

Universiteit
Antwerpen

Universiteit Antwerpen
Faculteit Wetenschappen
Departement Fysica
Academiejaar 2016-2017

GROEI EN EVOLUTIE VAN NETWERKEN
ALS STOCHASTISCH PROCES

Auteur:

Dietrich Elst

Promotor:

Prof. Dr. Jacques Tempere

PROEFSCHRIFT TER VERKRIJGING VAN DE GRAAD VAN
MASTER IN DE FYSICA

Samenvatting

Netwerken zijn alomtegenwoordig: de verbindingen tussen luchthavens, ‘the internet of things’ en zelfs het metabolisme in een cel hebben allemaal de structuur van een netwerk van nodes en links. Om deze netwerken te bestuderen, worden in hoofdstuk 1 enkele fundamentele eigenschappen van netwerken geïntroduceerd. Dit om twee belangrijke types netwerken te bespreken: het Bernoulli netwerk en het schaalvrij netwerk. Vooral het schaalvrij netwerk dankt haar populariteit aan haar goede overeenkomst met sociale en economische netwerken.

De mean-field aanname die nodig is in de beschrijving van het schaalvrij netwerk door het Barabási-Albert model, wordt in hoofdstuk 2 omzeild door het opstellen van een master equation en deze op te lossen via de genererende functiemethode. Dit wordt gedaan voor zowel random als preferentiële groei. Uit de kansverdeling die hieruit volgt, worden de verwachtingswaarde en variantie afgeleid. In het preferentiële geval komen de resultaten hiervan, waar mogelijk, overeen met die van het Barabási-Albert model.

Simulaties worden vervolgens gebruikt om deze theoretische resultaten te bevestigen. Voor nodes die vroeg geboren worden, is er een minder goede overeenkomst. Deze fout kan voor zeer grote netwerken echter verwaarloosd worden.

Nadat de groei van een netwerk besproken werd in hoofdstuk 2, kijkt hoofdstuk 3 naar de impact van het verplaatsen en verwijderen van links. Daartoe wordt eerst gekeken naar een algoritme dat het aantal nodes en links onveranderd laat, maar links willekeurig verplaatst. Hiervoor wordt verwacht dat het netwerk uiteindelijk naar een Bernoulli netwerk evolueert. Na het opstellen van een master equation voor de kans van een link tussen twee nodes, wordt dit inderdaad bevestigd. Meer nog, er kan een tijdconstante afgeleid worden die het exponentieel verval naar het Bernoulli netwerk karakteriseert.

Ook dit resultaat wordt numeriek bevestigd. Bovendien wordt er gekeken naar het verloop van de degree distributie en de clustering coëfficiënt. Deze laatste evolueert sneller dan voorspeld naar de limietwaarde, doordat ze voornamelijk bepaald wordt door nodes van lage degree.

Er wordt ook gekeken naar de simulatie van een algoritme van preferentiële verplaatsing, waar de clustering coëfficiënt logaritmisch lijkt te stijgen. De degree distributie wordt smaller en schuift traag op naar hogere degrees.

Ten slotte wordt in de laatste sectie van hoofdstuk 3 gekeken naar een gemodificeerd groeiproses waar ook een aantal links verloren gaan. Van vier mogelijke combinaties van het proces (preferentiële of random groei met preferentiële of random verwijdering) worden de gevolgen voor de degree distributie beschreven. Hier zijn uitgebreidere simulaties echter aanbevolen.

Abstract

Networks permeate our world: airport connections, the internet of things, even cell metabolisms all have the structure of a network, consisting of nodes and links. To study these networks, chapter 1 introduces some fundamental properties of networks, to eventually introduce two important network types: Bernoulli and scale free networks. The scale free network in particular has gained popularity due to its good correspondence with social and economic networks.

The mean-field approximation in the description of the scale free network by the Barabási-Albert model, is omitted in chapter 2. Instead, a master equation is formulated for both random and preferential growth and solved using the generating function method. The resulting probability distribution leads to an expectation value and variance. The results for preferential growth agree, where possible, with the Barabási-Albert model.

Subsequent simulations serve to confirm these theoretical results. A systematic error is found for the oldest nodes, which can be neglected for very large networks.

After discussing network growth in chapter 2, chapter 3 describes the consequences of moving and removing links. First of all, the number of links and nodes is kept constant, while links are randomly relocated. It is expected that the network will eventually evolve to a Bernoulli network. Solving the master equation for the probability of a link between two nodes indeed confirms this prediction. The result of the master equation even predicts a time constant that characterizes the decay to a Bernoulli network.

Again, simulations of this algorithm confirm these results. The degree distribution and clustering coefficient are also observed. The clustering coefficient shows a faster decay than predicted, being mostly determined by nodes of smaller degree.

Preferential relocation is also simulated, for which the clustering coefficient seems to increase as a logarithmic function. The degree distribution slowly shifts towards larger degrees and becomes narrower.

Finally, the last section of chapter 3 modifies the earlier described growth processes of chapter 2 with the removal of links. The effect of four algorithms is observed in the degree distribution, suggesting further simulations are necessary.

Voorwoord

Een masterthesis behoort tot één van de vruchten die voortkomen uit een masteropleiding. Daarnaast vormt het ook het slot van een hoofdstuk in het leven van de student die de wereld in trekt. Zo'n hoofdstuk wordt niet door één persoon geschreven, maar door iedereen in zijn naaste omgeving. Daarom lijkt het me terecht deze mensen hier te bedanken.

De eerste dankbetuiging gaat uit naar de promotor van deze thesis: Prof. dr. Jacques Tempere. Dit voor zijn grote steun, geduld, kennis en enthousiasme. Aan de hand van een heldere uiteenzetting wist hij mij steeds op het juiste pad te brengen.

Vervolgens wil ik graag de gehele TQC-groep aan de Universiteit Antwerpen bedanken voor de leuke sfeer tijdens het maken van deze thesis. Ik kreeg er een plek in een kantoor om ijverig te werken en de kans om mijn thesis te presenteren tijdens een Bami-lezing. Bovendien zorgde iedereen er voor vele leuke, interessante discussies en middagpauzes (Still can't work the perpetuator, though ...).

Tot slot wil ik mijn familie en vrienden bedanken voor alle steun. In het bijzonder mijn ouders en zussen voor de warme thuis, mijn vriendin voor alle liefde, rust en steun die ze me biedt en mijn vrienden die van tijd tot tijd voor de nodige ontspanning zorgen.

Inhoudsopgave

Samenvatting	i
Abstract	iii
Voorwoord	v
1 Netwerk Theorie	1
1.1 Fundamentele begrippen	1
1.1.1 Verbindingsmatrix	1
1.1.2 Degree & degree distributie	1
1.1.3 Gemiddelde padlengte & diameter	2
1.1.4 Clustering coëfficiënt	3
1.1.5 Andere eigenschappen	3
1.2 Erdős-Rényi model	4
1.3 Barabási-Albert model	5
1.3.1 Preferentiële groei	5
1.3.2 Random groei	8
1.4 Doelstellingen	9
2 Groeiende netwerken	11
2.1 Random groei	11
2.1.1 Master equation	11
2.1.2 Normering, verwachtingswaarde en variantie	14
2.1.3 Wachtijd T_k	15
2.1.4 Voorbeeld	16
2.2 Preferentiële groei	18
2.2.1 Kansverdeling voor de degree	18
2.2.2 Normering, verwachtingswaarde en variantie	20
2.3 Numerieke controle	22
2.3.1 Verwachtingswaarde en standaardafwijking	22
2.3.2 Kansverdeling	23

3	Link herschikking	25
3.1	Willekeurige verplaatsing	25
3.2	Simulaties willekeurige verplaatsing	28
3.2.1	De kans $p_{ij}(t)$	28
3.2.2	Tijdsverloop van de degree distributie	29
3.2.3	Clustering coëfficiënt	29
3.3	Simulatie met bias	30
3.4	Link verwijdering bij toevoeging van nodes	32
3.4.1	Random groei	32
3.4.2	Preferentiële groei	33
3.5	Conclusies link herschikking	35
4	Conclusie	37
	Bijlagen	39
A	Genererende functiemethode	41
A.1	Formele machtreeks	41
A.2	Oplossen van recursierelaties	42
A.3	Bepalen van momenten	43
B	Mathematica implementatie	45
B.1	Random groei	45
B.2	Preferentiële groei	45
B.3	Random link verwijdering	46
B.4	Preferentiële link verwijdering	46
B.5	Voorbeeld simulatie	46
B.6	Willekeurige verplaatsing	47
B.7	Verplaatsing met bias	47
B.8	Verplaatsing simulatie	48
	Bibliografie	49

Hoofdstuk 1

Netwerk Theorie

1.1 Fundamentele begrippen

Links die nodes verbinden, dat is de essentie van een netwerk. Deze simpele omschrijving bevat tegelijkertijd een grote kracht: wat de nodes en links voorstellen is niet gespecificeerd. Voor alle systemen met een onderliggend netwerk kunnen dus dezelfde eigenschappen gedefinieerd worden én (met enige voorzichtigheid) vergeleken worden. Daarom kan netwerktheorie in verscheidene onderzoeksgebieden gebruikt worden, van economie tot epidemiologie.¹

1.1.1 Verbindingsmatrix

Een netwerk kan mathematisch beschreven worden aan de hand van een matrix. Deze *verbindingsmatrix* G heeft elementen g_{ij} die gelijk aan 1 zijn wanneer er een link bestaat tussen nodes i en j . Indien niet, is het element gelijk aan 0. Bijgevolg is dit een $N_n \times N_n$ vierkante matrix, waar N_n het aantal nodes in het netwerk is (zie figuur 1.1) [8].

