vi vælger en knude v i kredsen og den kant

⁹Man siger, at grafen har en høj klyngekoefficient - den formelle definition kommer på side 28.



Figur 5.2 (a) regulær kreds, (b) Watts-Strogatz graf, (c) stokastisk graf. Kilde: Figur 1, [12].

e , som forbinder v til dens nærmeste nabo med uret. Med sandsynlighed $p \in]0, 1[$ ændrer vi kanten e og forbinder den i stedet fra v til en tilfældig anden knude fra kredsen (hvis det resulterer i to ensliggende kanter, undlader vi at ændre kanten). Fremgangsmåden gentages, indtil alle knuder i kredsen er blevet valgt. Herefter vælges igen knuden v i kredsen, men nu kanten e' , som forbinder v til dens næste nabo med uret. På samme måde som før ændres kanten e' med sandsynlighed p . Af figur 5.2 ses følgende tre grafer: den oprindelige regulære kreds, resultatet af den nævnte fremgangsmåde for et givet p samt den klassiske stokastiske graf, som fremkommer, hvis $p = 1$, da alle kanter i så fald ændres uniformt. Vi vil nu se nærmere på, hvilke egenskaber Watts-Strogatz grafen har.

I stedet for at se på den traditionelle diameter af en graf, ser Watts og Strogatz på den gennemsnitlige afstand mellem knuderne. Dette mål giver en bedre idé om, hvordan hele grafen hænger sammen, da man kigger på *alle* afstande i stedet for kun den maksimale. De to størrelser hænger dog nøje sammen.¹⁰ Den gennemsnitlige afstand L af en graf med knudemængde V er defineret ved:¹¹

$$L = \frac{1}{n(n-1)} \sum_{\{u,v\} \subset V, u \neq v} d(u,v).$$

Klyngekoefficienten C beskriver hvor stor en andel af knudernes naboer, der også selv er naboer. Lad v være en knude med mindst to naboer og lad G_v være delgraf af naboer til v . Hvis knude v har $d(v)$ naboer, da er antallet

¹⁰Se evt. Bollobás [11].

¹¹Antallet af kanter i en komplet graf er lig $n(n-1)/2$, men da hver kant bliver betragtet to gange (både afstanden $d(u,v)$ og $d(v,u)$), divideres med $n(n-1)$.

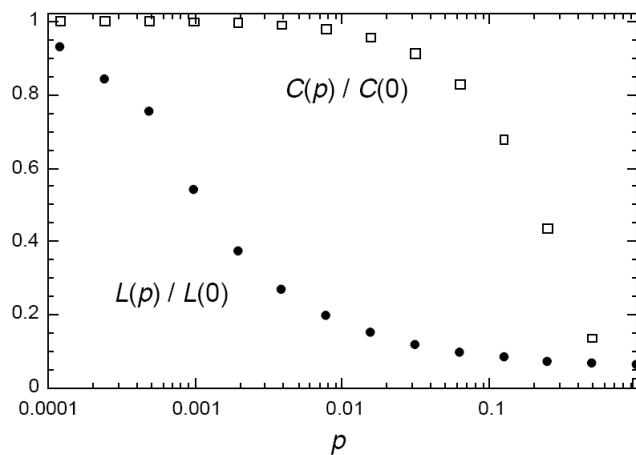
af knuder i G_v lig $d(v)$, og antallet af kanter er højst $\binom{d(v)}{2} = \frac{d(v) \cdot (d(v)-1)}{2}$. Koefficienten C_v er andelen af disse kanter, som eksisterer, når vi sletter knude v , og C er den gennemsnitlige værdi af C_v for alle knuder i netværket. Koefficienterne C_v og C er da defineret ved:

$$C_v = \frac{\text{antal kanter i } G_v}{d(v) \cdot (d(v) - 1)/2}$$

og

$$C = \frac{1}{n} \sum_{v \in V} C_v .$$

Lad $L(p)$ og $C(p)$ være henholdsvis den gennemsnitlige afstand og klyngekoefficienten for en Watts-Strogatz graf fremkommet ved ændringer af den regulære kreds med sandsynlighed p . Figur 5.3¹² afbilder funktionerne $L(p)/L(0)$ og $C(p)/C(0)$. Vi ser, at den gennemsnitlige afstand bliver kraftigt formind-



Figur 5.3 $L(p)/L(0)$ og $C(p)/C(0)$ som funktion af p . Kilde: Figur 2, [12].

sket i forhold til den oprindelige, så snart p vokser. Samtidig sker der stort set ingen ændring i klyngekoefficienten; først når p bliver stor, får det afgørende betydning for $C(p)$. Watts og Strogatz konkluderede således, at man ved at introducere en smule tilfældighed i en regulær graf kan reducere den gennemsnitlige afstand mellem knuder (og dermed også diameteren) betragteligt, samtidig med at den høje klyngekoefficient bliver bevaret.

Ved at anvende noget af teorien fra stokastiske grafer lykkedes det altså

¹²Kilde: Watts & Strogatz [12]

Watts og Strogatz at generere en graf med egenskaber, der kendetegner virkelige netværk. Imidlertid er arkitekturen af Watts-Strogatz grafen ikke helt sammenlignelig med de store virkelige netværk. Den helt afgørende forskel er, at virkelige netværk ofte indeholder store knudepunkter¹³, hvilket ikke er tilfældet for Watts-Strogatz grafen. Der er derfor siden blevet gjort meget for at finde frem til en model, der bedre passer med virkeligheden, og som stadig har de nødvendige egenskaber som Watts-Strogatz modellen. Det har bl.a. resulteret i såkaldte scale-free grafmodeller, som følgende afsnit omhandler.

5.3 Scale-free modeller

En stokastisk graf med en sandsynlighedsfunktion af typen $P(k) \sim k^{-\gamma}$, $\gamma \geq 0$ kaldes også for en scale-free graf. $P(k)$ udtrykker her sandsynligheden for, at en bestemt knude er forbundet med k naboer. Navnet 'scale-free' er opstået, da forholdet mellem sandsynlighederne $P(100)/P(10)$ er lig med forholdet $P(1000)/P(100)$, som igen er lig med $P(10000)/P(1000)$ osv.

Adskillige virkelige netværk såsom internettet, telefonnettet og infrastrukturen kan modelleres ved brug af scale-free grafmodeller.¹⁴ Der er to overordnede egenskaber ved en scale-free grafmodel, som gør, at den adskiller sig væsentligt fra både den klassiske stokastiske grafmodel og Watts-Strogatz modellen:

- 1) Antallet af knuder er ikke fastlagt på forhånd, som det var tilfældet med både den klassiske stokastiske graf og Watts-Strogatz grafen. Der er således mulighed for at tilføje nye knuder.
- 2) Knuderne forbindes enkeltvis med et vist antal eksisterende knuder under den regel, at knuder med stor valens har større sandsynlighed for at blive valgt end knuder med lille valens.

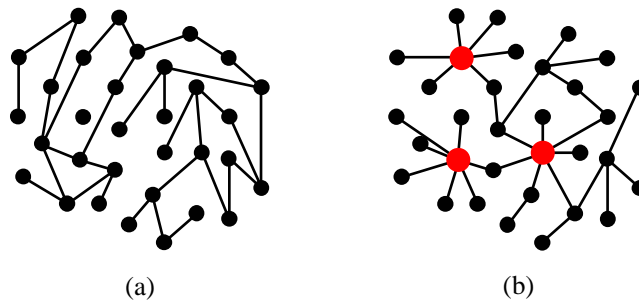
Den første egenskab gør, at grafmodellen hele tiden kan udvides med nye knuder. Dette er i høj grad medvirkende til, at modellen er anvendelig i praksis. Hovedparten af virkelige netværk er ikke stationære, men udvikler sig hele tiden med nye knuder; tænk f.eks. på internettet, hvor nye hjemmesider kommer til og på infrastrukturen, hvor der bl.a. bliver lavet nye motorveje og motorvejskryds. Hvis antallet af knuder på forhånd var fastlagt, var der således ikke mulighed for at modellere disse scenarier.

¹³Det vil sige knuder, som har forholdsmeæssig stor valens.

¹⁴F.eks. Faloutsos et al. [23].

Den anden egenskab betyder, at alle eksisterende knuder ikke har lige stor sandsynlighed for at blive valgt som nabo til en ny knude. Sandsynligheden afhænger af, hvor stor valens de eksisterende knuder har i forvejen - og dermed, hvor "vigtige" de er. Hvis en eksisterende knude har stor valens, er der større sandsynlighed for, at den nye knude bliver forbundet til netop denne knude, end hvis valensen var lille. Sagt med andre ord, bliver vigtige knuder vigtigere. På den måde skaber man forskellige store knudepunkter, som det f.eks. også er tilfældet med infrastrukturen. Hovedbanegården i København er et eksempel på en knude med meget stor valens, som derfor er vigtig, for at resten af netværket kan fungere optimalt. Internettet er et andet eksempel, hvor nye ukendte hjemmesider typisk har links til mere kendte sider som f.eks. Google, MSN osv., så også her bliver vigtige knuder vigtigere.

Figur 5.4 skitserer forskellen mellem en alm. stokastisk graf og en scale-free stokastisk graf. Man ser umiddelbart, at strukturen i en scale-free graf er meget forskellig fra den i en klassisk stokastisk graf. Knudepunkterne er illustreret ved de røde cirkler og optræder således kun i den scale-free graf. Helt generelt indeholder en klassisk stokastisk graf typisk ikke knuder med stor valens.



Figur 5.4 (a) klassisk stokastisk graf, (b) scale-free graf.

5.3.1 Barabási og Alberts model

Barabási og Albert [13] var blandt de første til at beskrive en fremgangsmåde, som resulterer i en scale-free graf.

Start med et lille antal (m_0) knuder. Tilføj en knude med $m' \leq m_0$ kanter, som forbinder den nye knude med m' af de eksisterende knuder. Antag at sandsynligheden $P(d(i))$ for at en ny knude er

forbundet med en eksisterende knude i afhænger af valensen af denne knude, sådan at:

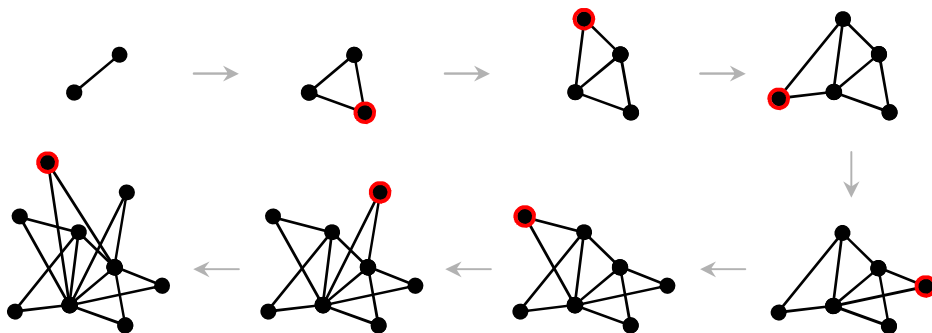
$$P(d(i)) = \frac{d(i)}{\sum_{j=1}^n d(j)}.$$

Når t knuder er blevet tilføjet på denne måde, har vi en graf med $m_0 + t$ knuder og $m = m_0 + t$ kanter. Resultatet af denne fremgangsmåde er en graf, som besidder de to vigtige egenskaber nævnt på side 30. Forudsætningerne, for at vi nu har en brugbar graf til at modellere et virkeligt netværk, er således opfyldt. Grafen udvikler sig til at være scale-free med en sandsynlighedsfunktion $P(k) \sim k^{-\gamma}$, for at en knude har k naboer. Den gennemsnitlige sammenhæng for en sådan graf er lig med $2m/n$, så hvis grafen f.eks. har 10 knuder og 15 kanter, er den gennemsnitlige sammenhæng lig med 3. Barabási og Albert har i [13] fundet, at $\gamma = 2,9 \pm 0,1$.¹⁵ For $m_0 = m = 2$ illustrerer figur 5.5 Barabási og Alberts procedure for generering af en scale-free graf. Vi starter med to knuder, hvorefter en ny knude med to tilhørende kanter kommer til. Denne forbindes til de knuder med størst valens (i dette tilfælde de to eneste knuder). Grafen har således på nuværende tidspunkt tre knuder. Vi tilføjer igen en ny knude med to tilhørende kanter. Da alle knuder på dette tidspunkt har samme valens, er det helt tilfældigt, hvilke naboer den nye knude får. Vores graf har nu fire knuder, hvoraf de to har valens 3 og de andre valens 2. Når vi tilføjer endnu en knude, vil sandsynligheden for, at denne knude forbindes med en af knuderne med valens 3 være lig $2 \cdot \frac{3}{2+2+3+3} = \frac{3}{5}$. I figur 5.5 illustreres tilfældet, hvor den nye knude forbindes med én af knuderne med valens 3 og én af knuderne med valens 2. Det resulterer i en graf med fem knuder; en med valens 4, to med valens 3 og to med valens 2. Hver gang proceduren gentages, forbindes en ny knude overvejende sandsynligt med de eksisterende knuder med størst valens. På den måde bliver grafen heterogen, og der opstår knudepunkter som afbildet i figur 5.5. Disse kendetegner bliver tydeligere, efterhånden som flere knuder tilføjes.

Barabási og Alberts scale-free grafmodel fra 1999 (også kaldet BA-modellen) er typisk den, der henvises til under emnet scale-free grafer, og det er derfor også den model, vi beskæftiger os med i kapitel 6. Bollobás og Riordan påpeger i [11] imidlertid en række forhold, som gør, at modellen matematisk set ikke er helt veldefineret.

Det første problem er starten; hvis der ikke i forvejen eksisterer en graf med et

¹⁵I de fleste artikler, sættes værdien af γ dog lig 3.



Figur 5.5 Tilføjelse af rød knude iflg. Barabási og Alberts procedure. Stor forskel på valensen af knuderne i den sidste graf.

antal kanter, kan vi ikke danne en sådan ud fra den nævnte fremgangsmåde. Det er f.eks. ikke muligt at betragte forholdet $\frac{d(i)}{\sum_{j=1}^n d(j)}$, $i \in \{1, \dots, m_0\}$, når valenserne er lig 0. Det er således en nødvendighed, at vi på forhånd er givet en graf. Selvom dette er tilfældet, kan det være af afgørende betydning for den resulterende graf, at begyndelsesgraphen har nogle bestemte egenskaber. Ser vi f.eks. på tilfældet $m = 1$, hvor hver ny knude kun bliver forbundet med én kant; da giver Barabási og Alberts fremgangsmåde et træ, men *kun* hvis begyndelsesgraphen også er et træ. Hvis begyndelsesgraphen ikke er sammenhængende, vil fremgangsmåden aldrig resultere i en sammenhængende graf, da hver ny knude kun bliver tilknyttet én af sammenhængskomponenterne.

Det andet problem har at gøre med måden, hvorpå nye knuder bliver forbundet med den eksisterende graf. Problemet opstår, når vi på tidspunkt $t + 1$ tilføjer en knude med $m \geq 2$ kanter til en mængde N_{t+1} bestående af m eksisterende knuder. For helt at kunne beskrive modellen skal vi kunne angive fordelingen for N_{t+1} eller med andre ord sandsynligheden for udfaldet af en bestemt mængde N_{t+1} med m knuder ud af de $\binom{t}{m}$ forskellige muligheder. Fordelingen har $\binom{t}{m} - 1$ frihedsgrader, da de $\binom{t}{m}$ sandsynligheder skal summe til 1. Der er imidlertid kun specificeret t forskellige marginale sandsynligheder (en for hver gang en ny knude tilføjes), for at en bestemt knude tilhører N_{t+1} , så fordelingen er ikke unikt bestemt.

For at undgå disse formelle problemer, definerede Bollobás og Riordan i [14] en alternativ scale-free grafmodel, den såkaldte LCD-model.

5.3.2 LCD-modellen

Dette afsnit beskriver en matematisk præcis fremgangsmåde (introduceret af Bollobás og Riordan i [14]) til generering af en scale-free stokastisk graf. Grafmodellen, som kaldes LCD-modellen¹⁶, tillader både løkker og parallelle kanter. Det rimelige heri skal ses i lyset af, at f.eks. hjemmesider sagtens kan have et link til egen side eller flere links til en anden side. På samme måde vil der i et trafiknetværk ofte være flere alternativer til at komme fra et sted til et andet - det kan f.eks. være med bus eller med S-tog. Der eksisterer således løkker såvel som parallelle kanter i virkelige netværk.

Vi betragter først situationen, hvor hver ny knude bliver tilknyttet grafen via én ny kant, dvs. $m = 1$. Betragt en følge af knuder v_1, v_2, \dots og lad $(G_1^t)_{t \geq 0}$ være en grafproces, som er defineret på følgende måde:

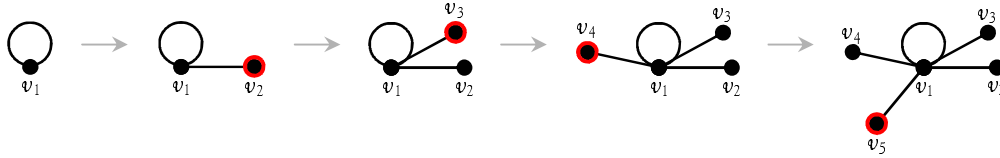
- Start med G_1^0 , som er den tomme graf uden knuder og kanter eller start med G_1^1 , som er grafen bestående af en knude og en løkke.
- Givet grafen G_1^{t-1} får vi G_1^t ved at lægge en knude v_t til G_1^{t-1} . Den nye knude v_t forbindes med en knude v_i , hvor i vælges blandt de t mulige med sandsynligheder:

$$P(i = s) = \begin{cases} d_{G_1^{t-1}}(v_s)/(2t - 1) & 1 \leq s \leq t - 1; \\ 1/(2t - 1) & s = t. \end{cases}$$

Ovenstående giver en graf G_1^t med t knuder og t kanter. Starten er entydigt bestemt (modsat Barabási og Alberts grafmodel beskrevet i det foregående afsnit). Samtidig forbindes nye knuder fortrinsvis med "vigtige" knuder, så det scale-free kendetegn er stadig opfyldt. Det er tilfældet, idet v_t forbindes til en knude v_i valgt med en sandsynlighed, der er proportional med dens valens på det pågældende tidspunkt. Den nye kant fra v_t bidrager på forhånd med én til valensen af v_t , hvilket fremgår af, at $P(i = t) = \frac{1}{2t-1}$. Figur 5.6 illustrerer, hvordan grafen G_1^5 kan se ud (eksemplet er lavet ud fra de mest sandsynlige udfald). Udgangspunktet er en graf med en knude v_1 og en løkke. Herefter tilføjes en knude v_2 , hvor sandsynligheden for en løkke er $\frac{1}{2 \cdot 2 - 1} = \frac{1}{3}$ og sandsynligheden for forbindelse til v_1 er $\frac{2}{2 \cdot 2 - 1} = \frac{2}{3}$. Vi forbinder v_2 til v_1 og tilføjer en ny knude v_3 . Sandsynligheden for en løkke er $\frac{1}{2 \cdot 3 - 1} = \frac{1}{5}$, sandsynligheden for forbindelse til v_2 er lig $\frac{1}{5}$, og sandsynligheden for forbindelse til v_1 er $\frac{3}{5}$. Vi forbinder v_3 til v_1 og fortsætter på samme måde indtil $t = 5$. Resultatet er ikke overraskende en graf med et knudepunkt v_1 . Hvis vi i stedet for

¹⁶Det bliver senere klart, hvorfor modellen kaldes LCD-modellen.

at vælge de mest sandsynlige udfald havde brugt et program til at generere grafen, ville vi muligvis få et lidt andet resultat; strukturen ville dog stadig være præget af, at v_1 fungerer som knudepunkt.



Figur 5.6 *Illustration af G_1^5 .*

Hvis hver ny knude i stedet tilknyttes grafen via flere kanter (dvs. $m > 1$), tilføjes de m kanter en ad gangen efter samme regel som før. Samtidig opdateres valenserne for de forskellige knuder undervejs i processen. For knuden v_t er valensen inden tilføjelsen af den første kant lig med 1, inden tilføjelsen af den næste kant lig med $d(v_t) + 1$ og så videre, indtil alle m kanter er tilføjet. På denne måde fremkommer grafen G_m^t med t knuder og mt kanter. Vi kan også definere processen $(G_m^t)_{t \geq 0}$ ved hjælp af G_1^{mt} (dvs. grafen fremkommet af processen for $m = 1$, hvor antallet af knuder er lig mt), hvilket er en stor fordel, da G_1^{mt} er en meget simplere graf, som dermed er lettere at arbejde med. Fremgangsmåden (illustreret i figur 5.7) er følgende:

- Lad $m > 1$. Kør processen $(G_1^t)_{t \geq 0}$ på knuderne $v'_1, v'_2, \dots, v'_{mt}$ og få grafen G_1^{mt} .
- Identificér knuderne v'_1, \dots, v'_m i G_1^{mt} ved at erstatte de m knuder med én knude v_1 og lade alle kanter incidente på v'_1, \dots, v'_m være incidente på v_1 . På tilsvarende vis identificeres v'_{m+1}, \dots, v'_{2m} til v_2 . Sådan fortsættes, indtil vi når $v'_{m(t-1)+1}, \dots, v'_{mt}$, som identificeres til v_t .
- Grafen bestående af v_1, \dots, v_t udgør nu G_m^t .

Vi har hermed beskrevet en matematisk veldefineret procedure til generering af en scale-free graf. Skal vi foretage en præcis matematisk analyse af robustheden, vil LCD-modellen således være at foretrække, men da vi i stedet bruger resultater fra simulation, er Barabási og Alberts model i princippet ligeså god. Selvom vi i dette projekt udelukkende koncentrerer os om ikke-orienterede grafer, kan modellen let modificeres til at tage højde for orienteringer; en naturlig og simpel metode vil være for $i > j$ at orientere enhver kant (v_i, v_j) fra v_i til v_j .¹⁷

¹⁷Bollobás og Riordan [11].



Figur 5.7 $G_1^{2.5} = G_1^{10}$ bliver til G_2^5 ved identifikation af (v'_1, v'_2) til v_1 , (v'_3, v'_4) til v_2 , (v'_5, v'_6) til v_3 , (v'_7, v'_8) til v_4 og (v'_9, v'_{10}) til v_5 .

En af de vigtigste egenskaber og fordele ved denne LCD-model frem for Barabási og Alberts model er, at den har en ækvivalent statistisk (dvs. ikke-rekursiv) formulering. Graferne G_1^t og G_m^t kan på ethvert tidspunkt $t = n$ således beskrives ved hjælp af en statistisk fordeling. I følgende afsnit beskriver vi for fuldstændighedens skyld metoden til dette, introduceret af Bollobás og Riordan i [14], som samtidig forklarer navnet ”LCD-modellen”.

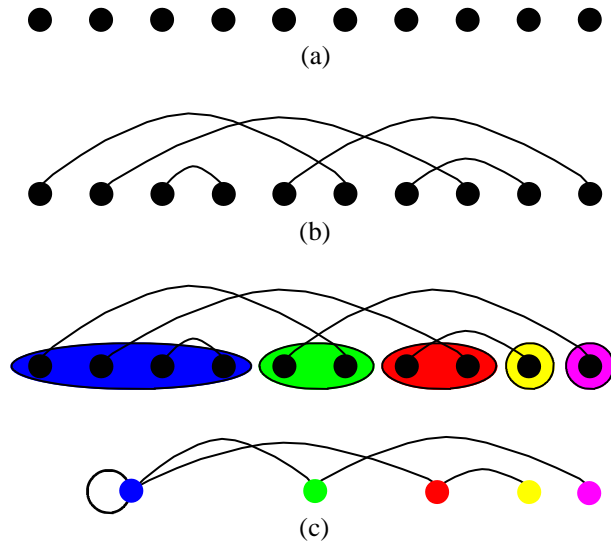
5.3.3 n -parring

For at simplificere notationen, sætter vi nu $i = v_i$, så en graf G_1^n består af knuderne $1, \dots, n$. Betragt mængden $\{1, \dots, 2n\}$ af knuder og lad de $2n$ knuder være placeret på en usynlig x -akse (se figur 5.8(a)). Generér en vilkårlig n -parring¹⁸ L af knuderne, således at knuderne på akserne forbindes parvist (figur 5.8(b)). Dette kaldes også et *Linearized Chord Diagram* - forkortet LCD. Gå nu fra venstre mod højre på x -aksen og identificér alle venstre endeknuder af LCD'en til og med den første højre endeknude; sådan fortsættes, indtil vi når endeknuden længst til højre på x -aksen. Af figur 5.8(c) ses den resulterende graf, som vi betegner $\phi(L)$.

Vi kan også generere L på følgende måde (se figur 5.9 for illustration):

- Vælg en tilfældig LCD L' med $n - 1$ kanter.
- Tilføj en kant e , hvor højre endeknude ligger til højre for alle knuder i L' , og venstre endeknude ligger i en af de $2n - 1$ mulige placeringer (imellem de eksisterende knuder), hvor hver placering vælges med lige stor sandsynlighed.

¹⁸En n -parring er en mængde af n kanter, der ikke har endeknuder til fælles.



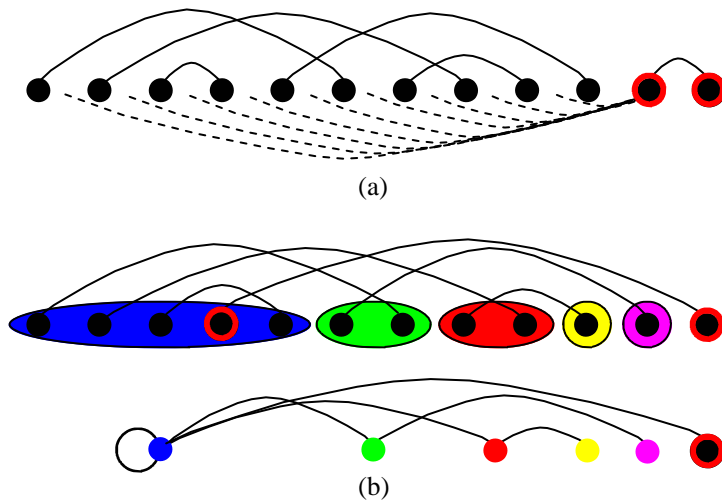
Figur 5.8 (a) $2n$ knuder på række ($n = 5$), (b) en n -parring L , (c) øverste graf: illustration af knuder, som skal identificeres; nederste graf: $\phi(L)$.

Dette svarer til at lægge en ny knude til $\phi(L')$ og forbinde den med en anden knude med sandsynligheder proportionale med valensen af knuderne. Dette svarer igen præcis til beskrivelsen af G_1^n , så fordelingen af G_1^n er med andre ord den samme som fordelingen af $\phi(L)$. Så hvis L vælges uniformt mellem alle de mulige LCD'er med n kanter, er fordelingen af $\phi(L)$ den samme som fordelingen af G_1^n . Det er således indlysende, hvorfor modellen beskrevet i kapitel 5.3.2 kaldes LCD-modellen.

5.3.4 Klemm-Eguíluz modellen

Vi har indtil nu set på tre forskellige modeller, som hver især besidder relevante egenskaber med henblik på virkelige netværk. Watts-Strogatz modellen udmærker sig ved en høj klyngekoefficient, men mangler de to egenskaber nævnt på side 30, som er særdeles vigtige i virkelige netværk. Disse to egenskaber er både BA- og LCD-modellen i besiddelse af, men til gengæld har ingen af dem en høj klyngekoefficient.

For at løse dette problem designede Klemm og Eguíluz i [17] en scale-free graf med en høj klyngekoefficient (fremover betegnet KE-graf). I deres algoritme gør Klemm og Eguíluz brug af en binær tilstandsvariabel, som enten er aktiv eller inaktiv. Algoritmen, der ligesom BA-modellen danner en scale-free



Figur 5.9 (a) en LCD med 5 kanter tilføjes en ny kant med to (røde) endeknuder; den ene knude placeres yderst til højre - den anden placeres via én af de stiplede linier, (b) venstre endeknude placeres vilkårligt, men med størst sandsynlighed i det blå felt i den øverste graf; med samme sandsynlighed forbindes den nye knude i den nederste graf med den blå knude.

graf med gennemsnitlig sammenhæng $2m/n$, er som følger:

- Lad en komplet graf med m aktive knuder være givet.
- Tilføj en ny knude med m kanter til den eksisterende graf.
- For enhver af de m kanter, besluttes det med sandsynlighed p , om kanten forbindes til en af de aktive knuder. Det betyder med andre ord, at med sandsynlighed $1 - p$ bliver kanten forbundet med en af de inaktive knuder.
- Hvis det blev besluttet at forbinde kanten til en inaktiv knude, vælges knuden med sandsynligheder proportionale med valensen af de inaktive knuder - dvs.

$$P(d(i)) = \frac{d(i)}{\sum_{j=1}^n d(j)}$$

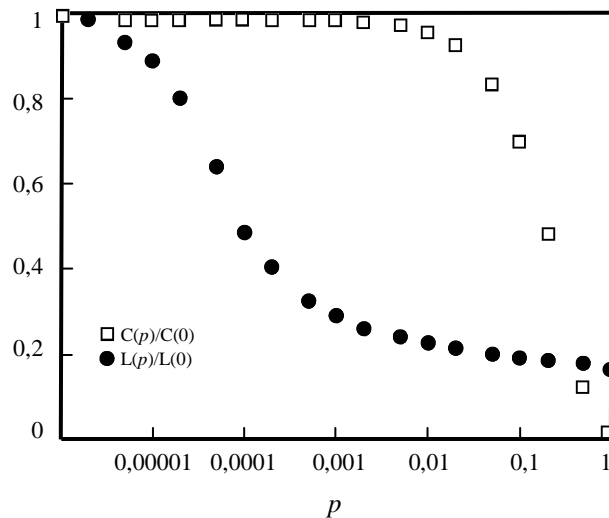
ligesom i BA-modellen.