Aangezien deze thesis niet kijkt naar de richting van een link (met andere woorden, een link van i naar j is ook een link van j naar i), geldt dat $g_{ij} = g_{ji}$. Ofwel, de verbindingsmatrix G is symmetrisch. De diagonaalelementen zullen 0 zijn ($g_{ii} = 0$), doordat links van nodes met zichzelf niet in rekening worden gebracht.

1.1.2 Degree & degree distributie

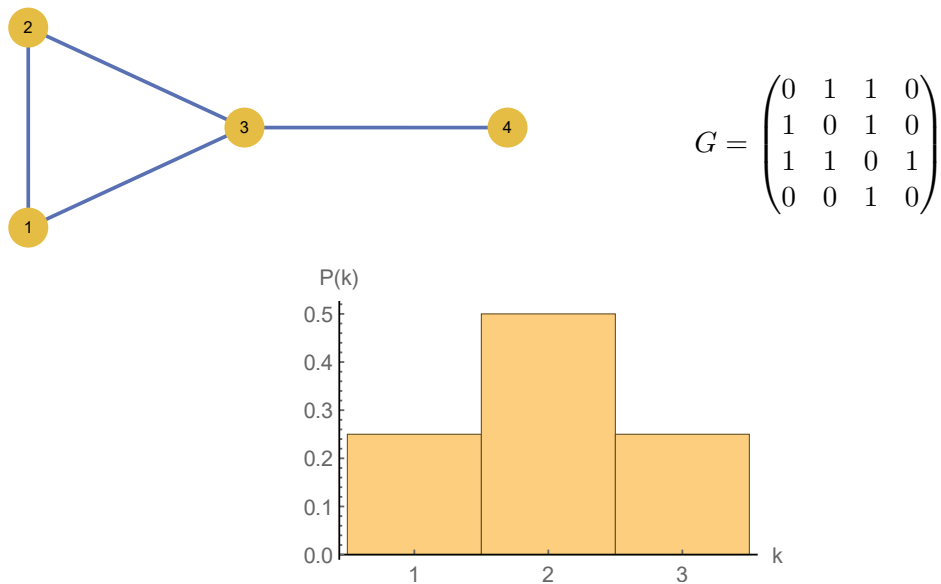
Een basiseigenschap van een node is het aantal links, of het aantal burens, waarmee ze verbonden is. Verder genoteerd als k , wordt dit de *degree* (of graad) genoemd. In termen van de verbindingsmatrix G , kan dit voor node i geschreven worden als

$$k_i = \sum_{j=1}^{N_n} g_{ij}. \quad (1.1)$$

¹Een voorbeeld van een toepassing op vlak van epidemiologie is het resultaat van Small & Chi (2004) [15]. Zij wisten reeds geobserveerde opflakeringen in een epidemie te reproduceren in een simulatie, waarbij de ziekte werd verspreid over een schaalvrij netwerk. Een ander resultaat van Alemi, Bierbaum, Myers & Sethna (2015)[1] toont de verspreiding van het fictieve zombie-virus doorheen de Verenigde Staten.

Bovenstaande vergelijking is in wezen het sommeren over de kolommen van de verbindingsmatrix. Aangezien G symmetrisch is, kan er ook gesommeerd worden over de rijen van de verbindingsmatrix [8].

Wanneer alle degrees in een netwerk bepaald zijn, kan er geteld worden hoe vaak iedere degree voorkomt (figuur 1.1). Dit wordt genoteerd als de *degree distributie* $P(k)$ die de frequentie van iedere degree k weergeeft (met normering $\sum_k P(k) = 1$). Op die manier geeft de degree distributie de kans weer dat een willekeurig aangeduide node degree k heeft.



Figuur 1.1: *Linksboven:* Een voorbeeld van een klein netwerk. Node 4 heeft bijvoorbeeld degree 1. *Rechtsboven:* De verbindingsmatrix van het voorbeeldnetwerk. Deze is duidelijk symmetrisch. *Onder:* De degree distributie van het netwerk, waarbij de frequentie $P(k)$ in functie van de degree k getoond wordt.

1.1.3 Gemiddelde padlengte & diameter

In netwerken kunnen paden en wandelingen worden afgelegd, waarbij er vanaf de ene node wordt getracht naar de andere te lopen via tussenliggende links. In een wandeling is het toegelaten om tussenliggende links meerdere keren te bewandelen, op een pad is dit niet toegelaten.

De afstand l_{ij} tussen twee nodes wordt dan gedefinieerd als de lengte van het kortste pad tussen deze nodes [8]. De afstanden tussen alle nodes in het netwerk kunnen vervolgens uitgemiddeld worden om de gemiddelde padlengte l te bekomen. Deze lengte geeft aan hoever twee nodes doorgaans van elkaar verwijderd zijn.

De grootste afstand tussen twee nodes wordt de diameter van het netwerk genoemd [8]. Deze geeft een bovengrens voor de afstand tussen twee nodes, maar kan een vertekend beeld geven wanneer er enkele nodes ver buiten het netwerk liggen.

Indien er een deel van het netwerk los hangt van de rest van het netwerk, zal de diameter ∞ zijn. Dit resultaat kan bij verdere berekeningen moeilijkheden geven, bijvoorbeeld bij het berekenen van de gemiddelde padlengte. Ook informatie over de afzonderlijke delen gaat dan verloren.

1.1.4 Clustering coëfficiënt

De degree distributie zegt niets over de fijne structuren die in het netwerk aanwezig kunnen zijn. Bijvoorbeeld, in het netwerk van sociale contacten vormen er zich snel vriendengroepen waarin iedereen elkaar kent. In het netwerk ontstaan er dus clusters die volledig verbonden zijn, terwijl dit in de degree distributie niet meteen hoeft op te vallen.

Kijk daarom naar alle combinaties van drie nodes waartussen ten minste twee links aanwezig zijn. De fractie hiervan waar drie links aanwezig zijn, geeft de clustering coëfficiënt van het netwerk.

Mathematisch is dit te berekenen aan de hand van

$$Cl = \frac{\sum_{i,j,k} g_{ij}g_{ik}g_{jk}}{\sum_{i,j,k \neq j} g_{ij}g_{ik}}, \quad (1.2)$$

waar g_{nm} een element is van G [8]. De teller $g_{ij}g_{ik}g_{jk}$ zal enkel 1 zijn indien de nodes i, j en k met elkaar verbonden zijn. De termen waar $j = i$, $k = i$ en $k = j$ zullen wegvallen, aangezien de diagonaalelementen 0 zijn. De noemer $g_{ij}g_{ik}$ is enkel 1 indien de links $i \leftrightarrow j$ en $j \leftrightarrow k$ aanwezig zijn. De derde (mogelijke) link $i \leftrightarrow k$ hoeft dus niet aanwezig te zijn. Hier mogen nodes j en k wel niet aan elkaar gelijk zijn, aangezien er dan g_{ij}^2 verschijnt, wat betekent dat er geen drie verschillende links bekeken worden. Net iets visueler kan dit ook voorgesteld worden als:

$$Cl = \frac{3 \cdot \# \text{ } \nabla}{\# \text{ } \curvearrowright}$$

De factor 3 verschijnt doordat er 3 unieke paden van lengte 2 (noemer) bestaan in een driehoek (teller).

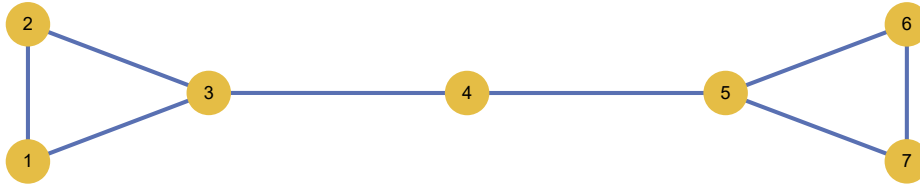
Als voorbeeld kan de clustering coëfficiënt van het netwerk in figuur 1.1 berekend worden. Het aantal driehoeken in het netwerk is 1, terwijl er 5 unieke paden van lengte 2 zijn. Dit geeft een clustering coëfficiënt van $\frac{3}{5}$.

1.1.5 Andere eigenschappen

De degree van een node is niet de enige eigenschap die bepaalt hoe belangrijk een node is voor de structuur van een netwerk. Zo is duidelijk te zien in figuur 1.2 dat node 4 een zeer centrale rol speelt in het bij elkaar houden van het netwerk, hoewel haar degree van 2 niet de hoogste is in het netwerk. Er zijn dus duidelijk extra *centraliteitseigenschappen* nodig voor de juiste klassificatie van nodes binnen het netwerk.

Eén van deze centraliteitseigenschappen is de *nabijheidscentraliteit* (closeness centrality), welke uitdrukt hoe belangrijk deze node is in het verbinden van andere nodes. Een andere eigenschap is de *verbindingscentraliteit* (betweenness centrality), waar gekeken wordt hoe vaak de node op de kortste paden tussen twee willekeurige nodes ligt [8].

De nodige begrippen voor deze thesis zijn bij deze geïntroduceerd. Er valt natuurlijk nog veel meer te vertellen, maar dat zou deze thesis onnodig doen afwijken van haar doelstellingen. Een grondige omlijsting van netwerk theorie kan gevonden worden in het boek *Social and Economic Networks* van Matthew O. Jackson [8]. Dit boek werd gebruikt als basis voor de inleidende begrippen van deze sectie en de omschrijving van twee belangrijke types netwerken in de volgende secties (1.2 en 1.3).



Figuur 1.2: Een voorbeeld van een netwerk waar de nodes met hoge degree niet de belangrijkste zijn. Node 4 heeft degree 2, terwijl nodes 3 en 5 degree 3 hebben. Zowel de nabijheidscentraliteit (0.6) en verbindingscentraliteit (9) van node 4 zijn de hoogste in het netwerk. De positie van een node kan dus belangrijker zijn dan haar degree.