- Efter tilføjelse af en knude, deaktiveres en af de andre knuder. Sandsynligheden, for at en knude i deaktiveres, er givet ved:

$$P(d(i)) = \frac{d(i)^{-1}}{\sum_{j=1}^n (d(j)^{-1})}.$$

Sidste trin betyder, at knuder med mindst valens har størst sandsynlighed for at blive deaktiveret; med andre ord forbliver de ”vigtigste” knuder aktive.

Værdien af p er afgørende for KE-grafens udseende og egenskaber (se figur 5.10¹⁹). For $p = 0$ resulterer KE-modellen i en graf med en høj klyngekoeffi-



Figur 5.10 $L(p)/L(0)$ og $C(p)/C(0)$ som funktion af p . Kilde: Figur 1, [17].

cient, men også en stor gennemsnitlig afstand. Omvendt medfører $p = 1$, at KE-modellen er identisk med BA-modellen, dvs. en graf med lille klyngekoefficient og kort afstand.²⁰ Så snart p antager en værdi større end nul, falder den gennemsnitlige afstand drastisk, samtidig med at den høje klyngekoefficient bevares. Hvis $p \neq 0$ og $p \ll 1$ er de ønskede betingelser om en scale-free graf med høj klyngekoefficient således opfyldt.²¹

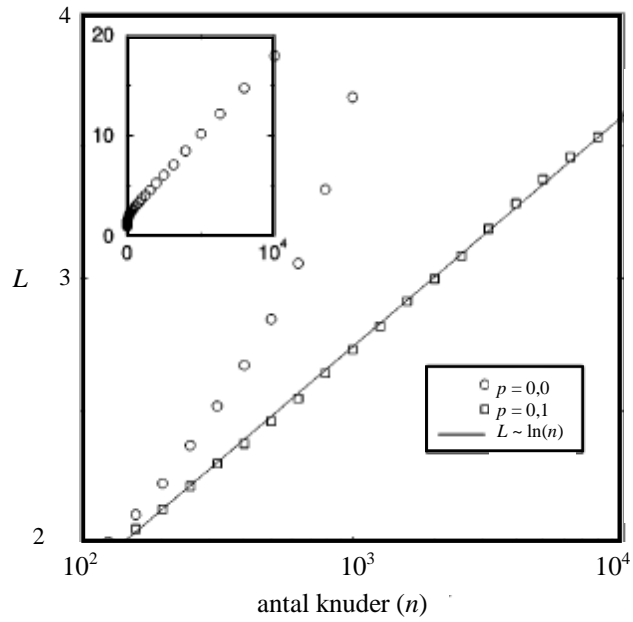
Antallet af knuder i grafen har stor betydning for indflydelsen af p . Af figur 5.11(a) fremgår det, at den gennemsnitlige afstand mellem knuderne vokser lineært med antallet af knuder for $p = 0$. Sætter vi i stedet $p = 0,1$, vokser afstanden kun logaritmisk med antallet af knuder. Tilsvarende kan vi af figur 5.11(b) aflæse klyngekoefficienten for grafer af forskellig størrelse med hhv. $p = 0$, $p = 0,1$ og $p = 1$. Klyngekoefficienten bliver gradvist mindre, når p vokser fra 0 til 0,1, og antallet af knuder samtidig bliver større. Klyngekoeffi-

¹⁹Kilde: Klemm og Eguíluz [17].

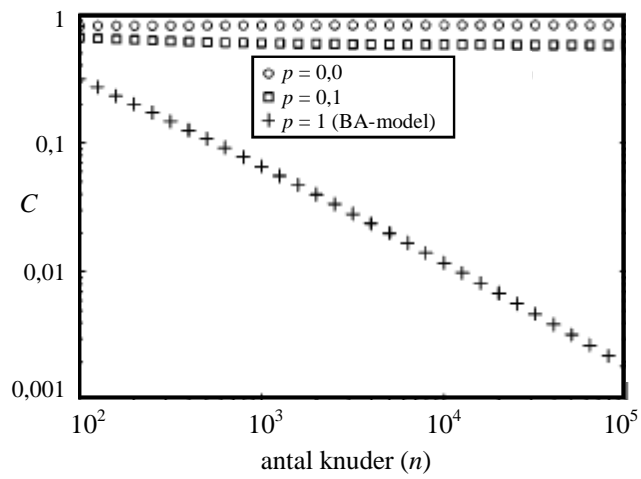
²⁰Bemærk at udviklingen af klyngekoefficienten og den gennemsnitlige afstand som funktion af p er tæt på identisk med den for Watts-Strogatz modellen (se figur 5.3).

²¹Vi sætter $p = 0,1$, når vi i kapitel 6 analyserer robustheden af KE-grafer.

cienten for $p = 0,1$ forbliver dog høj sammenlignet med tilfældet $p = 1$, hvor koefficienten falder drastisk i takt med et øget antal knuder. Vi kan heraf konkludere, at jo større grafen er, jo vigtigere er værdien af p med henblik på en analyse af grafens egenskaber.



(a)



(b)

Figur 5.11 (a) gennemsnitlig afstand som funktion af antal knuder for bestemte værdier af p , (b) klyngekoefficienten som funktion af antal knuder for bestemte værdier af p . Kilde: Figur 2 og 4a, [17].

Kapitel 6

Robusthed af scale-free grafer

Vi har indtil nu set på forskellige grafter, som hver især har både fordele og ulemper. Til modellering af store netværk har vi sluttet, at de scale-free grafer er mest velegnede. Men hvad sker der med et netværk, hvis der opstår en fejl, eller hvis det bliver udsat for et direkte angreb? Fungerer resten af netværket stadig, og er der forskel på effektiviteten af de forskellige netværkstyper? Disse spørgsmål vil blive forsøgt besvaret i dette kapitel. Vi beskæftiger os med de to scale-free grafmodeller af hhv. Barabási og Albert samt Klemm og Eguíluz, og drager undervejs sammenligninger med den klassiske stokastiske graf af Erdős og Rényi. Resultaterne er baseret på simulationer foretaget af Crucitti et al. i [16], hvor effektiviteten af et netværk analyseres efter henholdsvis fejl og angreb.¹

6.1 Effektiviteten af et netværk

I kapitel 3 blev begrebet effektivitet kort introduceret. Vi vil nu præcisere, hvad der forstås ved effektiviteten af et netværk og minder i den forbindelse om definitionerne for den gennemsnitlige afstand L og klyngekoefficienten C fra kapitel 5.2:

$$L = \frac{1}{n(n-1)} \sum_{\{u,v\} \subset V, u \neq v} d(u,v)$$
$$C = \frac{1}{n} \sum_{v \in V} C_v, \quad \text{hvor} \quad C_v = \frac{\text{antal kanter i } G_v}{d(v) \cdot (d(v) - 1)/2}.$$

¹Jeg har forsøgt at få adgang til simulations-software ved bl.a. at kontakte Dekker, der i [1] gør meget reklame for sit eget software CAVALIER, men han kunne desværre ikke hjælpe.

Definitionen af L giver kun mening, såfremt grafen er sammenhængende. I modsat fald vil afstanden $d(u, v) = \infty$ for mindst én kombination af $(u, v) \subset V$. Det medfører, at $L = \infty$, hvilket åbenlyst ikke giver mening. Når vi i dette kapitel studerer effekten af et nedbrud i netværket, vil det resulterende netværk ofte være usammenhængende, hvormed L bliver uegnet som robusthedsmål.

I stedet for at se på L og C undersøger Latora og Marchiori i [15], hvor *effektivt* strømmen flyder i netværket efter et nedbrud. Netværket bliver betragtet både *globalt* og *lokalt*; det vil sige, at man ud over at undersøge netværket som en helhed også kigger på, hvordan enkelte udsnit fungerer. Antag at al strømmen i netværket foregår langs kanterne, og at effektiviteten af en kant $e = (u, v)$ afhænger af afstanden mellem de to knuder u og v på den måde, at en større afstand medfører en dårligere effektivitet. Lader vi ε_{uv} betegne effektiviteten af kanten fra u til v , antager vi, at sammenhængen mellem ε_{uv} og afstanden $d(u, v)$ kan skrives:

$$\varepsilon_{uv} = \frac{1}{d(u, v)}.$$

Med denne definition undgår vi problemet med $L = \infty$, når grafen er usammenhængende. Hvis der ikke eksisterer en sti fra u til v , er $d(u, v) = \infty$ pr. definition, men så er effektiviteten $\varepsilon_{uv} = \frac{1}{\infty} = 0$ i fin overensstemmelse med den inverse sammenhæng. Vi har dermed ikke længere begrænset os til kun at kigge på sammenhængende grafer, hvilket er signifikant i forbindelse med en analyse af et nedbruds konsekvenser. Undersøger vi ε_{uv} , $\forall u, v \in V$ og tager gennemsnittet, får vi den globale effektivitet \mathcal{E}_{glob} af en graf G - dvs. effektiviteten af hele G . Vi har således:

$$\begin{aligned} \mathcal{E}_{glob}(G) &= \frac{1}{n(n-1)} \sum_{\{u,v\} \subset V, u \neq v} \varepsilon_{uv} \\ &= \frac{1}{n(n-1)} \sum_{\{u,v\} \subset V, u \neq v} \frac{1}{d(u, v)}. \end{aligned}$$

Det er værd at bemærke, at $0 \leq \mathcal{E}_{glob}(G) \leq 1$. Den størst mulige værdi for $\sum \frac{1}{d(u,v)}$ er lig $n(n-1)$ og indtræffer for den komplette graf, hvor $d(u, v) = 1$, $\forall u, v \in V, u \neq v$. Det medfører, at $\mathcal{E}_{glob}(G) \leq 1$, og da alle afstande er ikke-negative, har vi $\mathcal{E}_{glob}(G) \geq 0$, hvor lighedstegnet kun gælder, hvis antallet af kanter i G er lig med nul.

Den lokale effektivitet \mathcal{E}_{loc} defineres som den gennemsnitlige effektivitet i

alle delgrafer $G_v, v \in V$, hvor G_v er delgrafen af naboer til v . Da det størst mulige antal kanter i G_v er $d(v) \cdot (d(v) - 1) / 2$, og hver kant betragtes to gange, kan vi udregne effektiviteten $\mathcal{E}(G_v)$ af en delgraf G_v ved:²

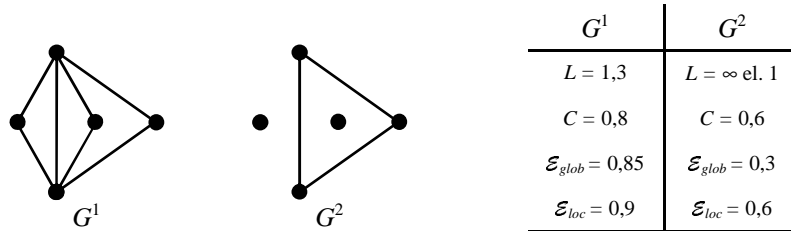
$$\mathcal{E}(G_v) = \frac{1}{d(v)(d(v) - 1)} \sum_{\{s,t\} \subset V_v, s \neq t} \frac{1}{d(s,t)}.$$

Udregner vi $\mathcal{E}(G_v)$ for alle knuder i grafen og tager gennemsnittet, får vi den lokale effektivitet af G . Det vil sige:

$$\mathcal{E}_{loc}(G) = \frac{1}{n} \sum_{v \in V} \mathcal{E}(G_v).$$

Som tilfældet var for $\mathcal{E}_{glob}(G)$, gælder der $0 \leq \mathcal{E}_{loc}(G) \leq 1$, hvor $\mathcal{E}_{loc}(G) = 1$, hvis G er komplet, og $\mathcal{E}_{loc}(G) = 0$, hvis antallet af kanter i G er nul.

Tænker vi på den gennemsnitlige afstand L og klyngekoefficienten C , hører \mathcal{E}_{glob} i en vis forstand sammen med L , mens \mathcal{E}_{loc} hører sammen med C . For at illustrere forskellene mellem de forskellige begreber betragtes et eksempel med to grafer G^1 og G^2 , som hver har fem knuder, men et forskelligt antal kanter (illustreret i figur 6.1).



Figur 6.1 To grafer med 5 knuder, men forskelligt antal kanter. Nøgletal i tabellen til højre.

Vi undersøger nu robustheden af de to grafer ved at beregne L og C samt \mathcal{E}_{glob} og \mathcal{E}_{loc} . For G^1 fås følgende:

$$\begin{aligned} \circ L &= \frac{1}{n(n-1)} \sum_{\{u,v\} \subset V, u \neq v} d(u,v) \\ &= \frac{1}{20} \left(2(7 \cdot 1 + 3 \cdot 2) \right) \\ &= \frac{13}{10} \end{aligned}$$

² V_v betegner knudemængden hørende til grafen G_v .

$$\begin{aligned}
\circ C &= \frac{1}{n} \sum_{v \in V} C_v \\
&= \frac{1}{5} \left(\frac{1}{2(2-1)/2} + \frac{3}{4(4-1)/2} + \frac{1}{2(2-1)/2} + \frac{1}{2(2-1)/2} + \frac{3}{4(4-1)/2} \right) \\
&= \frac{1}{5} \left(1 + \frac{3}{6} + 1 + \frac{3}{6} + 1 \right) \\
&= \frac{4}{5} \\
\circ \mathcal{E}_{glob} &= \frac{1}{n(n-1)} \sum_{\{u,v\} \subset V, u \neq v} \frac{1}{d(u,v)} \\
&= \frac{1}{5(5-1)} \left(2 \left(\frac{1}{1} + \frac{1}{1} + \frac{1}{1} + \frac{1}{1} + \frac{1}{1} + \frac{1}{1} + \frac{1}{1} + \frac{1}{2} + \frac{1}{2} + \frac{1}{2} \right) \right) \\
&= \frac{17}{20} \\
\circ \mathcal{E}_{loc} &= \frac{1}{n} \sum_{v \in V} \mathcal{E}(G_v) \\
&= \frac{1}{5} \left(\frac{1}{2(2-1)} \left(2 \cdot \frac{1}{1} \right) \cdot 3 + \frac{1}{4(4-1)} \left(2 \cdot \frac{1}{1} + 2 \cdot \frac{1}{1} + 2 \cdot \frac{1}{1} + 2 \cdot \frac{1}{2} + 2 \cdot \frac{1}{2} + 2 \cdot \frac{1}{2} \right) \cdot 2 \right) \\
&= \frac{1}{5} \left(\frac{1}{2} \cdot 6 + \frac{1}{12} \cdot 18 \right) \\
&= \frac{9}{10}
\end{aligned}$$

De tilsvarende resultater for G^2 er:

$$\begin{aligned}
\circ L &= \frac{1}{n(n-1)} \sum_{\{u,v\} \subset V, u \neq v} d(u,v) \\
&= \infty
\end{aligned}$$

For at undgå $L = \infty$ kan vi begrænse os til at kigge på den gennemsnitlige afstand i den sammenhængende del af grafen. I så fald får vi:

$$\begin{aligned}
L' &= \frac{1}{3(3-1)} \left(2(1 + 1 + 1) \right) \\
&= 1 \\
\circ C &= \frac{1}{n} \sum_{v \in V} C_v \\
&= \frac{1}{5} \left(\frac{1}{2(2-1)/2} + \frac{1}{2(2-1)/2} + \frac{1}{2(2-1)/2} \right) \\
&= \frac{3}{5} \\
\circ \mathcal{E}_{glob} &= \frac{1}{n(n-1)} \sum_{\{u,v\} \subset V, u \neq v} \frac{1}{d(u,v)}
\end{aligned}$$

$$\begin{aligned}
&= \frac{1}{5(5-1)} \left(2 \left(\frac{1}{1} + \frac{1}{1} + \frac{1}{1} \right) \right) \\
&= \frac{3}{10} \\
\circ \mathcal{E}_{loc} &= \frac{1}{n} \sum_{v \in V} \mathcal{E}(G_v) \\
&= \frac{1}{5} \left(\frac{1}{2(2-1)} \left(2 \cdot \frac{1}{1} \right) \cdot 3 \right) \\
&= \frac{3}{5}
\end{aligned}$$

Det ses tydeligt af figur 6.1, at G^1 er mere robust end G^2 . Denne observation bekræftes af udregningerne for den globale effektivitet, idet \mathcal{E}_{glob} for G^1 er væsentligt større end \mathcal{E}_{glob} for G^2 . G^1 fungerer således kun 15% dårligere end den komplette graf med fem knuder, hvorimod G^2 er hele 70% dårligere. Den gennemsnitlige afstand L er ikke anvendelig som robusthedsmål i dette tilfælde, da G^2 er usammenhængende, resulterende i $L = \infty$. I den alternative udregning for L i G^2 , hvor vi kun beskæftigede os med den sammenhængende del af grafen, fik vi $L = 1$. Det er mindre end L for G^1 , hvilket er i klar modstrid med, at G^1 er den mest robuste graf. Den globale effektivitet er dermed et bedre mål for robustheden af en graf end den gennemsnitlige afstand.

I visse tilfælde kan $1/L$ være en god approksimation for \mathcal{E}_{glob} . $1/L$ måler effektiviteten, hvis netværket er sekventielt (det vil sige, at der kun er én enhed i omløb ad gangen i hele netværket). \mathcal{E}_{glob} måler derimod effektiviteten i et netværk, hvor alle knuderne ekspederer enhederne samtidigt,³ hvilket f.eks. er tilfældet i vejnettet og internettet. Hvis alle afstandene i et netværk er næsten ens, gør det ikke den store forskel, hvorvidt man kun betragter en enkelt enhed eller flere enheder samtidig. Det medfører, at $1/L$ i det tilfælde vil være en rimelig approksimation for effektiviteten.

Lader vi $1/L$ være et mål for effektiviteten, opstår der imidlertid igen problemer med divergensen af L . Som tidligere nævnt medfører en usammenhængende graf, at $L = \infty$, hvilket medfører, at $1/L = 0$. Tilsvarende betyder en enkelt dårlig forbindelse, at $1/L$ vil falde betragteligt,⁴ mens den reelle situation i et stort netværk sandsynligvis vil være så godt som uændret af blot én dårlig forbindelse. I begge tilfælde giver $1/L$ således et misvisende

³Latora & Marchiori [18].

⁴Latora et al. gør i [18] opmærksom på, at L øges med ca. $\frac{1}{\varepsilon(n+1)}$ for hver ny knude, mens \mathcal{E}_{glob} falder med ca. $\frac{2}{n+1}$. Hvis den nye knude er forbundet dårligt til resten af netværket (dvs. ε lille), medfører det, at L bliver stor; dvs. $1/L$ falder. Til gengæld sker der ikke meget med \mathcal{E}_{glob} i et stort netværk.

resultat, da effektiviteten i et helt netværk åbenlyst ikke er nul (eller tæt på nul) pga. en enkelt manglende (eller dårlig) forbindelse. Disse problemer opstår ikke ved \mathcal{E}_{glob} som effektivitetsmål, hvilket er grunden til, at vi fremover foretrækker \mathcal{E}_{glob} i stedet for $1/L$ til analyse af robustheden. Forskellen på de to mål bliver igen belyst i forbindelse med analysen i kapitel 6.2 og 6.3.

Betragt nu den lokale effektivitet \mathcal{E}_{loc} og klyngekoefficienten C for hhv. G^1 og G^2 . Vi udregnede \mathcal{E}_{loc} til at være $\frac{9}{10}$ for G^1 og $\frac{3}{5}$ for G^2 . Det betyder således, at hvis en knude bryder ned, er G^1 kun 10% fra at fungere optimalt i resten af netværket. G^2 er derimod 40% dårligere end den optimale situation, hvilket igen stemmer fint overens med vores umiddelbare konklusion, at G^1 på alle måder er bedre sammenhængende end G^2 . Udregningerne for klyngekoefficienten understøtter dette resultat, idet $C = \frac{4}{5}$ for G^1 , og $C = \frac{3}{5}$ for G^2 . Vi bemærker, at værdierne for C og \mathcal{E}_{loc} er tæt på hinanden, så C er en god approksimation for den lokale effektivitet.⁵

6.2 Robusthed over for fejl

Alle virkelige netværk kan risikere at bryde ned. I dette afsnit analyseres hhv. BA-grafen og KE-grafen (beskrevet i kapitel 5) med henblik på robusthed over for fejl. Ved fejl menes tilfældige nedbrud, som, vi antager, indtræffer uafhængigt af knudernes valens. Det kan diskuteres, hvorvidt dette er en rimelig antagelse, idet knuder med stor valens er mere belastet end knuder med lille valens. I infrastrukturen vil sandsynligheden for et uheld (eller lang ventetid pga. kø) typisk være betydeligt større i vigtige knudepunkter frem for eksempelvis i et trafikkrøds ude på landet. Omvendt kan antagelsen forsvares ved, at der er gjort mere for at undgå nedbrud i de vigtige knuder, så disse to faktorer vil i et vist omfang udligne hinanden. Nedbrud i knuderne med størst valens bliver studeret nærmere i kapitel 6.3.

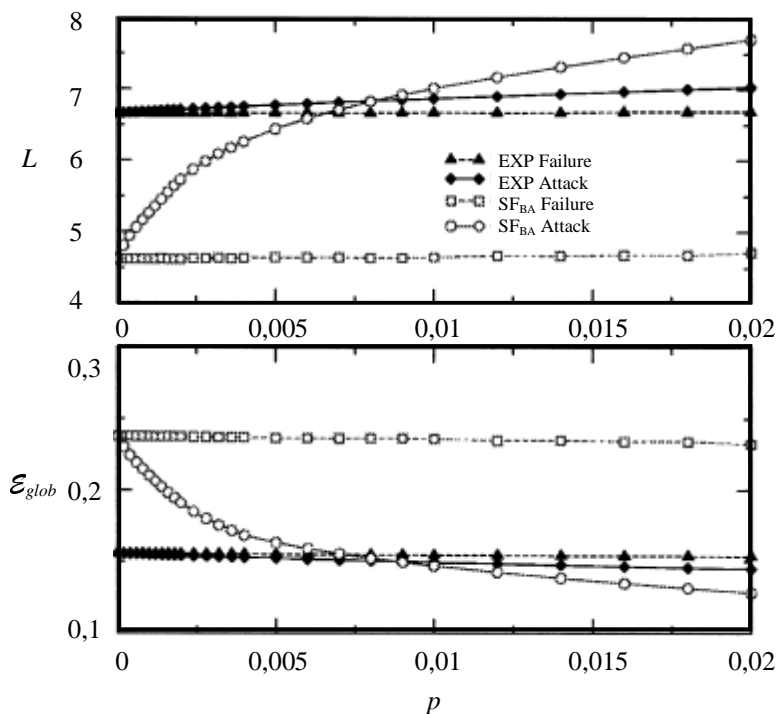
6.2.1 Fejl i BA-graf

Betragt et netværk med 5000 knuder og 10000 kanter opbygget efter BA-modellen beskrevet i kapitel 5.3.1 med dertil hørende sammenhæng $\frac{2 \cdot 10000}{5000} = 4$. Vi simulerer en fejl i netværket som et nedbrud i en tilfældig knude. Et sådant nedbrud betyder normalt, at netværkets effektivitet falder; vi ser i dette afsnit på både den globale og lokale effektivitet efter fjernelse af en

⁵dette er tilfældet, når delgraferne G_v består af få knuder. Jo mindre en graf er, jo mindre er den indbyrdes afstand mellem knuderne, hvormed summen af $\frac{1/d(u,v)}{2}$, $\forall u \neq v$ nærmer sig antallet af kanter i grafen, hvilket igen betyder en mindre forskel på \mathcal{E}_{loc} og C .

andel af knuderne. Resultaterne sammenlignes med effekten af et nedbrud i en sædvanlig stokastisk graf efter Erdős-Rényi modellen.⁶

Crucitti et al. simulerer i [16] et nedbrud i netværket, hvor andelen af det samlede antal knuder, der fjernes, er lig p . Effekten af nedbruddet, som en funktion af p , er afbildet i figur 6.2.



Figur 6.2 Effekten på L og \mathcal{E}_{glob} af et nedbrud i hhv. en ER- og BA-graf (med 5000 knuder og 10000 kanter) som funktion af andel nedbrudte knuder ($p \leq 0,02$). Kilde: Figur 3, [16].

I figuren betegner ”EXP Failure” og ”SF_{BA} Failure” tilfældige nedbrud i hhv. ER-grafen og BA-grafen.⁷ Resultaterne er gennemsnittet over ti realisationer med 5000 knuder og 10000 kanter for såvel ER-grafen som for BA-grafen.

For $p = 0$, dvs. udgangspunktet, hvor ingen knuder er fjernet, har BA-grafen en gennemsnitlig afstand mellem to knuder på ca. 4,6 mod ER-grafens på ca. 6,7. Når p stiger svagt, har det ikke nogen betydelig effekt på L for hverken

⁶Data for simulationen findes i [16], og alle resultater er et gennemsnit over ti forskellige realisationer.

⁷”EXP Attack” og ”SF_{BA} Attack” studeres først i næste afsnit.

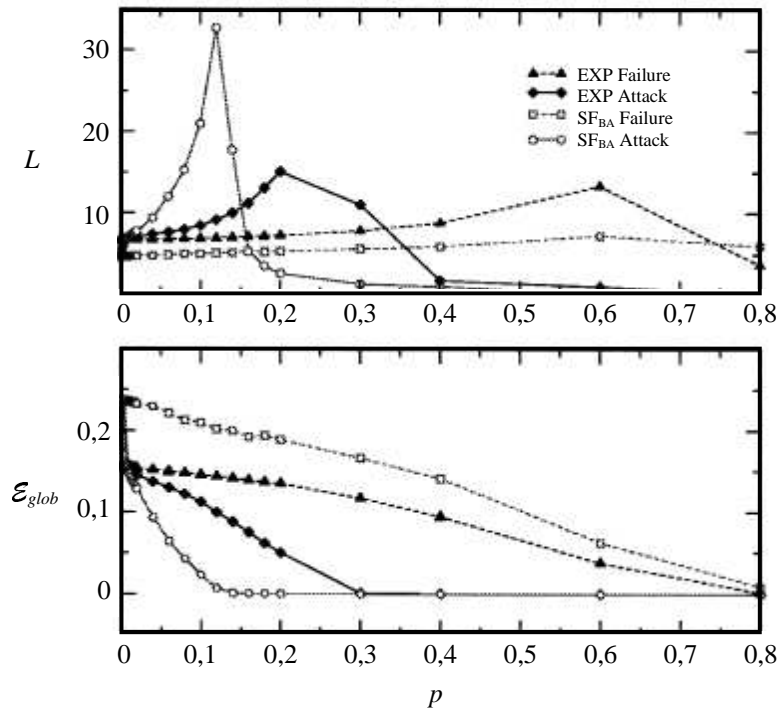
ER- eller BA-grafen; således er værdien for ER-grafen uændret på ca. 6,7 for $p = 0,02$, mens L for BA-grafen er steget med ca. 0,12 til 4,72. Ser vi på den globale effektivitet \mathcal{E}_{glob} beskrevet i kapitel 6.1, er situationen den samme. For BA-grafen er effektiviteten i udgangspunktet ca. 0,24, hvilket er væsentligt højere end værdien på ca. 0,16 for ER-grafen. Det betyder med andre ord, at graferne har en effektivitet på hhv. 24% og 16% sammenlignet med den komplette graf. Fjernes en lille andel af knuderne ($p \leq 0,02$), ses igen en stort set uændret robusthed af graferne. Forklaringen på de ubetydelige ændringer i robustheden skal findes i, at graferne består af mange knuder med forholdsvis lille valens og kun få store knudepunkter. Når andelen af knuder, der fjernes, ikke er større end 0,02, er sandsynligheden for fjernelse af en vigtig knude derfor lille.

Vi vender nu blikket mod større værdier af p . Indtil videre har vi kun set på $p \leq 0,02$, men hvilken betydning har det for robustheden, hvis andelen af fjernede knuder pludselig stiger væsentligt? I figur 6.3 er værdier af L og \mathcal{E}_{glob} for $0 \leq p \leq 0,8$ afbildet for både ER-grafen og BA-grafen.

Af figuren ses, at den gennemsnitlige afstand for BA-grafen ikke ændres betragteligt, selvom op imod 80% af knuderne fjernes. Således sker der kun en stigning fra ca. 4,6 til ca. 6. Ser vi derimod på ER-grafen, sker der en betydelig stigning indtil $p = 0,6$, hvorefter L falder til en værdi på ca. 4, som ligger under startværdien på 6,7. Det vil dermed sige, at bruger vi L som robusthedsmål, er ER-grafen mere robust efter 80% af knuderne er fjernet, hvilket åbenlyst giver et forkert billede af virkeligheden. Dette sker, da man for at undgå divergens af L kun kigger på den sammenhængende del af grafen (beskrevet i kapitel 6.1). Man kan dermed aflæse af figuren, at ER-grafen begynder at blive usammenhængende for $p \approx 0,6$, da L antager sit maksimum i dette punkt.