1.2 Erdős-Rényi model

Het bekijken van netwerken die at random gegenereerd worden, maakt het mogelijk om algemene uitspraken te doen over de eigenschappen van dergelijke netwerken. Ook empirische netwerken kunnen verondersteld worden zich te manifesteren aan de hand van stochastische processen. Zo bespraken Erdős & Rényi (1959) [4] een model waarbij een vast aantal links willekeurig verdeeld wordt over een vast aantal nodes. Een tweede model, geïntroduceerd door Gilbert (1959) [7], start evenwel met een vast aantal nodes N_n , maar geeft iedere mogelijke link een kans p tot bestaan.

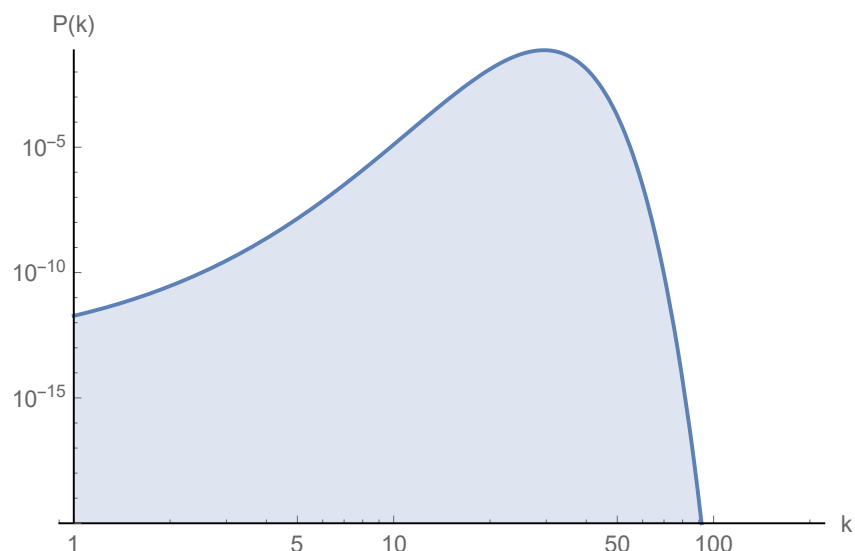
Aangezien het tweede model de stochastische aard van de vorming van links expliciet omschrijft, zal dit model (zoals doorgaans in de literatuur) bedoeld worden wanneer er verder over het Erdős-Rényi model wordt gesproken. Een netwerk gegenereerd door dit model wordt ook een Bernoulli netwerk genoemd, naar de onafhankelijke binaire kansprocessen die het netwerk vormen.

Het zoeken naar de degree distributie kan vertaald worden naar het zoeken naar de kans dat een willekeurige node exact k links heeft. Aangezien er $N_n - 1$ andere nodes zijn om uit te kiezen, wordt dit de kans dat er verbindingen zijn met k andere nodes (p^k) en géén verbindingen met $(N_n - 1) - k$ nodes ($(1 - p)^{N_n - 1 - k}$). Er moet bovendien rekening worden gehouden met alle mogelijke combinaties van k nodes die uit $N_n - 1$ nodes gekozen kunnen worden. Daarom wordt de degree distributie uiteindelijk een binomiaalverdeling[8]:

$$P(k) = \binom{N_n - 1}{k} p^k (1 - p)^{N_n - 1 - k}. \quad (1.3)$$

Het achterhalen van de clustering coëfficiënt wordt geholpen door de ongecorreleerde aanwezigheid van de links. Dat wil zeggen, de aanwezigheid van een link hangt niet af van het bestaan van andere links. De clustering coëfficiënt kan ook worden geïnterpreteerd als de kans dat een link in een collectie van drie nodes aanwezig is, gegeven dat de andere twee mogelijke links reeds aanwezig zijn. Bijgevolg zal de clustering coëfficiënt gelijk zijn aan p in een Bernoulli netwerk[12].

De nodes in een Bernoulli netwerk liggen doorgaans niet ver van elkaar verwijderd. De gemiddelde padlengte van een Bernoulli netwerk schaalte zelfs volgens $\ln N_n$ [6], wat de vereiste is voor de ‘small-world’ eigenschap. Deze term werd in de context van netwerken in 1967 aangehaald door Milgram [10] in zijn experiment, waarin brieven werden doorgestuurd tot de juiste ontvanger bereikt werd.



Figuur 1.3: Plot van de degree distributie $P(k)$ t.o.v. de degree k in een Bernoulli netwerk (vergelijking 1.3) met $N = 1000$ en $p = 0.03$. De aanwezigheid van grote degrees wordt, voor grote netwerken, exponentieel onderdrukt.

1.3 Barabási-Albert model

1.3.1 Preferentiële groei

Empirische data volgt echter niet het stramien van Bernoulli netwerken (zie figuur 1.4). Al in 1965 vond Price [14] een ‘fat-tail’ distributie (of machtswet) in het netwerk van citaties tussen wetenschappelijke artikels. Met andere woorden, de degree distributie had de vorm

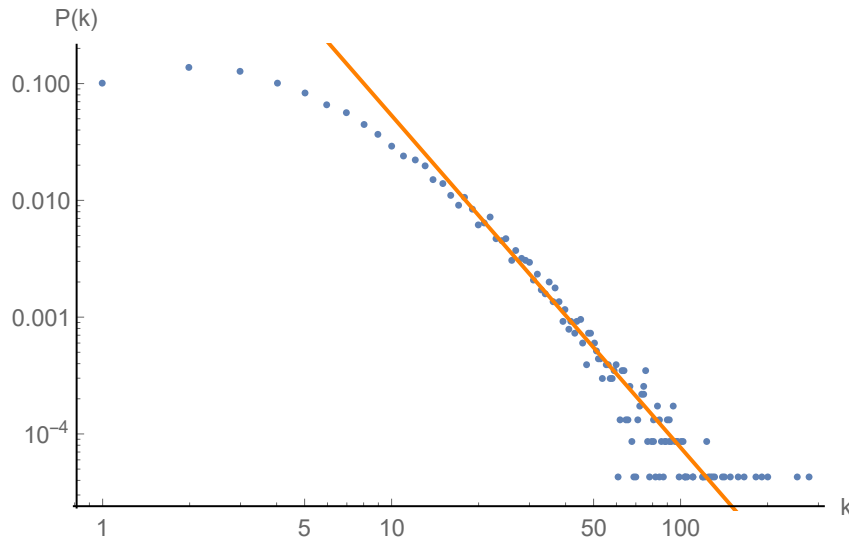
$$P(k) \propto k^{-\gamma}. \quad (1.4)$$

Hetzelfde verband werd later teruggevonden in een waaier aan netwerken [11, 16, 5, 9] met exponent γ typisch tussen 2 en 3. Netwerken met dergelijke degree distributies worden *schaalvrije netwerken* genoemd. Bovendien werd de clustering coëfficiënt van deze netwerken relatief hoog bevonden, zeker in vergelijking met wat te verwachten is voor een Bernoulli netwerk [11]. Ook de gemiddelde padlengte (en de diameter) van deze netwerken werden zeer klein bevonden [9]. Dit wil zeggen dat het om small-world netwerken gaat, met een schaalvrije degree distributie. Price introduceerde in 1976 [13] het begrip ‘cumulative advantage’ (cumulatief voordeel), waarmee hij de geobserveerde machtswetten kon voorspellen. Pas 23 jaar later verkreeg het model populariteit dankzij Barabási en Albert [2] die het onderliggende groeiproces *preferentiële aanhechting* (preferential attachment) doopten. Hun BA-model steunt op 3 aannames:

1. Netwerken groeien doorheen de tijd. Op iedere tijdstap wordt er één node toegevoegd aan het netwerk. Deze node wordt gelabeld met het tijdstip van haar geboorte.
2. Iedere nieuwe node verbindt zich met een vast aantal links m aan het reeds bestaande netwerk.

3. De links waarmee de nieuwe node zich verbindt, worden volgens preferentiële aanhechting verspreid over het netwerk. Dit wil zeggen, de kans p_i dat een node i in het netwerk op tijdstip t een nieuwe link verkrijgt, is evenredig met haar degree k_i :

$$p_i = \frac{k_i}{\sum_{j=1}^t k_j} \quad (1.5)$$



Figuur 1.4: De degree distributie van het netwerk van wetenschappelijke collaboraties rond gecondenseerde materie, gaande van 1993 tot 2003[9]. Dit netwerk bestaat uit auteurs als nodes die verbonden worden indien zij hebben samengewerkt aan een publicatie. De oranje fit toont een machtwet ($\gamma = 2.84$) bij grote degrees, die niet overeenstemt met een Bernoulli netwerk. Het netwerk heeft dus een schaalvrij karakter. De afwijking bij kleine degree is een gevolg van de groei van het netwerk die niet volledig via preferentiële aanhechting verloopt. Bij zeer grote degrees komt de eindigheid van het netwerk tevoorschijn in de vorm van telruis, waardoor de afwijkingen van de schaalvrije curve groter worden. De log-log schaal maakt het eenvoudiger schaalvrije netwerken te herkennen aan hun lineaire curve.

Verder zijn er nog twee benaderingen nodig in het BA-model. De eerste benadering (mean-field) kijkt enkel naar het verwachte gedrag van het netwerk. Er wordt niet gekeken naar alle mogelijke uitkomsten van het groeiproces, alleen de verwachte uitkomst.

De tweede benadering veronderstelt dat het netwerk reeds voldoende gegroeid is, zodat de tijdsevolutie als continu kan worden beschouwd. Een netwerk dat al lange tijd heeft gehad om te groeien, zal inderdaad nog weinig invloed ondervinden van één enkele tijdstap waarin één node wordt toegevoegd.