Betragter vi i stedet den nederste del af figur 6.3, ser vi, at den globale effektivitet for såvel BA-grafen som ER-grafen er støt faldende for $0 \leq p \leq 0,8$, hvilket intuitivt stemmer overens med virkeligheden. Niveaue for BA-grafen ligger for små værdier af p væsentligt højere end for ER-grafen, men den globale effektivitet for de to grafer nærmer sig hinanden, jo større andel af knuderne der fjernes. Fjernes halvdelen af knuderne, falder \mathcal{E}_{glob} for BA-grafen fra 0,24 til ca. 0,11; modsat falder \mathcal{E}_{glob} for ER-grafen til ca. 0,07. Effektiviteten er altså stadig væsentligt højere for BA-grafen, selvom 50% af knuderne bryder ned. Sker det derimod for 80% af knuderne, er effektiviteten for begge grafer tæt på 0.

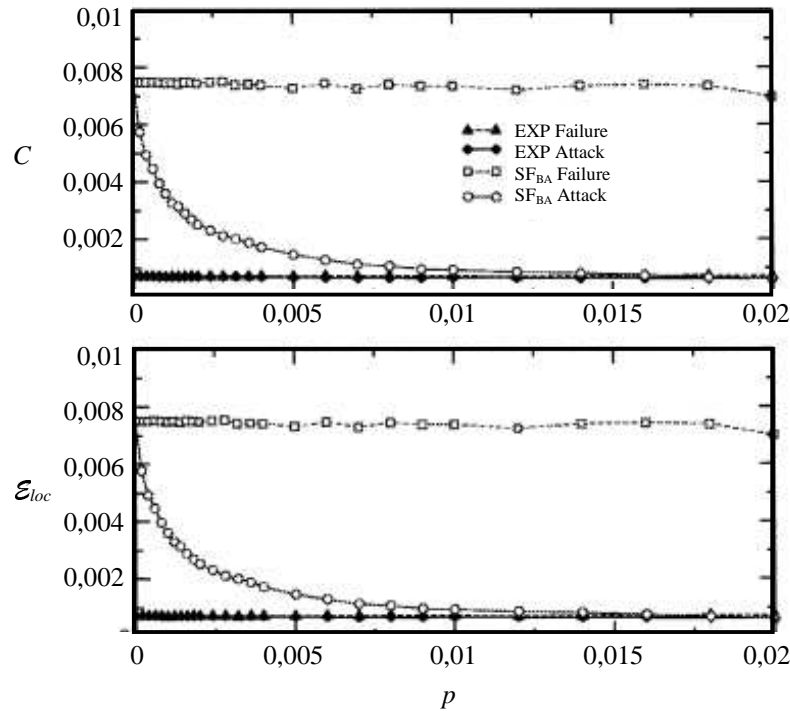
I figur 6.4 er værdierne for klyngekoefficienten og den lokale effektivitet af-



Figur 6.3 Effekten på L og \mathcal{E}_{glob} af et nedbrud i hhv. en ER- og BA-graf (med 5000 knuder og 10000 kanter) som funktion af andel nedbrudte knuder ($p \leq 0,8$). Kilde: Figur 4, [16].

bildet for små værdier af p . Begge grafer har pr. konstruktion en lille klyngekoefficient, hvilket betyder, at udgangspunktet for BA-grafen og ER-grafen er hhv. $C \approx 0,007$ og $C \approx 0,0008$. Det betyder samtidig, at værdien for den lokale effektivitet \mathcal{E}_{loc} er tilsvarende lille for begge grafer. Dette er tilfældet, idet hovedparten af knuderne i såvel BA-grafen som ER-grafen har lille valens, hvormed delgraferne G_v bliver små.⁸ Sker der en fejl i en knude, får det således stor negativ indflydelse for naboknudernes indbyrdes effektivitet, selvom det globalt set ikke har afgørende betydning for grafens robusthed. Af figur 6.4 ses, at der for små værdier af p ikke er en signifikant forskel i den lokale robusthed for nogen af graferne. For $p > 0,5$ er der en tilnærmelsesvist lineær nedadgående tendens for BA-grafen med henblik på C og \mathcal{E}_{loc} , mens situationen er uændret for ER-grafen (dette skal dog ses i lyset af den i forvejen meget lave værdi for ER-grafen). Udviklingen er illustreret i figur 6.5.

⁸Se evt. fodnote 5 i kapitel 6.1.



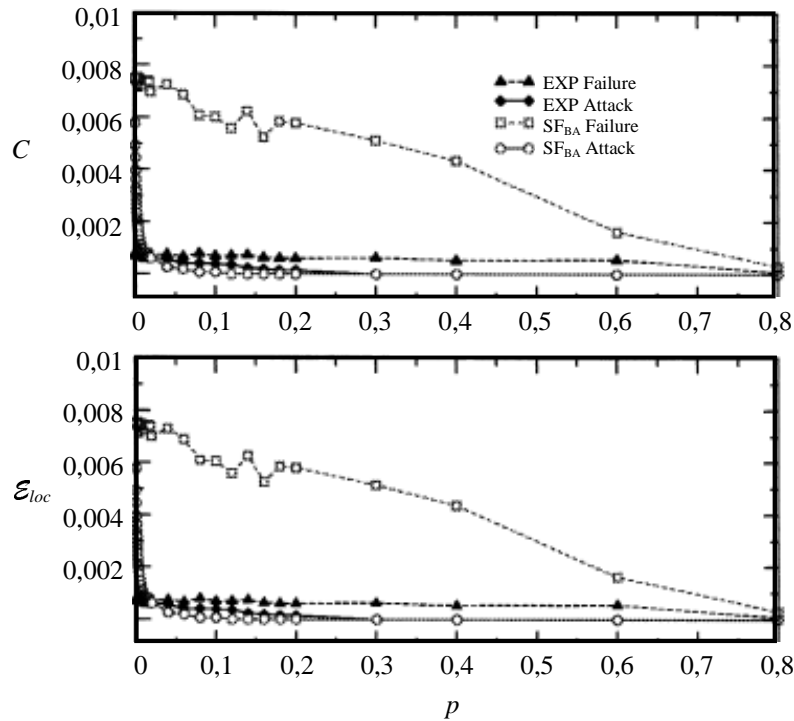
Figur 6.4 Effekten på C og ε_{loc} af et nedbrud i hhv. en ER- og BA-graf (med 5000 knuder og 10000 kanter) som funktion af andel nedbrudte knuder ($p \leq 0,02$). Kilde: Figur 5, [16].

6.2.2 Fejl i KE-graf

Betragt igen et netværk med 5000 knuder og 10000 kanter - denne gang konstrueret ud fra algoritmen beskrevet i kapitel 5.3.4 med $p = 0,1$. Fremgangsmåden resulterer i en scale-free KE-graf, som foruden de sædvanlige kendetegn ved scale-free grafer besidder den særlige egenskab, at klyngekoeficienten er høj (jvf. Watts-Strogatz modellen). Vi udfører en tilsvarende analyse for KE-grafer som for BA-grafer og sammenligner resultaterne.⁹

Figur 6.6 afbilder udviklingen af robustheden globalt set som funktion af $p \leq 0,02$ (De relevante grafer i dette afsnit er "EXP Failure" og "SF_{KE} Failure"). Værdierne for ER-grafen er naturligvis på samme niveau som før, så vi koncentrerer os om værdierne for KE-grafen. Her ses en initial værdi af den gennemsnitlige afstand på ca. 9,4, hvilket ligger betydeligt højere end

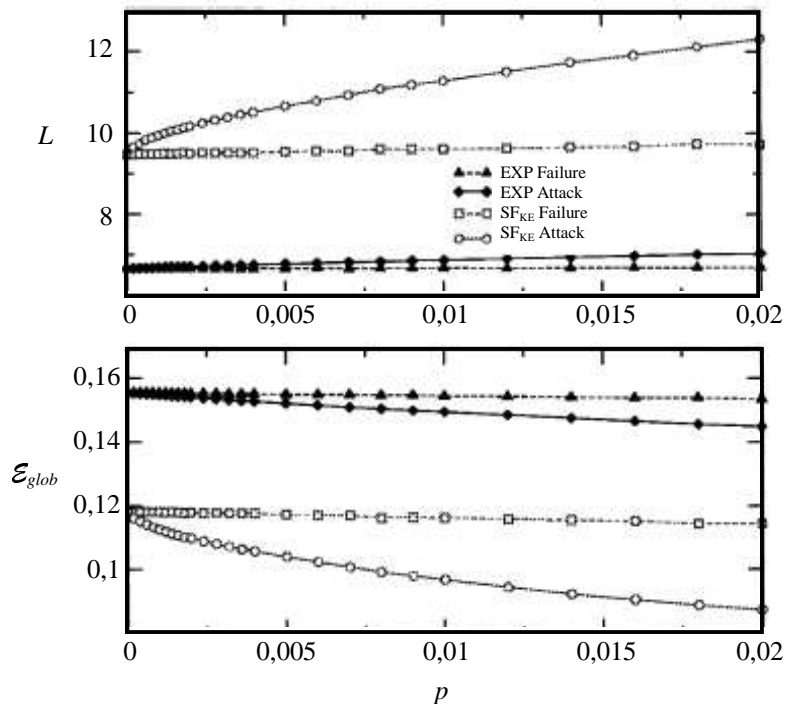
⁹Data for simulationen findes i [16], og alle resultater er et gennemsnit over ti forskellige realisationer.



Figur 6.5 Effekten på C og \mathcal{E}_{loc} af et nedbrud i hhv. en ER- og BA-graf (med 5000 knuder og 10000 kanter) som funktion af andel nedbrudte knuder ($p \leq 0,8$). Kilde: Figur 6, [16].

værdierne for både ER- og BA-grafen. Den gennemsnitlige afstand for BA-grafen var kun på 4,6, så det kræver altså i gennemsnit fem kanter mere at forbinde to knuder i KE-grafen end i BA-grafen, når størrelsen af graferne er ens. Det betyder samtidig, at den globale effektivitet også er lavere for KE-grafen end for både ER- og BA-grafen. Således fungerer KE-grafen som udgangspunkt kun ca. 12% så godt som den komplette graf sammenlignet med 24% for BA-grafen. Når antallet af nedbrudte knuder udgør under 2% af den samlede andel, er den globale effektivitet nær ved uændret, hvilket også var tilfældet for BA-grafen.

Figur 6.7 afbilder udviklingen for værdier af $p \leq 0,8$. Som det også var tilfældet for BA-grafen, ses en lineær nedgang i den globale effektivitet for p voksende. KE-grafen når dog allerede niveauet nul, når 60% af knuderne fjernes, modsat 80% for BA-grafen. Ser vi på værdien for den gennemsnitlige afstand, bliver det igen tydeligt, hvorfor L ikke er et optimalt mål for robustheden. For små værdier af p følger L og \mathcal{E}_{glob} fint hinanden, men for $p > 0,4$

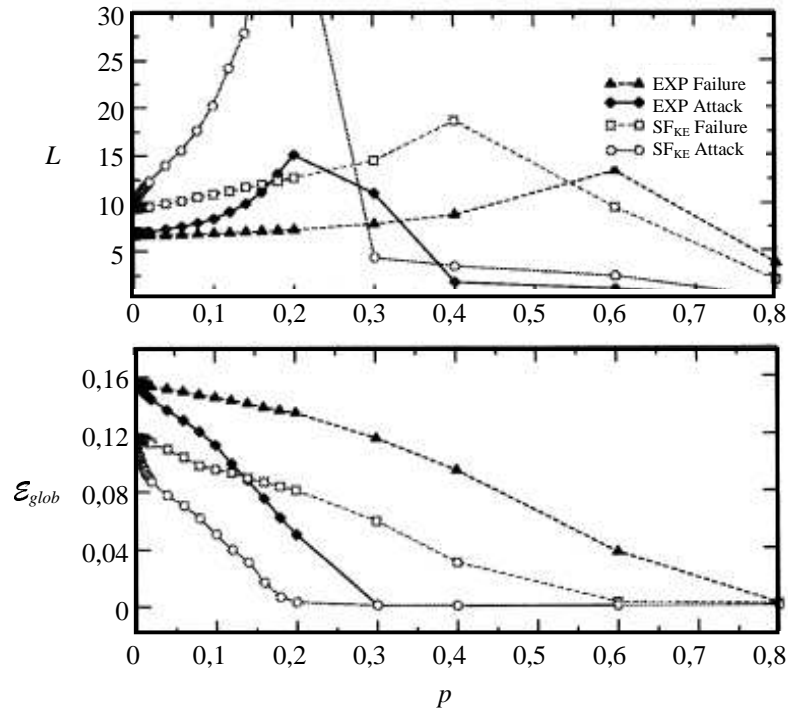


Figur 6.6 Effekten på L og \mathcal{E}_{glob} af et nedbrud i hhv. en ER- og KE-graf (med 5000 knuder og 10000 kanter) som funktion af andel nedbrudte knuder ($p \leq 0,02$). Kilde: Figur 7, [16].

er grafen for L aftagende, hvilket ikke er ensbetydende med en mere robust graf, men i stedet et udtryk for, at KE-grafen bliver usammenhængende, når 40% af knuderne bryder ned. Af resultaterne afbildet i figur 6.6 og figur 6.7 kan vi tydeligt konkludere, at prisen for en højere klyngekoefficient betales med en væsentligt forringet global effektivitet med henblik på tilfældige nedbrud.

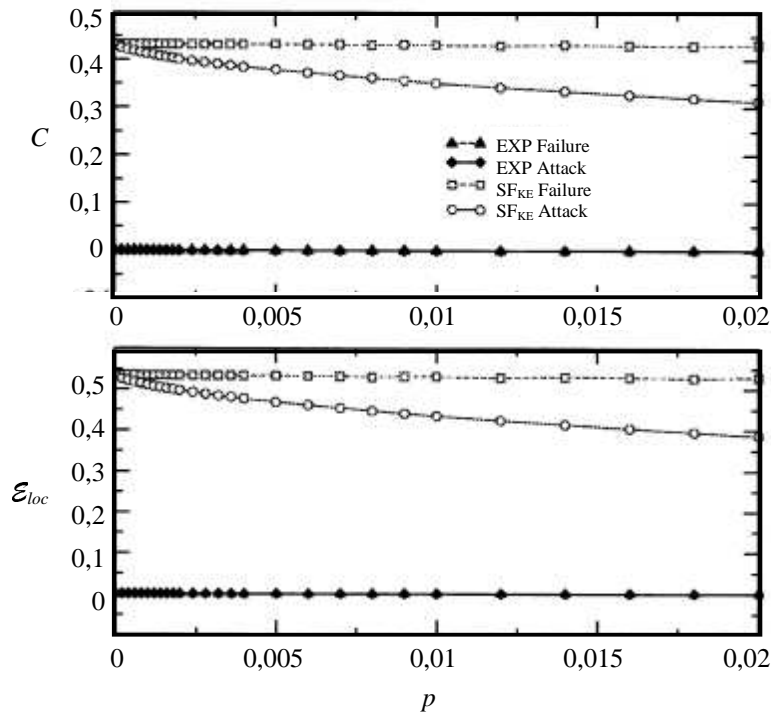
Det interessante er nu at undersøge, hvorvidt KE-grafen er mere robust lokalt set end ER- og BA-grafen. Figur 6.8 illustrerer den lokale effektivitet \mathcal{E}_{loc} og klyngekoefficienten C , så længe andelen af nedbrudte knuder højst udgør 2% af det samlede antal.

Såvel \mathcal{E}_{loc} som C er i al væsentlighed uændret for KE-grafen, når $p \leq 0,02$, hvilket også var tilfældet for BA- og ER-grafen. Der er imidlertid en betydelig forskel i niveauet for de to scale-free grafer. Den lokale effektivitet for KE-grafen er således på ca. 0,54 i udgangspunktet, hvilket er hele 0,533 højere



Figur 6.7 Effekten på L og \mathcal{E}_{glob} af et nedbrud i hhv. en ER- og KE-graf (med 5000 knuder og 10000 kanter) som funktion af andel nedbrudte knuder ($p \leq 0,8$). Kilde: Figur 8, [16].

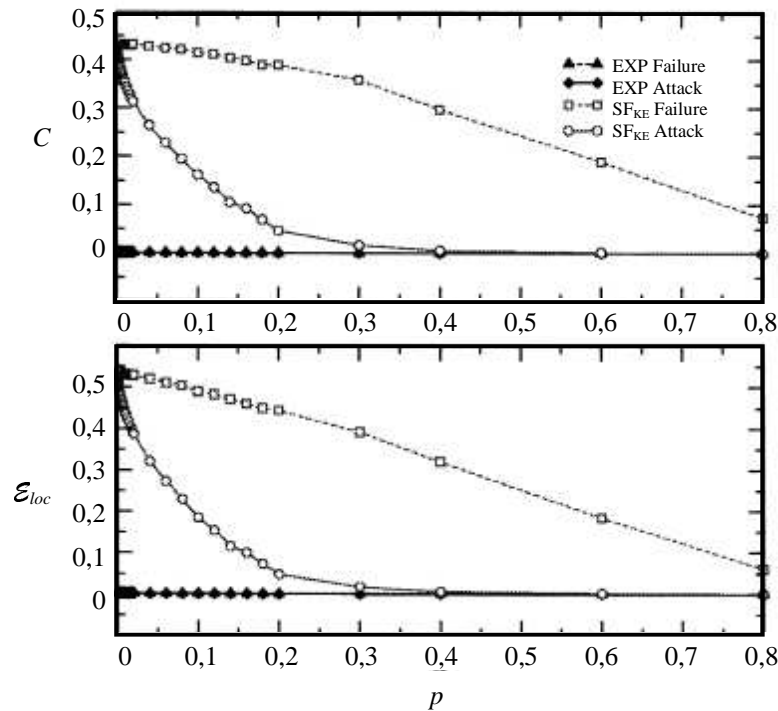
end værdien for BA-grafen. Hvor BA-grafen altså var særdeles sårbar over for fejl lokalt set, er KE-grafen kun 46% dårligere end den komplette graf. Da vi sammenlignede BA- med ER-grafen, argumenterede vi for, at C var en god approksimation for \mathcal{E}_{loc} pga. små delgrafer. Det gælder ikke i samme grad for KE-grafen, da delgrafen af naboer til en bestemt knude ikke altid er lille pga. konstruktionen af KE. Værdien af C på 0,43 giver dog stadig en vis indikation om den lokale effektivitet, samtidig med at udviklingen for C afspejler udviklingen for \mathcal{E}_{loc} . Udviklingen for $p > 0,02$ (afbildet i figur 6.9) stemmer fint overens med vores umiddelbare forestilling om et gradvist fald i C og \mathcal{E}_{loc} . Faldet sker dog forholdvist langsomt, så selvom der opstår fejl i 30% af knuderne, betyder det kun en tilbagegang på 14 procentpoint i den lokale effektivitet. Således opretholder KE-grafen også et væsentligt højere niveau for \mathcal{E}_{loc} for store p . Værdien af \mathcal{E}_{loc} for KE-grafen med $p = 0,8$ er for eksempel højere end \mathcal{E}_{loc} for BA-grafen med $p = 0$!



Figur 6.8 Effekten på C og \mathcal{E}_{loc} af et nedbrud i hhv. en ER- og KE-graf (med 5000 knuder og 10000 kanter) som funktion af andel nedbrudte knuder ($p \leq 0,02$). Kilde: Figur 9, [16].

6.3 Robusthed over for angreb

Det er med tiden blevet mere og mere relevant at undersøge forskellige netværks robusthed over for angreb. Ved angreb forstås nedbrud, som ikke sker tilfældigt, men som derimod er forårsaget med den hensigt at forvolde størst mulig skade. Visse netværk er naturligvis mere udsat for angreb end andre; f.eks. er flynettet et interessant mål for terrorister, mens internettet er et mål for hackere. Modsat er f.eks. telefonnettet ikke i øjeblikket et truet netværk; her er det i stedet først og fremmest fejl, som er skyld i nedbrud. Dette afsnit omhandler på samme måde som det forrige robustheden af henholdsvis ER-, BA- og KE-grafen, men nu med henblik på angreb. Et angreb simuleres som fjernelse af den knude i grafen med størst valens. Idet vi kigger på ikke-vægtede grafer, har vi dermed antaget, at eventuelle terrorister vælger sit mål ud fra en betragtning om, hvor effekten af angrebet bliver størst. Det kan naturligvis diskuteres, hvorvidt denne antagelse er rimelig, idet terroristerne i så fald skal have fuldt kendskab til hele netværket. Samtidig må



Figur 6.9 Effekten på C og \mathcal{E}_{loc} af et nedbrud i hhv. en ER- og KE-graf (med 5000 knuder og 10000 kanter) som funktion af andel nedbrudte knuder ($p \leq 0,8$). Kilde: Figur 10, [16].

man formode, at sikkerheden er betydeligt bedre på de største knudepunkter, hvilket medfører sværere vilkår for terroristerne. Alt andet lige må det dog formodes, at risikoen for et terrorangreb i f.eks. Heathrow Airport er betydeligt større end i Billund Lufthavn pga. den større effekt, hvis angrebet skulle lykkes. Derfor kan det med rimelighed retfærdiggøres, at et angreb betyder fjernelse af knuden med størst valens, hvilket er udgangspunktet for simulationerne i dette afsnit.¹⁰

6.3.1 Angreb i BA-graf

Lad en BA-graf med 5000 knuder og 10000 kanter være givet. Figur 6.2 i kapitel 6.2.1 illustrerer udviklingen i den gennemsnitlige afstand samt den globale effektivitet. I figuren betragter vi nu kurverne "SF_{BA} Attack" og "EXP Attack", som henholdsvis betegner angreb i BA-grafen og angreb i ER-grafen, hvor ER-grafen igen er den klassiske stokastiske graf med samme

¹⁰Data findes i [16].

antal knuder og kanter som BA-grafen.

I tilfældet med nedbrud pga. fejl så vi ikke nogen betydelig udsving for $p \leq 0,02$ for hverken BA- eller ER-grafen. Dette billede har ændret sig betragteligt, når vi ser på angreb. Den initiale værdi for BA-grafen på 0,24 falder således kraftigt, så snart knuden med størst valens fjernes. Sker der noget med 1% af de vigtigste knuder, medfører det et fald på ca. 10 procentpoint eller med andre ord et fald på over 40% i grafens globale effektivitet! Ser vi i stedet på den klassiske graf, har det ikke nogen betydelig effekt på hverken L eller \mathcal{E}_{glob} , at de vigtigste knuder fjernes, så længe $p \leq 0,02$. Selvom startniveauet for BA-grafen var betydeligt højere end for ER-grafen (0,24 mod 0,16), skal der fjernes under 1% af de vigtigste knuder, for at niveauet for ER-grafen er højest.

Figur 6.3 viser udviklingen af effektiviteten efter et angreb for $p \leq 0,8$. Vi ser, at \mathcal{E}_{glob} for BA-grafen allerede når niveauet nul for $p \approx 0,14$, mens ER-grafen først bryder helt sammen, når 30% af de vigtigste knuder er fjernet. Resultaterne her står i skarp kontrast til dem for tilfældige nedbrud, hvor 80% af knuderne kunne fjernes, før niveauet nåede nulpunktet. Vi ser nu i endnu højere grad end tidligere, at den gennemsnitlige afstand ikke er et velegnet robusthedsmål. L stiger godt nok planmæssigt for små p , hvilket svarer til faldet i \mathcal{E}_{glob} , men for $p \approx 0,1$ falder L pludselig igen, indikerende en bedre robusthed, hvilket åbenlyst er forkert. F.eks. vil BA-grafen ifølge L -målet være over dobbelt så god, når 20% af knuderne er angrebet, som hvis ingen knuder var angrebet.

I figur 6.4 og figur 6.5 er den lokale effektivitet afbildet. Husk på at både BA- og ER-grafen som udgangspunkt besidder en meget lav lokal effektivitet/klyngekoefficient med værdier på henholdsvis 0,007 og 0,0008. Vi så tidligere, at \mathcal{E}_{loc} for BA-grafen aftog tilnærmelsesvist lineært som en funktion af p i intervallet $[0; 0,8]$, så længe der var tale om tilfældige nedbrud. Ved angreb sker faldet næsten øjeblikkeligt; angribes således 0,5% af knuderne, falder effektiviteten fra 0,007 til under 0,002, og angribes yderligere 0,5% er niveauet på højde med effektiviteten for ER-grafen.

Når der fokuseres på robusthed over for angreb klarer ER-grafen sig generelt særdeles godt sammenlignet med BA-grafen. Grunden er, at ER-grafen er ret homogen og uden de store knudepunkter; det gør derfor ikke den store forskel, hvilke knuder i en ER-graf, der fjernes, da ingen knuder er vitale for grafen. Selvom den globale effektivitet er væsentligt lavere som udgangspunkt i ER-grafen end i BA-grafen, er ER-grafen altså interessant, hvis ønsket er

at designe et netværk, som er bedst muligt beskyttet mod angreb. Det kan f.eks. være tilfældet i en krigssituation, hvor risikoen for et angreb på en militærbase er specielt stor. Det kan derfor være af helt afgørende betydning, at hele militærnetværket ikke bryder sammen efter et angreb.¹¹

6.3.2 Angreb i KE-graf

Vi undersøger nu effekten af et angreb i KE-grafen (med $p = 0,1$), der ligesom BA- og ER-grafen består af 5000 knuder og 10000 kanter. Kurven "SF_{KE} Attack" i figur 6.6 og figur 6.7 i kapitel 6.2.2 illustrerer udviklingen i den globale effektivitet som følge af et angreb på de vigtigste knuder.

Tidligere så vi, at den initiale globale robusthed for KE-grafen lå lavere end for både ER- og BA-grafen. Ved tilfældige fejl opførte alle tre grafer sig i al væsentlighed ens, hvilket vil sige ingen ændring for små p og et gradvist fald for større p . Når nedbruddet i stedet sker grundet et angreb på de vigtigste knuder, sker faldet i den globale effektivitet betydeligt hurtigere. Hvis 1% af de vigtigste knuder i KE-grafen angribes, betyder det et fald i \mathcal{E}_{glob} fra 0,12 til 0,1. Selvom et angreb på 1% af knuderne således betyder et fald i effektiviteten på ca. 17%, er det væsentligt mindre end for BA-grafen, hvis effektivitet faldt med over 40% efter et tilsvarende angreb. BA-grafen nåede nulpunktet for \mathcal{E}_{glob} ved $p \approx 0,14$ og trods KE-grafens lavere initiale effektivitet, kan der her fjernes op til 20% af knuderne, før \mathcal{E}_{glob} bliver nul. Selvom KE-grafen er følsom over for angreb, er den altså mere robust end BA-grafen; faktisk følger udviklingen for KE-grafen samme mønster som ER-grafen. For små værdier af p er faldet ganske vist mindst for ER-grafen, men set over en større skala, aftager \mathcal{E}_{glob} med næsten samme procentsats. Grundet en højere startværdi, varer det dog lidt længere, inden ER-grafen når niveau nul.

Den lokale effektivitet er afbildet i figur 6.8 og 6.9. Vi husker, at startniveauet for KE-grafen er betragteligt højere end for såvel ER- som BA-grafen (KE-grafen blev jo netop designet med henblik på at få en scale-free graf med høj klyngekoefficient). For små værdier af p blev den lokale effektivitet ikke påvirket af tilfældige nedbrud. Når det i stedet er de vigtigste knuder, der fjernes, medfører det et fald i \mathcal{E}_{loc} fra 0,54 til 0,39 for $p = 0,02$. Sammenligner vi med udviklingen for BA-grafen, sker faldet betydeligt langsommere for KE-grafen, ligesom det var tilfældet med den globale effektivitet. Der skal således fjernes 40% af de vigtigste knuder i KE-grafen, for at den lokale effektivitet er lig nul sammenholdt med ca. 5% for BA-grafen.

¹¹Anthony Dekker skriver mere om militærnetværk i bl.a. [1] og [2].

6.3.3 Konklusion på analysen

Ved at benytte robusthedsmålet introduceret i kapitel 6.1 kan vi ud fra analysen om konsekvenserne af tilfældige nedbrud konkludere, at de tre grafer, ER, BA og KE, hver især har både fordele og ulemper. Den klassiske stokastiske graf, ER-grafen, udmærker sig både globalt og lokalt ved at holde et særdeles stabilt niveau efter nedbrud i et vist antal knuder. Til gengæld er niveauet meget lavt sammenlignet med den komplette graf, samtidig med at ER-grafen sjældent optræder i virkelige netværk. Modsat ER-grafen er begge scale-free grafer relevante i virkeligheden, så det er sammenligningen mellem disse grafer, der er mest interessant. Her fandt vi, at BA-grafen klarede sig betydeligt bedre globalt set end KE-grafen, mens det omvendte gjorde sig gældende på det lokale plan.

De to scale-free grafer klarer sig tilfredsstillende, når nedbruddene i netværket sker tilfældigt. Begge grafer består overvejende af knuder med lille valens og kun få knudepunkter. Når der sker en fejl, er der således størst sandsynlighed for, at fejlen sker i en knude med lille valens, hvilket betyder, at det kun er et begrænset antal knuder, der bliver berørt. Situationen ved angreb er den direkte modsatte; her er det netop knuderne med størst valens, der bliver fjernet, hvilket hurtigt påvirker en stor del af det resterende netværk. Af de to scale-free grafer klarer KE-grafen sig bedst, da den er lidt mindre følsom i dette tilfælde. ER-grafen er dog KE-grafen overlegen, når det gælder global effektivitet efter angreb, men dette skyldes udelukkende en højere initial værdi.

Hvis et nyt netværk skal designes for at sikre optimal robusthed med hensyn til både fejl og angreb, er der således flere faktorer, der spiller ind. Først skal risikoen for angreb overvejes. Hvis der konstant er en direkte trussel som f.eks. i en krigssituation, vil en konstruktion efter ER-modellen være at foretrække. Vurderes det, at risikoen for et angreb er minimal, er spørgsmålet, om global eller lokal effektivitet har højest prioritet. Hvis global effektivitet er vigtigst, er BA-grafen overlegen, og i modsat fald vil KE-grafen være det bedste valg.