Dit alles mee in rekening gebracht, zorgt voor een uitdrukking voor de verandering van de degree k_i van node i op tijdstip t :

$$\frac{dk_i(t)}{dt} = m \frac{k_i(t)}{\sum_{j=1}^t k_j(t)}$$

De totale som van alle degrees in het netwerk is gelijk aan twee maal het aantal links, aangezien iedere link twee keer wordt geteld. Het aantal links in dit model wordt gegeven door $m \cdot t$, doordat op ieder tijdstip m links worden bijgevoegd. Bovenstaande uitdrukking wordt dan

$$\frac{dk_i(t)}{dt} = \frac{k_i(t)}{2t}. \quad (1.6)$$

De differentiaalvergelijking (met randvoorwaarde $k_i(i) = m$, de node heeft m links bij haar geboorte) is eenvoudig op te lossen en geeft als resultaat

$$k_i(t) = m\sqrt{\frac{t}{i}}. \quad (1.7)$$

Ieder label van een node komt dus op een welbepaald tijdstip overeen met een unieke degree (en vice versa). Die vaststelling is een grote hulp in de zoektocht naar de degree distributie, welke opgevat kan worden als de frequentie distributie van de degrees van alle nodes. De frequentie distributie van de degrees kan op haar beurt afgeleid worden uit de cumulatieve distributiefunctie (CDF). De CDF $F(k)$ geeft de fractie nodes met een degree kleiner dan k , wat gegeven wordt door

$$F_t(k) = \frac{N_n - N(k)}{N_n},$$

waar $N(k)$ het aantal nodes is met degree *groter* of gelijk aan k en N_n het totaal aantal nodes, gegeven door t .

Stel nu dat node 10 degree k heeft op tijd t . De vraag hoeveel nodes een degree groter of gelijk aan k hebben kan dan geïllustreerd worden door een rangschikking van alle nodes te maken volgens hun degree:

$$\begin{array}{l} \text{Degree:} \\ \text{Node:} \end{array} \quad \begin{array}{cccccccccccccc} & & & & & \overbrace{\hspace{2cm}} & & & & & \overbrace{\hspace{2cm}} & & & & & \\ & & & & & >k & & & & & k & & & & & <k \\ \dots & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 & 13 & 14 & \dots \end{array}$$

Vergelijking 1.7 zegt inderdaad dat iedere node geboren na node 10 een kleinere degree dan k zal hebben (want $\sqrt{\frac{1}{i}}$ is een monotoon dalende functie). Nodes 1 – 10 hebben een degree $\geq k$. Het zoeken van dit aantal komt dus overeen met het zoeken naar de node die exact degree k heeft. M.a.w. $N(k) = i(k)$, wat voor de CDF betekent dat

$$F_t(k) = \frac{t - i(k)}{t}. \quad (1.8)$$

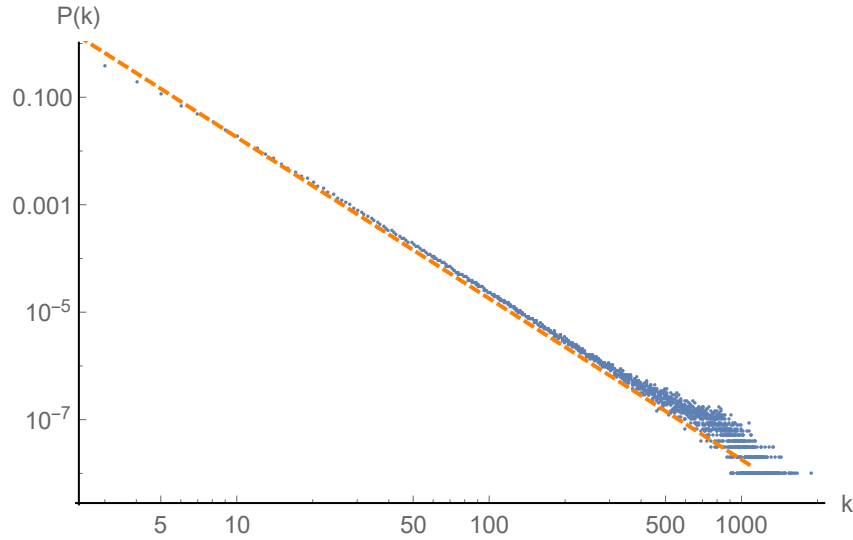
Gebruik makend van vergelijking 1.7 wordt dit

$$\begin{aligned} i(k) &= tm^2k^{-2} \\ F_t(k) &= 1 - m^2k^{-2} \end{aligned} \quad (1.9)$$

Aangezien de frequentie distributie gelijk is aan de afgeleide van de CDF naar k , is de degree distributie van een schaalvrij netwerk in het BA model

$$P(k) = 2m^2k^{-3}. \quad (1.10)$$

Een eenvoudige integratie over k met ondergrens m en bovengrens $+\infty$ toont dat bovenstaand resultaat inderdaad genormeerd is. Bovendien wordt voorspeld dat $\gamma = 3$ voor grote netwerken. Figuur 1.5 toont de uitmiddeling van 1000 gegenereerde schaalvrije netwerken van 100 000 nodes, vergeleken met vergelijking 1.10. Hier wordt duidelijk dat het BA mean-field model beter geldt voor zeer grote netwerken.



Figuur 1.5: De degree distributie $P(k)$ t.o.v. de degree k van 1000 gegenereerde schaalvrije netwerken van 100 000 nodes volgens het BA-model (blauw), vergeleken met het resultaat van figuur 1.10 (oranje onderbroken lijn). De helling van het theoretische resultaat is bijna goed, al is er een klein verschil merkbaar. Aan de continue tijdsbenadering is dus nog niet volledig voldaan.

1.3.2 Random groei

Een netwerk kan ook zonder preferentiële aanhechting gegroeid worden: de nieuwe node verdeelt dan haar m links willekeurig over het netwerk. De verandering van de degree van node i op tijdstip t wordt dan

$$\frac{dk_i(t)}{dt} = \frac{m}{t},$$

na het toepassen van de mean-field en continue-tijdsbenadering. Bovenstaande differentiaalvergelijking heeft dezelfde randvoorwaarde als voorheen ($k_i(i) = m$). De oplossing wordt dan

$$k_i(t) = m + m \ln\left(\frac{t}{i}\right). \quad (1.11)$$

Analoog aan de afleiding mét preferentiële aanhechting is ook hier de CDF gegeven door

$$F_t(k) = \frac{t - i(k)}{t}, \quad (1.12)$$

met

$$i(k) = t \exp\left(-\frac{k - m}{m}\right). \quad (1.13)$$

Dit geeft als CDF

$$F_t(k) = 1 - \exp\left(-\frac{k - m}{m}\right). \quad (1.14)$$

De degree distributie van een groeiend netwerk zonder preferentiële aanhechting wordt dan exponentieel dalend in de degree:

$$P(k) = \frac{1}{m} \exp\left(-\frac{k - m}{m}\right) \quad (1.15)$$

Er is dus duidelijk nog een verschil tussen de willekeurige groei hier en de constructie van een Bernoulli netwerk. Dit resultaat zal later in de thesis ook herhaald worden wanneer er verder gekeken wordt naar de groei van netwerken.

1.4 Doelstellingen

De vorige sectie introduceerde de netwerk theorie nodig voor de rest van deze thesis. Verder werden twee zeer fundamentele netwerken ingeleid: het Bernoulli en schaalvrije netwerk. Hier werden ook de ingevoerde benaderingen besproken die tot een analytisch resultaat leidden.

Hoofdstuk 2 behandelt een poging tot het elimineren van één van de benaderingen in het Barabási-Albert model: de mean-field benadering. Door middel van een master equation en de genererende functiemethode wordt hier de kansverdeling voor de degree afgeleid voor een random en preferentieel groeiend netwerk. Met behulp van deze kansverdeling worden de verwachtingswaarde, variante en kansverdeling voor de wachttijd bepaald. Deze resultaten worden vervolgens vergeleken met numerieke simulaties van schaalvrije netwerken.

De gedachte dat fysieke netwerken voortdurend evolueren, leidt tot het derde hoofdstuk, waar bekeken wordt wat de invloed is van een algoritme dat speelt met de links van een netwerk. Een theoretische behandeling van het willekeurig verplaatsen van links wordt hier opgesteld aan de hand van de genererende functiemethode, welke gecontroleerd en besproken wordt aan de hand van numerieke simulaties. Daarbij worden onder andere de evolutie van de degree distributie en clustering coëfficiënt geobserveerd. Aanvullende simulaties bekijken ook de invloed van algoritmes waarbij een bias werd ingesteld in het relocatie proces, aangezien deze bias het afleiden van een analytisch resultaat bemoeilijkt. Ook wordt er gekeken naar het effect van een aangepast groeiproces, waar steeds een fractie links verwijderd wordt.

Hoofdstuk 2

Groeiende netwerken

Hoofdstuk 1 beschreef alvast het groeiende schaalvrij netwerk in de mean-field en continue tijdsbenadering, zoals reeds beschreven in het BA-model. De continue tijdsbenadering werd ingevoerd om gebruik te kunnen maken van analytische operaties op continue functies, in plaats van discrete tijdstappen te bekijken. De mean-field benadering werd gebruikt om over te gaan op het verwachte gedrag van een node: in plaats van iedere mogelijke realisatie van het kansproces te bekijken, werd er gekeken naar de verwachte uitkomst. Dit vereenvoudigt de berekening, maar zorgt ook voor een verlies van informatie over de mogelijke fluctuaties rond het verwachte gedrag.