Der er dog endnu en faktor, som har indflydelse på konstruktionen af et velfungerende netværk. Et nedbrud i et netværk medfører en øget risiko for, at den resterende del af netværket bliver overbelastet og i løbet af kort tid bryder sammen. Denne vigtige faktor vil blive beskrevet nærmere i følgende kapitel.

Kapitel 7

Kaskadeeffekter efter nedbrud

Vi har indtil nu studeret den statiske tilstand efter ét eller flere nedbrud som følge af enten fejl eller angreb, men hvad sker der i perioden efter et nedbrud, indtil netværket bliver genoprettet? Man kan let forestille sig, at et uheld i trafikken kan få konsekvenser for andre dele af infrastrukturen. Antag at der sker et uheld i et motorvejskryds i myldretiden, og at al trafik herigennem går i stå. I den efterfølgende periode spredes nyheden om trafikuheldet over radioen og GPS'en, så bilisten kan nå at vælge en alternativ rute frem for at holde i kø. I nyere navigationsanlæg findes det såkaldte TMC-system, der via FM-båndet informerer GPS'en om uheldet, som derefter guider bilisten udenom ulykkesstedet. Der er således en ret stor sandsynlighed for, at adskillige bilister vælger at passere det pågældende motorvejskryds via en alternativ rute. Denne rute følger antageligt mindre veje, hvis konstruktion ikke er egnet til den pludselige ekstra trafik. Det medfører en væsentligt langsommere trafik ad disse veje end normalt. Uheldet i motorvejskrydset har dermed ikke blot haft en *umiddelbar* negativ effekt, men har i løbet af kort tid også forårsaget en lavere effektivitet på andre dele af netværket. En tilsvarende kaskadeeffekt kan f.eks. opstå i et elnetværk, hvis kablerne ikke har nok kapacitet til at kunne håndtere den ekstra belastning, der opstår ved et uheld andetsteds i netværket. Det er således særdeles relevant at studere disse kaskadeeffekter nærmere med henblik på at komme dem i forkøbet.¹

For at kunne foretage en undersøgelse af kaskadeeffekterne beskrives i dette kapitel en model introduceret af Crucitti et al. i [20]. Lad en graf G være givet ved n knuder og m kanter. For at analysere belastningen af G lader vi G være en vægtet graf, hvor hver knude $i \in V$ på tidspunkt t har tilknyttet en kapacitet C_i og en belastning $L_i(t)$. Lad endvidere hver indgang $\{\varepsilon_{ij}\}$ i

¹Emnet er desværre ret overset, så antallet af artikler er meget begrænset, men jeg har valgt at tage udgangspunkt i [20].

$(n \times n)$ -matricen for G være tildelt en værdi mellem nul og én indikerende effektiviteten af kanten mellem knude i og j . Hvis der ikke eksisterer en kant mellem i og j , sættes $\varepsilon_{ij} = 0$. I starttidspunktet (dvs. for $t = 0$) sættes $\varepsilon_{ij} = 1$ for alle eksisterende kanter, hvilket svarer til, at alle kanter i starten fungerer optimalt.² Når der sker et nedbrud, ændres fordelingen af strømmen i grafen, og her er det vigtigt, at kapaciteten af en knude forbliver større end belastningen. Når det er tilfældet, er effektiviteten af de incidente kanter stadig én. Når belastningen derimod overstiger kapaciteten, resulterer det i en lavere effektivitet på kanterne.

Antag nu at knuderne i og j anvender den mest effektive sti imellem dem til at kommunikere med hinanden ($\forall i, j \in V$). Den mest effektive sti mellem i og j findes ved at udregne den *harmoniske* middelværdi \mathcal{H} for alle de mulige stier fra i til j . Den sti, hvis harmoniske middelværdi er tættest på én, vælges som den mest effektive. Den harmoniske middelværdi af tallene a_1, \dots, a_n udregnes ved:

$$\mathcal{H} = \frac{n}{\sum_{i=1}^n \frac{1}{a_i}} .$$

Belastningen $L_i(t)$ af knude i på tidspunkt t udregnes som det samlede antal effektive stier, der på tidspunkt t går igennem knude i . Vi antager, at kapaciteten C_i af knude i afhænger af belastningen på starttidspunktet på følgende måde:

$$C_i = \alpha L_i(0), \quad i = 1, \dots, n, \quad \alpha \in [1, \infty[.$$

Kapaciteten er således mindst ligeså stor som belastningen i udgangspunktet, hvilket betyder, at alle knuder er i stand til at håndtere den nødvendige strøm i grafen på tidspunkt $t = 0$. Forskellen på C_i og $L_i(t)$ er en faktor α , der kan betragtes som en toleranceparameter for knuderne; jo højere værdi af α for en knude i , jo større belastning kan knude i tolerere. Sammenhængen mellem C_i og $L_i(t)$ betyder samtidig, at der er begrænset kapacitet til rådighed, idet $\alpha \neq \infty$. Det stemmer fint overens med langt de fleste virkelige netværk, der pga. omkostningerne forbundet ved det ikke kan have ubegrænset kapacitet.

Når der sker et nedbrud i en knude (f.eks. pga. et trafikuheld), fjernes knuden (ligesom i simulationerne i kapitel 6), og der sker en omfordeling af strømmen til den resterende del af netværket. De mest effektive stier på tidspunkt t er ikke nødvendigvis de mest effektive på tidspunkt $t + 1$, hvilket også betyder, at belastningen af de enkelte knuder muligvis ændrer sig over tid. Det

²Da alle indgange i den tilhørende matrix har værdien nul eller en, svarer situationen i starttidspunktet til en ikke-vægtet graf.

er denne ændring, som i værste fald fører til overbelastninger visse steder i netværket, hvis knuderne ikke har tilstrækkelig kapacitet. Lad $\varepsilon_{ij}(t)$ betegne effektiviteten af kanten mellem knude i og j på tidspunkt t . På hvert tidspunkt t benytter vi da følgende regel til at korrigere effektiviteten af de enkelte kanter:

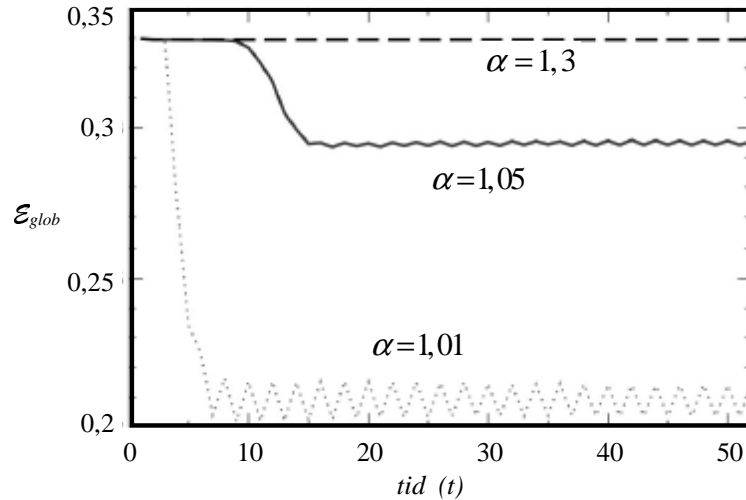
$$\varepsilon_{ij}(t+1) = \begin{cases} \varepsilon_{ij}(0) \frac{C_i}{L_i(t)}, & \text{hvis } L_i(t) > C_i; \\ \varepsilon_{ij}(0), & \text{hvis } L_i(t) \leq C_i. \end{cases}$$

hvor j betegner knude i 's direkte naboer. Ved at benytte denne algoritme, reduceres effektiviteten på de kanter, som er incidente på en overbelastet knude. Dette gør det mindre attraktivt at vælge de pågældende kanter, hvorfor strømmen i stedet følger en alternativ sti. Når belastningen er aftaget, sættes effektiviteten igen til den oprindelige værdi. Dette svarer nøjagtigt til situationen med trafikeksemplet, hvor bilisterne vælger alternative veje ved kødannelser.

For at illustrere modellen og undersøge hvornår kaskadeeffekterne indtræffer i hhv. ER- og BA-grafer, ser vi på et eksperiment udført vha. simulation af Crucitti et al. i [20]. Dette skal endvidere hjælpe os til at finde en nødvendig værdi for toleranceparameteren og dermed kapaciteten, for at knuderne med rimelighed er rustet til at kunne klare en eventuel ekstra belastning.

Vi betragter en ER-graf og en BA-graf, som begge har 2000 knuder og 10000 kanter. Alle resultater er udregnet som et gennemsnit af værdierne for 10 forskellige realisationer. Figur 7.1³ viser effekten over tid af et nedbrud i en tilfældig knude i BA-grafen på tidspunkt $t = 0$. Et nedbrud i grafen medfører straks en ændring af strømmen, hvor ændringen sker ifølge førnævnte algoritme. Vi ser, at denne ændring ikke påvirker den samlede robusthed, hvis toleranceparameteren er 1,3. Det vil med andre ord sige, at hvis kapaciteten for hver knude er 30% større end den initiale belastning, da er netværket tilstrækkeligt robust til at kunne håndtere omfordelingen af strømmen ved nedbrud i en enkelt knude. Effektiviteten af grafen er i dette tilfælde konstant ca. 0,33. Størrelsen af toleranceparameteren er imidlertid vital for netværkets evne til håndtering af den ekstra strøm, hvilket illustreres af kurverne for $\alpha = 1,05$ og $\alpha = 1,01$. Hvis kapaciteten er 5% større end belastningen, falder effektiviteten over tid og stabiliserer sig på et niveau, som er godt 10% lavere end effektiviteten på tidspunkt $t = 0$. Er kapaciteten derimod kun 1% større end den initiale belastning (dvs. $\alpha = 1,01$), da falder niveauet med ca. 40% over tid.

³Kilde: Crucitti et al. [20].

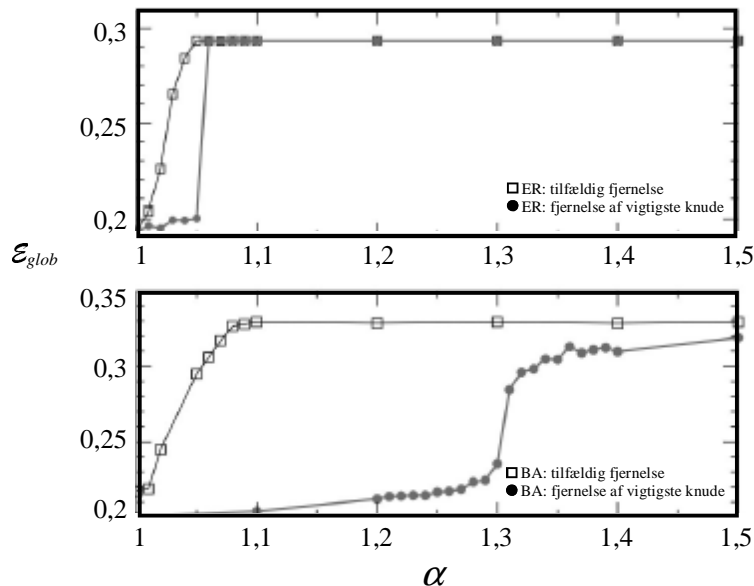


Figur 7.1 Global effektivitet efter et nedbrud på tidspunkt $t = 0$ i en BA-graf (med 2000 knuder og 10000 kanter) som funktion af tiden for forskellige værdier af toleranceparameteren α . Kilde: Figur 1, [20].

I figur 7.2⁴ understreges den vigtige sammenhæng mellem værdien af toleranceparameteren og robustheden på sigt efter et nedbrud. Figurerne afbilder den globale effektivitet (efter niveauet er stabiliseret) som en funktion af toleranceparameteren med henblik på at finde det nødvendige forhold mellem kapacitet og initial belastning, så kaskadeeffekterne begrænses. Den øverste del af figuren afbilder sammenhængen i ER-grafen, og den nederste del afbilder sammenhængen i den tilsvarende BA-graf. Kurverne med hvide kvadrater illustrerer sammenhængen ved tilfældig fjernelse af en knude, mens kurverne med sorte cirkler svarer til fjernelse af den vigtigste knude (dvs. knuden med størst belastning⁵). Ved ER-grafen er den optimale effektivitet efter et nedbrud ca. 0,29 - altså lidt lavere end ved BA-grafen, hvilket stemmer overens med vores tidligere analyse af robustheden i kapitel 6. Hvis kapaciteten kun er marginalt større end belastningen, falder effektiviteten i en ER-graf på sigt med op imod 35% efter fjernelse af en knude. En kapacitet på 5% over belastningen ($\alpha = 1,05$) betyder imidlertid, at effektiviteten på sigt ved tilfældig fjernelse er næsten uændret fra sit oprindelige niveau på ca. 0,29. Derimod er en toleranceparameter på 1,05 ikke tilstrækkelig, hvis der sker et nedbrud i den vigtigste knude; dette medfører således en effektivitet

⁴Kilde: Crucitti et al. [20].

⁵Dette svarer til et angreb, jvf. kapitel 6



Figur 7.2 Global effektivitet som funktion af toleranceparameteren α efter et nedbrud i en ER- og en BA-graf (begge med 2000 knuder og 10000 kanter). Kilde: Figur 2, [20].

på 0,20, hvilket svarer til et fald på ca. 30%. Som det fremgår af figuren, er en svag stigning på 2% i toleranceparameteren dog tilstrækkelig til at forhindre en kaskadeeffekt. Med en kapacitet på ca. 7% over den initiale belastning er ER-grafen således sikret mod kaskadeeffekter som følge af nedbrud i en enkelt knude, uanset om nedbruddet er tilfældigt eller et angreb.

Vi observerer en markant forskel fra ER- til BA-grafen, som igen skyldes den heterogene konstruktion af BA-grafen. Som tidligere nævnt resulterer en toleranceparameter på 1,01 i et fald på ca. 40% over tid ved nedbrud i en tilfældig knude. For at undgå kaskadeeffekter ved tilfældig fjernelse af en enkelt knude ser vi af den nederste del af figur 7.2, at det er nødvendigt med en kapacitet på 10% over den initiale belastning. Dette er højere end for ER-grafen, men den helt store forskel skal findes i fjernelse af den vigtigste knude. For ER-grafen undgik man kaskadeeffekter ved en toleranceparameter på ca. 1,07, mens en tilsvarende parameter for BA-grafen ikke har nogen nævneværdig positiv effekt. Selv med en kapacitet, som er 20% højere end nødvendigt i starttidspunktet, er effektiviteten på sigt kun ca. 0,21 efter et enkelt angreb - eller med andre ord ca. 35% lavere end den umiddelbare effekt. Først når toleranceparameteren overstiger 1,3, sker der noget med effektiviteten, som

hurtigt nærmer sig 0,3, hvilket i de fleste tilfælde vil være et acceptabelt niveau, da faldet er under 10%.