Die fluctuaties op het reeds gekende groeiproces worden in dit hoofdstuk op een theoretische wijze bekeken door een master equation op te stellen. Daarbij wordt de degree op een bepaald tijdstip opgevat als een toestand waarin de node zich bevindt. De overgangswaarschijnlijkheden tussen toestanden leggen bijgevolg het groeiproces vast.

2.1 Random groei

2.1.1 Master equation

Hier wordt eerst de methode getoond voor een netwerk dat groeit *zonder* preferentiële aanhechting. De links van nieuwe nodes worden dus willekeurig verspreid over het netwerk.

De degree van een node op een bepaald tijdstip kan opgevat worden als een telling van het aantal gebeurtenissen waarbij de node een link wint, sinds de geboorte van desbetreffende node. Dergelijke gebeurtenissen zijn (voor een random groei) onafhankelijk van elkaar, waardoor er verwacht kan worden dat de kansverdeling van de degree gegeven is door een Poissonverdeling. In wat volgt wordt er getracht de Poissonverdeling inderdaad aan te tonen voor de degree van een gekozen node i op tijdstip t .

Als node i op tijdstip $t \geq i$ zich in een toestand bevindt met degree $k > 0$, heeft ze een bepaalde kans $0 \leq \lambda(t) \leq 1$ om na een tijdstap Δt naar een toestand over te gaan met degree $k + 1$. Het is in dit groeiproces onmogelijk om te dalen in degree.

Dit kan genoteerd worden als de volgende overgangswaarschijnlijkheden:

$$P_i(k, t \rightarrow k - 1, t + \Delta t) = 0 \quad (2.1)$$

$$P_i(k, t \rightarrow k + 1, t + \Delta t) = \lambda(t)\Delta t \quad (2.2)$$

$$P_i(k, t \rightarrow k, t + \Delta t) = 1 - \lambda(t)\Delta t \quad (2.3)$$

De kans $\lambda(t)$ wordt in de context van een nucleair verval (een vertrouwd Poisson proces) de *activiteit* genoemd. Aangezien λ voor dit proces beschrijft hoeveel links er iedere tijdstap gemiddeld bijkomen, is 'activiteit' in de context van een groeiproces een passende term die verder gebruikt zal worden.

Voor de random groei werd de activiteit hier enkel afhankelijk gemaakt van de tijd t , de grootte van het netwerk. Zo is λ gelijk voor iedere node en wordt er geen bias geïntroduceerd in het groeiproces. Dit gebeurt wél bij preferentiële aanhechting: de activiteit is er afhankelijk van de degree, wat later behandeld wordt in sectie 2.2.

De eerder beschreven overgangswaarschijnlijkheden kunnen gebruikt worden om de kans uit te rekenen dat node i , geboren op tijdstip i met degree $m > 0$, na een tijd $t + \Delta t$ degree k heeft. Deze kans is gelijk aan de kans dat de node een degree $k - 1$ bereikt op tijd t en vervolgens in de tijdstap Δt een nieuwe link krijgt, *plus* de kans dat de node reeds op tijd t een degree k heeft en vervolgens haar degree in de tijdstap Δt onveranderd blijft:

$$\begin{aligned} P_i(m, i \rightarrow k, t + \Delta t) &= P_i(m, i \rightarrow k - 1, t)P_i(k - 1, t \rightarrow k, t + \Delta t) \\ &\quad + P_i(m, i \rightarrow k, t)P_i(k, t \rightarrow k, t + \Delta t) \end{aligned}$$

Na invullen van vergelijkingen 2.2 en 2.3, wordt dit

$$P_i(m, i \rightarrow k, t + \Delta t) = P_i(m, i \rightarrow k - 1, t)\lambda(t)\Delta t + P_i(m, i \rightarrow k, t)(1 - \lambda(t)\Delta t).$$

Dit kan geschreven worden als

$$\frac{P_i(m, i \rightarrow k, t + \Delta t) - P_i(m, i \rightarrow k, t)}{\Delta t} = \lambda(t) [P_i(m, i \rightarrow k - 1, t) - P_i(m, i \rightarrow k, t)].$$

Wanneer er genoeg tijd is verlopen, kan Δt als infinitesimaal beschouwd worden. Met andere woorden, de continue tijdsbenadering wordt hier toegepast, zodat bovenstaande vergelijking herleid wordt naar

$$\frac{\partial P_i(m, i \rightarrow k, t)}{\partial t} = \lambda(t) [P_i(m, i \rightarrow k - 1, t) - P_i(m, i \rightarrow k, t)]. \quad (2.4)$$

Deze partiële differentiaalvergelijking heeft bovendien als beginvoorwaarde dat een node wordt toegevoegd op tijdstip i met m links, oftewel

$$P_i(m, i \rightarrow k, i) = \delta_{m,k}. \quad (2.5)$$

Door de term $P_i(m, i \rightarrow k - 1, t)$ lijkt vergelijking 2.4 niet meteen oplosbaar. De reeds gekende genererende functiemethode biedt een mogelijkheid tot het oplossen van dit probleem. Bijlage A geeft een inleiding tot deze handige techniek en verdere referenties voor de geïnteresseerde lezer.

In dit probleem wordt de genererende functie gedefinieerd als

$$G_i(s, t) = \sum_{k=m}^{\infty} s^k P_i(m, i \rightarrow k, t) \quad (2.6)$$

Er is geen mogelijkheid dat eender welke node links verliest in dit model, zodat $P_i(m, i \rightarrow k, t) = 0$ voor $k < m$. Daarom begint deze som niet vanaf 0, maar vanaf m . Na vermenigvuldiging van 2.4 met s^k en sommatie over k , verschijnt er

$$\frac{\partial}{\partial t} \sum_{k=m}^{\infty} s^k P_i(m, i \rightarrow k, t) = \lambda(t) \left[\sum_{k=m}^{\infty} s^k P_i(m, i \rightarrow k - 1, t) - \sum_{k=m}^{\infty} s^k P_i(m, i \rightarrow k, t) \right]$$

De eerste term in het rechterlid kan herschreven worden als

$$\begin{aligned} \sum_{k=m}^{\infty} s^k P_i(m, i \rightarrow k-1, t) &= s \sum_{k=m}^{\infty} s^{k-1} P_i(m, i \rightarrow k-1, t) \\ &= s \cdot s^{m-1} P_i(m, i \rightarrow m-1, t) \\ &\quad + s \sum_{k=m+1}^{\infty} s^{k-1} P_i(m, i \rightarrow k-1, t), \end{aligned}$$

of na substitutie van $k' = k - 1$

$$\begin{aligned} \sum_{k=m}^{\infty} s^k P_i(m, i \rightarrow k-1, t) &= s \sum_{k'=m}^{\infty} s^{k'} P_i(m, i \rightarrow k', t) \\ &= s G_i(s, t) \end{aligned}$$

Zo wordt vergelijking 2.4 een partiële differentiaalvergelijking in $G_i(s, t)$

$$\frac{\partial G_i(s, t)}{\partial t} = \lambda(t)(s-1)G_i(s, t) \quad (2.7)$$

De beginvoorwaarde uit vergelijking 2.5 wordt voor voor de genererende functie vertaald naar

$$G_i(s, i) = \sum_{k=m}^{\infty} s^k \delta_{k,m} = s^m. \quad (2.8)$$

De oplossing voor vergelijking 2.7 is

$$G_i(s, t) = G_i(s, i) \exp\{(s-1)\Lambda_i(t)\}, \quad (2.9)$$

met de verkorte notatie

$$\Lambda_i(t) = \int_i^t \lambda(t') dt' \quad (2.10)$$

Door gebruik te maken van vergelijking 2.8, wordt de oplossing van vergelijking 2.7

$$G_i(s, t) = s^m \exp\{(s-1)\Lambda_i(t)\} \quad (2.11)$$

Om terug te gaan naar de kans $P_i(m, i \rightarrow k, t)$, kan er gebruik gemaakt worden van de maclaurinreeks (taylorreeks rond 0) voor de exponentiële functie en een substitutie ($k' = k + m$)

$$\begin{aligned} G_i(s, t) &= s^m \exp\{(s-1)\Lambda_i(t)\} \\ &= s^m \exp\{-\Lambda_i(t)\} \exp\{s\Lambda_i(t)\} \\ &= \exp[-\Lambda_i(t)] \sum_{k=0}^{\infty} \frac{s^{m+k} [\Lambda_i(t)]^k}{k!} \\ &= \exp[-\Lambda_i(t)] \sum_{k'=m}^{\infty} \frac{s^{k'} [\Lambda_i(t)]^{k'-m}}{(k'-m)!} \end{aligned}$$

In de laatste vergelijking kan de definitie van de genererende functie herkend worden (vergelijking 2.6). Na gelijkstelling van beide vergelijkingen wordt dit

$$P_i(m, i \rightarrow k, t) = \frac{[\Lambda_i(t)]^{k-m}}{(k-m)!} \exp[-\Lambda_i(t)] \quad (2.12)$$

Bovenstaande formule is een mooie algemene uitdrukking voor de kans dat een node een degree k heeft op tijd t bij een groeiproces dat afhangt van de tijd zonder preferentiële aanhechting. Bovendien is het duidelijk dat deze kansverdeling een (vershoven) Poissonverdeling is, met verwachtingswaarde $\Lambda_i(t)$.

2.1.2 Normering, verwachtingswaarde en variantie

Vergelijking 2.12 geeft de kansverdeling voor de degree k . De kans dat node i echter eender welke degree heeft na deze tijd moet gelijk zijn aan 1, aangezien dit alle mogelijke uitkomsten omvat voor node i . Met andere woorden, de norm van vergelijking 2.12 moet gelijk zijn aan 1:

$$\begin{aligned} \|P_i(m, i \rightarrow k, t)\| &= \sum_{k=m}^{\infty} \frac{[\Lambda_i(t)]^{k-m}}{(k-m)!} \exp[-\Lambda_i(t)] \\ &= \exp[\Lambda_i(t)] \exp[-\Lambda_i(t)] \\ &= 1, \end{aligned}$$

waarbij nogmaals gebruik werd gemaakt van de substitutie $k' = k - m$ en de maclaurinreeks voor de exponentiële functie. Merk op dat deze normering geldt voor alle waarden van $t > i$ en alle waarden van $k \geq m$.

De verwachtingswaarde van de degree, bepaald door de gevonden kansverdeling, wordt gegeven door

$$\begin{aligned} \mathbb{E}[k] &= \sum_{k=m}^{\infty} k P_i(m, i \rightarrow k, t) \\ &= \sum_{k=m}^{\infty} k \frac{[\Lambda_i(t)]^{k-m}}{(k-m)!} \exp[-\Lambda_i(t)] \\ &= \sum_{k=m}^{\infty} (k-m) \frac{[\Lambda_i(t)]^{k-m}}{(k-m)!} \exp[-\Lambda_i(t)] \\ &\quad + m \sum_{k=m}^{\infty} \frac{[\Lambda_i(t)]^{k-m}}{(k-m)!} \exp[-\Lambda_i(t)] \end{aligned}$$

In de sommatie van de eerste term in bovenstaande vergelijking kan weer de substitutie $k' = k - m$ toegepast worden. Bovendien kan er opgemerkt worden dat de term voor $k' = 0$ wegvalt. De nieuwe index k' start dan bij 1. De sommatie in de tweede term komt overeen met de normering van de kansverdeling, welke 1 is. Al deze manipulaties geven

$$\begin{aligned} \mathbb{E}[k] &= m + \sum_{k'=1}^{\infty} k' \frac{[\Lambda_i(t)]^{k'}}{k'!} \exp[-\Lambda_i(t)] \\ &= m + \Lambda_i(t) \sum_{k'=1}^{\infty} \frac{[\Lambda_i(t)]^{k'-1}}{(k'-1)!} \exp[-\Lambda_i(t)] \end{aligned}$$

De sommatie in de tweede term komt overeen met de normering van de kansverdeling voor $m = 1$ en geeft bijgevolg 1. Het eindresultaat voor de verwachtingswaarde van de degree is dan

$$\mathbb{E}[k] = m + \Lambda_i(t) \tag{2.13}$$

De activiteit $\lambda(t)$, die het groeiproces vastlegt, komt dus in de verwachtingswaarde bovendrijven als $\Lambda_i(t)$. Deze term kan geïnterpreteerd worden als het verwachte aantal geaccumuleerde links in het tijdsinterval i tot t .

De eerste term komt voort uit de begintoestand van iedere node. Elke node wordt initieel met m links verbonden aan het netwerk. Daardoor heeft geen enkele node een degree lager dan m .

Niet alleen de verwachte degree, maar ook de spreiding van de fluctuaties erop kan bepaald worden. De variantie wordt gegeven door

$$\sigma^2 [k] = \mathbb{E} [k^2] - (\mathbb{E} [k])^2. \quad (2.14)$$

Hoewel het tweede moment $\mathbb{E} [k^2]$ berekend kan worden aan de hand van de formele definitie (zoals eerder werd gedaan voor de verwachtingswaarde van k), is dit ook mogelijk via de genererende functie (zie bijlage A voor een gedetailleerde afleiding):

$$\mathbb{E} [k^2] = \left. \frac{\partial}{\partial s} \left[s \frac{\partial}{\partial s} G_i(s, t) \right] \right|_{s=1} \quad (2.15)$$

De afgeleide van de genererende functie (vergelijking 2.11) is alvast

$$s \frac{\partial}{\partial s} G_i(s, t) = m s^m \exp [(s-1)\Lambda_i(t)] + s^{m+1} \Lambda_i(t) \exp [(s-1)\Lambda_i(t)].$$

Het tweede moment van k is dan

$$\begin{aligned} \mathbb{E} [k^2] &= (m^2 s^{m-1} \exp [(s-1)\Lambda_i(t)] + m s^m \Lambda_i(t) \exp [(s-1)\Lambda_i(t)] \\ &\quad + (m+1)s \Lambda_i(t) \exp [(s-1)\Lambda_i(t)] + s^{m+1} \Lambda_i(t)^2 \exp [(s-1)\Lambda_i(t)]) \Big|_{s=1} \\ &= [m + \Lambda_i(t)]^2 + \Lambda_i(t) \end{aligned}$$

Ingevuld in vergelijking 2.14 geeft dit als variantie

$$\sigma^2 [k] = \Lambda_i(t). \quad (2.16)$$

Deze resultaten geven een mooi algemeen resultaat voor de activiteit λ die enkel expliciet afhankelijk is van de tijd/grootte van het netwerk. In wat volgt wordt er in een voorbeeld een specifieke vorm voor λ bekeken.

2.1.3 Wachtijd T_k

Indien het concreet verloop van de degree van node i gegeven wordt door $K_i(t)$, kan de nodige tijd T_k om degree k worden bepaald aan de hand van

$$T_k = K_i^{-1}(k),$$

waar $K_i^{-1}(k)$ de inverse functie is van $K_i(t)$, die het tijdstip geeft waarop node i de degree k bereikt.

Vergelijking 2.12 geeft nu kansverdeling van node i om degree k te bereiken, gegeven het tijdstip t , met $K_i(t)$ een stochastische variabele gegenereerd door deze kansverdeling. De tijd T_k is bijgevolg ook een stochastische variabele die gerealiseerd wordt door de kansverdeling $f_i(T_k)$, die de kans geeft dat node i een tijd T_k nodig heeft om de degree k te bereiken.

De kansverdeling $f_i(T_k)$ wordt gegeven door haar cumulatieve distributiefunctie $F_i(T_k)$ die de kans geeft dat node i minstens degree k op tijdstip T_k . Dit wordt geschreven als

$$\begin{aligned} F_i(T_k) &= 1 - \text{Prob} [K_i(T_k) < k] \\ &= 1 - \sum_{n=m}^{k-1} \text{Prob} [K_i(T_k) = n] \\ &= 1 - \sum_{n=m}^{k-1} P_i(m, i \rightarrow n, T_k) \\ &= 1 - \sum_{n=m}^{k-1} \frac{[\Lambda_i(T_k)]^{n-m}}{(n-m)!} \exp [-\Lambda_i(T_k)] \end{aligned}$$

De substitutie $n' = n - m$ geeft dan

$$F_i(T_k) = 1 - \exp[-\Lambda_i(T_k)] \sum_{n'=0}^{k-m-1} \frac{[\Lambda_i(T_k)]^{n'}}{n'!}. \quad (2.17)$$

Gebruik makend van de incomplete gammafunctie

$$\Gamma(k, x) = \int_x^\infty z^{k-1} e^{-z} dz \quad (2.18)$$

en haar eigenschap [17]

$$e^x \frac{\Gamma(k, x)}{\Gamma(k)} = \sum_{n=0}^{k-1} \frac{x^n}{n!}, \quad (2.19)$$

wordt de cumulatieve distributiefunctie

$$F_i(T_k) = 1 - \frac{\Gamma[k - m, \Lambda_i(T_k)]}{\Gamma[k - m]}. \quad (2.20)$$

Voor de frequentie distributie geldt $f_i(T_k) = \frac{dF_i(T_k)}{dT_k}$. Om de afleiding te vereenvoudigen (en het afleiden van incomplete gammafunctie te vermijden), is vergelijking 2.17 een goed begin:

$$f_i(T_k) = \frac{d\Lambda_i(T_k)}{dT_k} \left[\exp[-\Lambda_i(T_k)] \sum_{n'=0}^{k-m-1} \frac{[\Lambda_i(T_k)]^{n'}}{n'!} - \exp[-\Lambda_i(T_k)] \sum_{n'=1}^{k-m-1} \frac{[\Lambda_i(T_k)]^{n'-1}}{(n'-1)!} \right]$$

Aangezien $\frac{d\Lambda_i(T_k)}{dT_k} = \lambda(T_k)$ en aan de hand van een substitutie in de tweede term wordt dit

$$f_i(T_k) = \lambda(T_k) \exp[-\Lambda_i(T_k)] \left[\sum_{n'=0}^{k-m-1} \frac{[\Lambda_i(T_k)]^{n'}}{n'!} - \sum_{n'=0}^{k-m-2} \frac{[\Lambda_i(T_k)]^{n'}}{n'!} \right],$$

met als resultaat

$$f_i(T_k) = \lambda(T_k) \exp[-\Lambda_i(T_k)] \frac{[\Lambda_i(T_k)]^{k-m-1}}{(k-m-1)!}. \quad (2.21)$$

2.1.4 Voorbeeld

De vorm van λ die hier wordt voorgesteld, is gebaseerd op een groeiend netwerk zonder preferentiële aanhechting en waarbij iedere node begint met m links die willekeurig worden verspreid over de t nodes in het bestaande netwerk (random groei sectie 1.3.2). Met andere woorden, de activiteit is gegeven door $\lambda(t) = \frac{m}{t}$. Allereerst kan hieruit $\Lambda_i(t)$ afgeleid worden:

$$\begin{aligned} \Lambda_i(t) &= \int_i^t \frac{m}{t'} dt' \\ &= m \ln \left(\frac{t}{i} \right) \end{aligned}$$

Dit betekent voor de kansverdeling, verwachtingswaarde en variantie:

$$P_i(m, i \rightarrow k, t) = \frac{[m \ln(\frac{t}{i})]^{k-m}}{(k-m)!} \left(\frac{i}{t}\right)^m \quad (2.22)$$

$$\mathbb{E}[k] = m + m \ln\left(\frac{t}{i}\right) \quad (2.23)$$

$$\sigma^2[k] = m \ln\left(\frac{t}{i}\right) \quad (2.24)$$

Het resultaat voor de verwachtingswaarde van de degree k stemt volledig overeen met het resultaat van sectie 1.3.2, waar vanaf het begin gewerkt werd met de mean-field benadering. Figuur 2.1 toont de tijdsevolutie van de gevonden kansverdeling: naarmate de tijd vordert, wordt de verdeling een bredere Poissonverdeling, waarbij het gemiddelde logaritmisches groter wordt doorheen de tijd.