Ved design af et netværk, har vi i dette kapitel fundet, at det er særdeles vigtigt at tage højde for risikoen for kaskadeeffekter. Igen har opbygningen af netværket afgørende betydning for risikoen og omfanget af en eventuel kaskadeeffekt, da BA-grafer er væsentligt mere udsatte end ER-grafer. Som tidligere nævnt optræder BA-grafer hyppigst i virkelighedens netværk, så resultaterne om disse grafer er særligt vigtige. Vi fandt, at knuderne bør have en kapacitet på mindst 10% over det nødvendige for at kunne håndtere den ekstra trafik, som måtte opstå som følge af et tilfældigt nedbrud. Hvis man vurderer, at risikoen for nedbrud i de vigtigste knuder er stor, bør man i stedet sikre sig med en kapacitet på mindst 30% over det nødvendige, da man i modsat fald risikerer et totalt sammenbrud af netværket, indtil fejlen er udbedret.

Et aktuelt eksempel, som er genstand for megen debat for tiden, er den forestående motorvej omkring Silkeborg. En linieføring igennem byen betyder lukning af dele af ringvejen, mens arbejdet står på. Det er meget problematisk for myldretidstrafikken, som bliver ledt ud på små omfartsveje, der ikke har den fornødne kapacitet til at kunne klare den ekstra belastning, hvilket vil resultere i massive kødannelser. Det svarer i teorien til, at man fjerner en vigtig knude i en BA-graf, som ikke har en tilstrækkelig høj toleranceparameter, og som konsekvens mister op imod 60% af dens oprindelige effektivitet.

Kapitel 8

Konklusion

Specialet har haft til formål at give et overblik over forskellige grafter brugt til modellering af virkelige netværk og samtidig vurdere robustheden af disse grafter over for såvel fejl (uniformt tilfældige nedbrud) som angreb (nedbrud i den vigtigste knude - dvs. knuden med størst valens ved ikke-vægtede grafer).

Den klassiske stokastiske graf, ER-grafen (efter Erdős og Rényi), var tidligere den mest udbredte graftype til modellering af netværk. En ER-graf består groft sagt af et bestemt antal knuder forbundet tilfældigt via et bestemt antal kanter. Resultatet er en homogen graf, som er meget robust over for både fejl og angreb. Kendetegnet ved en ER-graf er netop, at alle knuder er (næsten) lige vigtige, så forskellen på tilfældige nedbrud og angreb er væsentligt mindre end ved andre grafer.

Mange virkelige netværk er imidlertid ikke homogene, da der kan være stor forskel på vigtigheden af de enkelte knuder. Barabási og Albert introducerede derfor i 1999 en scale-free graf (BA-grafen) som et alternativ til ER-grafen. BA-grafen har siden vundet stor udbredelse i kraft af en meget realistisk og samtidig simpel struktur. Når nye knuder skal forbindes med de eksisterende knuder i en BA-graf, sker det med sandsynligheder proportionale med de eksisterende knuders valens; på den måde bliver vigtige knuder vigtigere - ligesom det ofte vil være tilfældet i virkelige netværk. Dette resulterer i en graf, som er meget robust over for fejl, da der er stor sandsynlighed for, at et tilfældigt nedbrud sker i en mindre vigtig knude. Omvendt er BA-grafen sårbar over for angreb, da nedbrud i et vigtigt knudepunkt vil påvirke en stor del af netværket.

Der er stor forskel på, hvor effektivt de forskellige grafter fungerer; til

det formål så vi på den gennemsnitlige afstand L , klyngekoefficienten C , den globale effektivitet \mathcal{E}_{glob} samt den lokale effektivitet \mathcal{E}_{loc} . Det blev tydeliggjort, at \mathcal{E}_{glob} og \mathcal{E}_{loc} var de bedst velegnede effektivitetsmål. BA-grafen fungerede i udgangspunktet mere effektivt end ER-grafen, hvilket også var gældende efter flere tilfældige nedbrud. Hvis nedbruddene derimod skete som følge af angreb, ville ER-grafen fungere bedst.

	ER-graf	BA-graf
Robust over for fejl	+ / -	+
Robust over for angreb	+	-
Robust over for kaskadeeffekter	+	-
Udbredelse i virkelige netværk	-	+

Figur 8.1 Fordele og ulemper ved ER- og BA-grafer.

Endelig betragtede vi vægtede grafer for at vurdere risikoen for kaskadeeffekter efter et nedbrud i et netværk. Hver knude blev tildelt en kapacitet C_i og en belastning $L_i(t)$ på tidspunkt t . Så længe kapaciteten er større end belastningen, fungerer netværket tilfredsstillende. For en BA-graf så vi, at en kapacitet på 10% over den initiale belastning var tilstrækkelig for at undgå kaskadeeffekter ved et tilfældigt nedbrud. Hvis nedbruddet i stedet skete i den vigtigste knude, var kravet 30% ekstra kapacitet i forhold til den initiale belastning. For en ER-graf fandt vi, at det kun var nødvendigt med 7% ekstra kapacitet for at være sikret mod både fejl og angreb. Figur 8.1 illustrerer den overordnede forskel på ER- og BA-grafer.

For at kunne analysere et netværk korrekt, er det således af afgørende betydning at vælge den graftype, som bedst passer på beskrivelsen af netværket. I modsat fald vil analysen kunne give meget misvisende resultater.

Bilag A

Stikordsregister

Stikord	Defineret på side
A	
Afstand	16
Automorfi	19
B	
BA-graf	31
Belastning	60
Bro	15
D	
Delgraf	8
Diameter	16
E	
Endeknude	7
Effektivitet	17,42
ER-graf	23

Stikord	Defineret på side
G	
Gennemsnitlig afstand	28
Global effektivitet	43
Graf	7
Grafproces	24
H	
Harmonisk middelværdi	61
I	
Incident	7
K	
Kant	7
Kant-snit	14
Kantdisjunkt	9
Kantsammenhæng	14
Kant-transitiv	19
Kapacitet	10
Kaskadeeffekter	60
KE-graf	37
Klyngekoefficient	28
Knude	7
Knude-snit	13
Knudedisjunkt	9
Knudesammenhæng	14
Knude-transitiv	19
Komplet graf	8
Kreds	9
Kæmpekomponent	17

Stikord	Defineret på side
L	
LCD	34
Lokal effektivitet	43
Lukket tur	9
Løkke	7
M	
Maksimal delgraf	8
Med stor sandsynlighed	24
Mengers sætning	15
Minimal delgraf	8
Monoton egenskab	24
N	
n -parring	36
Nabo	7
O	
Optimalt sammenhængende	18
Orienteret graf	11
P	
Parallelle kanter	7
R	
Regulær graf	8

Stikord	Defineret på side
S	
Sammenhængende graf	9
Sammenhængskomponent	9
Scale-free graf	30
Simpel graf	7
Sti	9
Stokastisk graf	21
Stoptid	25
Strøm	10
Symmetrisk graf	19
T	
Toleranceparameter	61
Trivial graf	8
Tur	9
Tærskelfunktion	25
V	
Valens	8
Vægtet graf	9
W	
Watts-Strogatz graf	27
Æ	
Ægte delgraf	8
Å	
Åben tur	9

Bilag B

Notation

Symbol	Forklaring	Defineret på side
u, v, v_i	Knuder i en graf	7
e, e_l, e_{ij}	Kanter i en graf	7
V	Mængden af knuder: $\{v_1, \dots, v_n\}$	7
E	Mængden af kanter: $\{e_1, \dots, e_m\}$	7
G	En graf $G = (V, E)$ med n knuder og m kanter	7
$\Gamma(v)$	Mængden af v 's naboknuder	7
K_n	Den komplette graf med n knuder	8
$d(v), d_G(v)$	Valensen af knude v i en graf G	8
$\delta(G)$	Den mindste valens i G	8
$\Delta(G)$	Den største valens i G	8
\mathcal{Q}	En graf-egenskab	8
P_i	En sti i en graf	9
$\kappa(G)$	Knudesammenhængen	14
$\kappa'(G)$	Kantsammenhængen	14
e	En bro	15
$d(u, v)$	Afstanden mellem knude u og v	16
$\text{diam}(G)$	Diameteren af G	16
π	En automorfi	19
$\mathcal{G}(n, m)$	Udfaldsrum for stokastiske grafer	23

Symbol	Forklaring	Defineret på side
$G_{n,m}, G_m$	En stokastisk graf i $\mathcal{G}(n, m)$	23
$(G_t)_0^\infty$	En grafproces	24
τ	Stoptid for en grafproces	25
p	Sandsynlighed $p \in]0, 1[$	28
L	Gennemsnitlig afstand	28
C	Klyngekoeficient	28
G_v	Delgrafen af naboer til knude v	28
$\phi(L)$	Graf fremkommet af en n -parring L	36
ε_{uv}	Effektiviteten af kanten fra u til v	43
\mathcal{E}_{glob}	Den globale effektivitet af G	43
\mathcal{E}_{loc}	Den lokale effektivitet af G	43
V_v	Knudemængden hørende til G_v	44
C_i	Kapacitet af knude i	60
$L_i(t)$	Belastning af knude i på tidspunkt t	60
\mathcal{H}	Den harmoniske middelværdi	61
α	Toleranceparameter	61

Litteratur

- [1] Anthony H. Dekker, Bernard D. Colbert (2004). *Network Robustness and Graph Topology*, Proceedings of the 27th conference on Australasian computer science - Volume 26, 359-368.
Elektronisk tilgængelig på:
<http://crpit.com/confpapers/CRPITV26Dekker.pdf>

- [2] Anthony H. Dekker (2005). *Simulating network robustness for critical infrastructure networks*, Proceedings of the 28th Australasian conference on Computer Science - Volume 38, 59-67.
Elektronisk tilgængelig på:
<http://crpit.com/confpapers/CRPITV38Dekker.pdf>

- [3] Piet Van Mieghem (2005). *Robustness of Large Networks*, 2005 IEEE International Conference on Systems, Man and Cybernetics, 2372-2377.
Elektronisk tilgængelig på:
<http://www.nas.ewi.tudelft.nl/publications/2005/IEEE.SMC05.Hawai.PhysicsOfNetworks.pdf>

- [4] T. Tanizawa, G. Paul, R. Cohen, S. Havlin, H. E. Stanley (2005). *Optimization of network robustness to waves of targeted and random attacks*, Physical Review E 71, 047101-1-047101-4.
Elektronisk tilgængelig på:
<http://polymer.bu.edu/hes/articles/tpchs05.pdf>

- [5] Paolo Crucitti, Vito Latora, Massimo Marchiori, Andrea Rapisarda (2004). *Error and attack tolerance of complex networks*, Physica A 340, 388-394.
Elektronisk tilgængelig på:
http://www.w3.org/People/Massimo/papers/2004/tolerance_phya_04.pdf

- [6] Béla Bollobás, Christian Borgs, Jennifer Chayes, Oliver Riordan (2003). *Directed scale-free graphs*, Proceedings of the fourteenth annual ACM-SIAM symposium on Discrete algorithms, 132-139.
Elektronisk tilgængelig på:
<http://research.microsoft.com/~borgs/Papers/dirSCgrph.pdf>
- [7] K. Thulasiraman, M. N. S. Swamy (1992). *Graphs: Theory and Algorithms*, Wiley-Interscience.
- [8] Abdol-Hossein Esfahanian. *On the Evolution of Graph Connectivity Algorithms*, Computer Science and Engineering Department, Michigan State University.
Elektronisk tilgængelig på:
http://www.cse.msu.edu/~esfahani/book_chapter/Graph_connectivity_chapter.pdf
- [9] Béla Bollobás, Oliver Riordan (2003). *Robustness and Vulnerability of Scale-Free Random Graphs*, Internet Mathematics Vol 1, No. 1, 1-35.
Elektronisk tilgængelig på:
http://www.internetmathematics.org/volumes/1/1/pp001_035.pdf
- [10] Béla Bollobás (2001). *Random Graphs, second edition*, Cambridge University Press.
- [11] Béla Bollobás, Oliver Riordan (2002). *Mathematical results on scale-free random graphs*, In Handbook of Graphs and Networks. Wiley-VCH, 1-34.
- [12] D. J. Watts, S. H. Strogatz (1998). *Collective dynamics of 'small-world' networks*, Nature 393, 440-442.
Elektronisk tilgængelig på:
http://tam.cornell.edu/SS_nature_smallworld.pdf
- [13] Albert-László Barabási, Réka Albert (1999). *Emergence of Scaling in Random Networks*, Science Vol 286, 509-512.
- [14] Béla Bollobás, Oliver Riordan (2003). *The diameter of scale-free random graph*, Combinatorica, 1-24.
Elektronisk tilgængelig på:
<http://www.dpmms.cam.ac.uk/~omr10/diam/diam.pdf>

- [15] Vito Latora, Massimo Marchiori (2001). *Efficient Behavior of Small-World Networks*, Phys. Rev. Lett. 87, 198701.
Elektronisk tilgængelig på:
www.w3.org/People/Massimo/papers/2001/efficiency_prl.01.pdf
- [16] Paolo Crucitti, Vito Latora, Massimo Marchiori, Andrea Rapisarda (2003). *Efficiency of Scale-free Networks: Error and Attack Tolerance*, Physica A 320, 622-642.
Elektronisk tilgængelig på:
www.w3.org/People/Massimo/papers/2003/tolerance_physicaA_03.pdf
- [17] Konstantin Klemm, Víctor M. Eguíluz (2002). *Growing Scale-free Networks with Small-world Behavior*, Physical Review E 65, 057102.
Elektronisk tilgængelig på:
www.imedeia.uib.es/~victor/Networks/sw.pdf
- [18] Vito Latora, Massimo Marchiori (2002). *Economic Small-World Behavior in Weighted Networks*, cond-mat 0204089.
Elektronisk tilgængelig på:
www.arxiv.org/abs/cond-mat/0204089
- [19] Romualdo Pastor-Satorras, Alessandro Vespignani (2002). *Immunization of complex networks*, cond-mat 0107066v2.
Elektronisk tilgængelig på:
www.fen.upc.es/~romu/Papers/immuno.pdf
- [20] Paolo Crucitti, Vito Latora, Massimo Marchiori (2004). *Model for cascading failures in complex networks*, Physical Review E 69, 045104, 1-4.
Elektronisk tilgængelig på:
www.w3.org/People/Massimo/papers/2004/cascading_pre.04.pdf
- [21] Chris Godsil, Gordon Royle (2004). *Algebraic Graph Theory*, Springer.
- [22] László Babai (1994). *Automorphism groups, isomorphism, reconstruction*, Technical Report TR-94-10.
- [23] Michalis Faloutsos, Petros Faloutsos, Christos Faloutsos (1999). *On Power-Law Relationships of the Internet Topology*, Comm. rev. 29, 251-262.