De kansverdeling voor de wachttijd T_k wordt voor dit geval

$$f_i(T_k) = \frac{m^{k-m} i^m}{T_k^{m+1} (k-m-1)!} \left[\ln\left(\frac{T_k}{i}\right) \right]^{k-m-1} \quad (2.25)$$

en staat samen met haar cumulatieve distributiefunctie geplot in figuur 2.2.

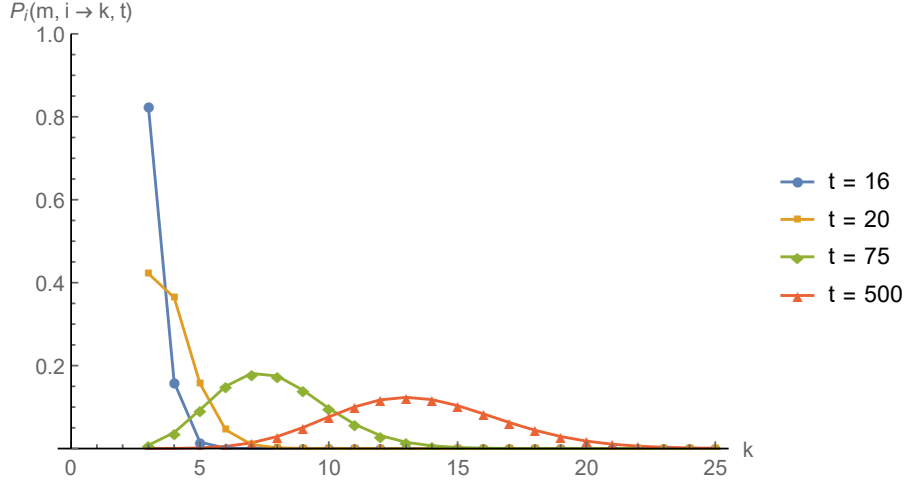
De verwachte wachttijd is dan

$$\mathbb{E}[T_k] = \sum_{T_k=i}^{\infty} \frac{m^{k-m} i^m}{T_k^m (k-m-1)!} \left[\ln\left(\frac{T_k}{i}\right) \right]^{k-m-1} \quad (2.26)$$

Volgens vergelijking 1.11, is deze wachttijd in het mean-field scenario

$$T_k = i \exp\left[\frac{k-m}{m}\right], \quad (2.27)$$

wat duidelijk niet overeenkomt de verwachtingswaarde uit vergelijking 2.26. Wat de oorzaak is van deze slechte overeenkomst is niet duidelijk, aangezien er geen simulaties werden uitgevoerd om de kansverdeling voor de wachttijd uitvoerig te controleren.



Figuur 2.1: De gevonden Poissonverdeling uit vergelijking 2.22 voor de degree k van node $i = 15$ in een netwerk dat groeit volgens het random groeiproces voor 4 verschillende tijdstippen. Iedere nieuwe node hecht zich met $m = 3$ links aan het netwerk. De verwachtingswaarde en de breedte van de kansverdeling stijgen logaritmisch in de tijd, wat zichtbaar is als de kansverdeling die breder wordt en opschuift.

2.2 Preferentiële groei

2.2.1 Kansverdeling voor de degree

Voor een netwerk dat groeit met nodes die zich via preferentiële aanhechting verbinden, kan er op dezelfde manier worden gewerkt. Hier dient wel opgemerkt te worden dat dit proces géén Poissonverdeling zal beschrijven, aangezien de preferentiële aanhechting ervoor zorgt dat de waarschijnlijkheid van een nieuwe link met een node afhankelijk is van het aantal links die de node reeds heeft. De gebeurtenissen zijn bijgevolg *niet* onafhankelijk. De overgangswaarschijnelijkheden tussen de toestanden van de node i worden nu gegeven door

$$P_i(k, t \rightarrow k - 1, t + \Delta t) = 0 \quad (2.28)$$

$$P_i(k, t \rightarrow k + 1, t + \Delta t) = \frac{k}{2t} \Delta t \quad (2.29)$$

$$P_i(k, t \rightarrow k, t + \Delta t) = 1 - \frac{k}{2t} \Delta t, \quad (2.30)$$

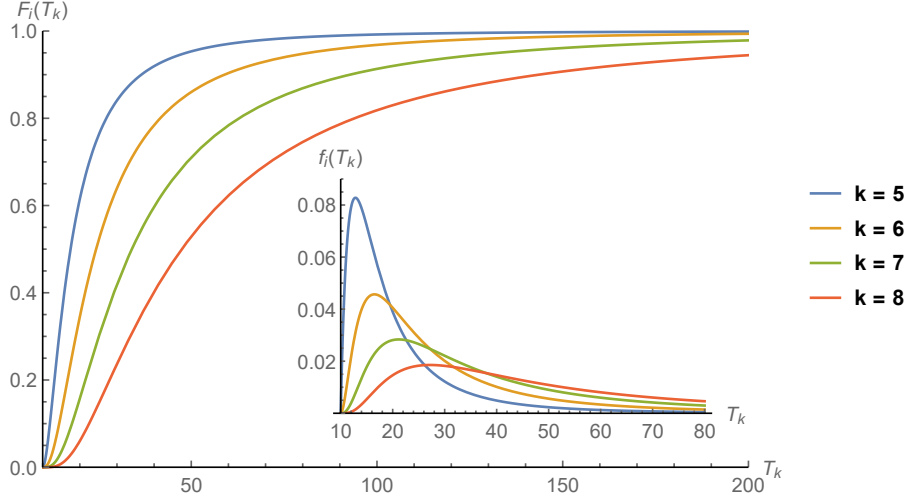
naar analogie met vergelijking 1.6. De master equation voor de kans dat node i op tijdstip $t + \Delta t$ degree k heeft, wordt dan

$$P_i(m, i \rightarrow k, t + \Delta t) = P_i(m, i \rightarrow k - 1, t) \frac{k - 1}{2t} \Delta t + P_i(m, i \rightarrow k, t) \left(1 - \frac{k}{2t} \Delta t\right).$$

Gebruik makend van de continue tijdsbenadering (en na wat herschikking) wordt dit

$$\frac{\partial}{\partial t} P_i(m, i \rightarrow k, t) = \frac{1}{2t} [k P_i(m, i \rightarrow k - 1, t) - k P_i(m, i \rightarrow k, t) - P_i(m, i \rightarrow k - 1, t)].$$

Deze vergelijking kan weer opgelost worden aan de hand van de genererende functiemethode. De genererende functie heeft hier dezelfde definitie als in het niet-preferentiële geval (vergelijking 2.6).



Figuur 2.2: De CDF voor de wachttijd T_k bij 4 verschillende waarden voor k . De overeenkomstige kansverdeling staat voor de zelfde waarden k geplot in de inzet.

Vervolgens wordt bovenstaande vergelijking ook weer langs beide zijden vermengvuldigd met s^k , waarna er gesommeerd wordt over k :

$$\frac{\partial}{\partial t} G_i(s, t) = \sum_{k=m}^{\infty} \frac{s^k}{2t} [kP_i(m, i \rightarrow k-1, t) - kP_i(m, i \rightarrow k, t) - P_i(m, i \rightarrow k-1, t)] \quad (2.31)$$

In het rechterlid moeten de volgende drie termen daarbij beter bekeken worden:

$$\begin{aligned} \sum_{k=m}^{\infty} s^k P_i(m, i \rightarrow k-1, t) &= s \sum_{k=m}^{\infty} s^{k-1} P_i(m, i \rightarrow k-1, t) \\ &= sG_i(s, t) \\ \sum_{k=m}^{\infty} ks^k P_i(m, i \rightarrow k, t) &= s \frac{\partial}{\partial s} \left[\sum_{k=m}^{\infty} s^k P_i(m, i \rightarrow k, t) \right] \\ &= s \frac{\partial}{\partial s} [G_i(s, t)] \\ \sum_{k=m}^{\infty} ks^k P_i(m, i \rightarrow k-1, t) &= s \frac{\partial}{\partial s} \left[\sum_{k=m}^{\infty} s^k P_i(m, i \rightarrow k-1, t) \right] \\ &= s \frac{\partial}{\partial s} [sG_i(s, t)] \\ &= sG_i(s, t) + s^2 \frac{\partial}{\partial s} G_i(s, t) \end{aligned}$$

Bovenstaande uitkomsten zorgen na invullen in vergelijking 2.31 en wat herwerking voor een partiële differentiaalvergelijking in $G_i(s, t)$:

$$\frac{\partial}{\partial t} G_i(s, t) = \frac{s(s-1)}{2t} \frac{\partial}{\partial s} G_i(s, t), \quad (2.32)$$

met nog altijd dezelfde randvoorwaarde $G_i(s, i) = s^m$.

De oplossing van bovenstaande vergelijking wordt gegeven door

$$G_i(s, t) = \left[\frac{s}{s - (s-1)\sqrt{\frac{t}{i}}} \right]^m. \quad (2.33)$$

Deze genererende functie kan vervolgens weer geschreven worden als een machtreeks. Hiervoor is eerst wat herwerking nodig:

$$\begin{aligned} G_i(s, t) &= s^m \left[\frac{1}{\sqrt{\frac{t}{i}} + \left(1 - \sqrt{\frac{t}{i}}\right) s} \right]^m \\ &= s^m \left(\frac{i}{t}\right)^{\frac{m}{2}} \frac{1}{\left[1 + \left(\sqrt{\frac{i}{t}} - 1\right) s\right]^m} \end{aligned}$$

Aan de hand van het binomiaal theorema van Newton¹

$$(1+x)^{-r} = \sum_{k=0}^{\infty} \binom{k+r-1}{k} (-x)^k,$$

wordt de genererende functie

$$G_i(s, t) = s^m \left(\frac{i}{t}\right)^{\frac{m}{2}} \sum_{k=0}^{\infty} s^k \binom{k+m-1}{k} \left(1 - \sqrt{\frac{i}{t}}\right)^k.$$

Met de substitutie $k' = k + m$ wordt dit uiteindelijk de machtreeks

$$G_i(s, t) = \left(\frac{i}{t}\right)^{\frac{m}{2}} \sum_{k'=m}^{\infty} s^{k'} \binom{k'-1}{k'-m} \left(1 - \sqrt{\frac{i}{t}}\right)^{k'-m}, \quad (2.34)$$

waarvan de coëfficiënten de gezochte kansverdeling bepalen:

$$P_i(m, i \rightarrow k, t) = \left(\frac{i}{t}\right)^{\frac{m}{2}} \binom{k-1}{k-m} \left(1 - \sqrt{\frac{i}{t}}\right)^{k-m} \quad (2.35)$$

Figuur 2.3 toont deze kansverdeling voor node 15 en een preferentiële groei, waarbij $m = 3$ voor verschillende tijdstippen. Zeer duidelijk is de snel groeiende breedte van de kansverdeling.

2.2.2 Normering, verwachtingswaarde en variantie

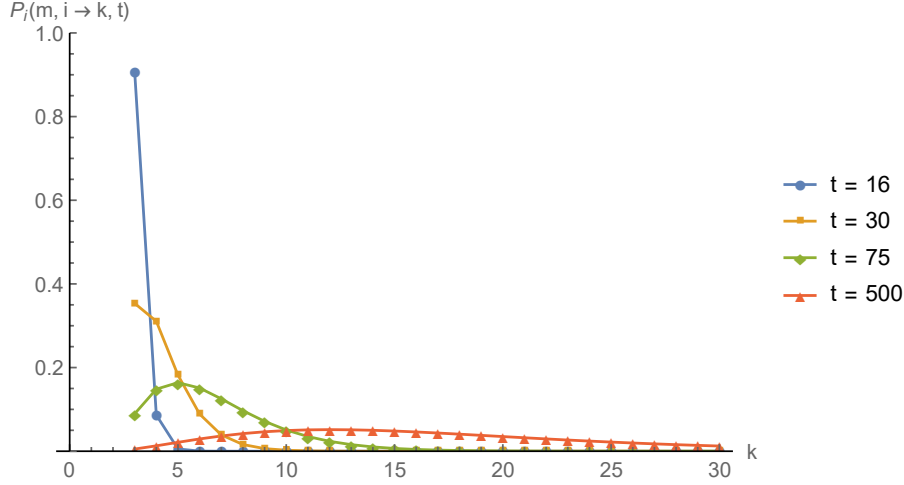
Het is eenvoudig aan te tonen dat de normering van de gevonden kansverdeling gelijk is aan 1. Aangezien

$$G_i(1, t) = \sum_{k=m}^{\infty} P_i(m, i \rightarrow k, t),$$

toont vergelijking 2.33 aan dat de kansverdeling aan de correcte normering voldoet:

$$\begin{aligned} G_i(1, t) &= \left[\frac{1}{1 - (1-1)\sqrt{\frac{t}{i}}} \right]^m \\ &= 1. \end{aligned}$$

¹Het binomiaal theorema van Newton is een veralgemening van het binomiaal theorema $(1+x)^r = \sum_{k=0}^r \binom{r}{k} x^k$, waar de binomiaalcoëfficiënt veralgemeend wordt naar het complexe vlak [3].



Figuur 2.3: Het resultaat uit vergelijking 2.35 geplot voor node $i = 15$. Het aantal links waarmee nodes zich preferentieel aanhechten werd ingesteld op $m = 3$. Hier worden 4 tijdstippen getoond, waaruit duidelijk wordt dat de breedte van de kansverdeling sneller groeit dan in het niet-preferentiële geval.

De genererende functie geeft een directe methode om de verwachtingswaarde van de degree te vinden:

$$\begin{aligned}
 \mathbb{E}[k] &= \left. \frac{\partial}{\partial s} G_i(s, t) \right|_{s=1} \\
 &= m \sqrt{\frac{t}{i}} \frac{s^{m-1}}{\left[s - (s-1) \sqrt{\frac{t}{i}} \right]^{m+1}} \Bigg|_{s=1} \\
 &= m \sqrt{\frac{t}{i}}
 \end{aligned} \tag{2.36}$$

Bovenstaand resultaat komt volledig overeen met het resultaat van het BA-model (vergelijking 1.7). Daar bovenop kan de spreiding rond dit resultaat gevonden worden aan de hand van de variantie

$$\sigma^2[k] = \mathbb{E}[k^2] - (\mathbb{E}[k])^2. \tag{2.37}$$

Het tweede moment kan hierbij nogmaals berekend worden aan de hand van de genererende functie:

$$\mathbb{E}[k^2] = \left. \frac{\partial}{\partial s} \left[s \frac{\partial}{\partial s} G_i(s, t) \right] \right|_{s=1} \tag{2.38}$$

De eerste afgeleide van de genererende functie is alvast

$$s \frac{\partial}{\partial s} G_i(s, t) = m \sqrt{\frac{t}{i}} \frac{s^m}{\left[s - (s-1) \sqrt{\frac{t}{i}} \right]^{m+1}}$$

De tweede afgeleide is dan na wat rekenwerk

$$\frac{\partial}{\partial s} \left[s \frac{\partial}{\partial s} G_i(s, t) \right] = s^{m-1} m \sqrt{\frac{t}{i}} \frac{m \sqrt{\frac{t}{i}} - s + s \sqrt{\frac{t}{i}}}{\left[s - (s-1) \sqrt{\frac{t}{i}} \right]^{m+2}},$$

waardoor het tweede moment gelijk is aan

$$\begin{aligned}\mathbb{E}[k^2] &= \frac{\partial}{\partial s} \left[s \frac{\partial}{\partial s} G_i(s, t) \right] \Big|_{s=1} \\ &= m^2 \frac{t}{i} + m \frac{t}{i} - m \sqrt{\frac{t}{i}}.\end{aligned}$$

Uiteindelijk wordt de variantie op de degree dan

$$\sigma^2[k] = m \sqrt{\frac{t}{i}} \left(\sqrt{\frac{t}{i}} - 1 \right). \quad (2.39)$$

Deze variantie is, voor grote $\frac{t}{i}$, groter dan voor het resultaat van een random groeiend netwerk (vergelijking 2.24). Dit is ook duidelijk in figuur 2.3. Preferentiële groei zorgt er voor dat een kleine afwijking in de degree een drastisch verschil kan geven op een later tijdstip, als gevolg van het groeiproces dat afhankelijk is van de eerdere degree van de node. Aan de hand van het BA-model kon de variantie op het verwachte gedrag niet voorspeld worden, aangezien daar vanaf het begin de mean-field benadering werd toegepast.

2.3 Numerieke controle

De simulaties die in deze sectie gebruikt worden ter bevestiging van de theoretische resultaten, werden uitgevoerd voor preferentiële groei. In een eerste simulatie wordt er gekeken naar de verwachte degree van een node en de spreiding erop. Een tweede deel bekijkt de effectieve vorm van de kansverdeling van de degree.

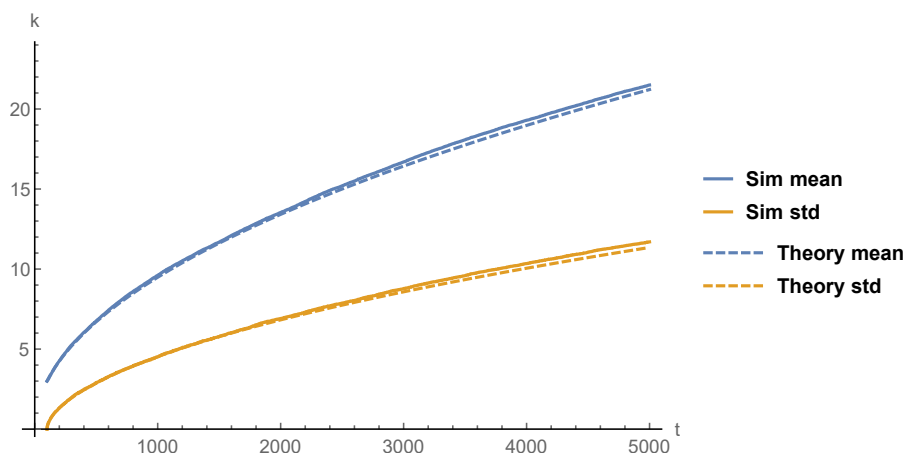
2.3.1 Verwachtingswaarde en standaardafwijking

In een gesimuleerd netwerk kan op ieder tijdstip de degree van een gekozen node uitgerekend worden (bijvoorbeeld aan de hand van de verbindingsmatrix G op desbetreffend tijdstip). Door dit groeiproces te herhalen, wordt voor ieder tijdstip een lijst met bekomen degrees opgesteld, waaruit het gemiddelde bepaald kan worden. Ook de spreiding wordt dan duidelijk door de standaardafwijking te berekenen van dezelfde lijst.

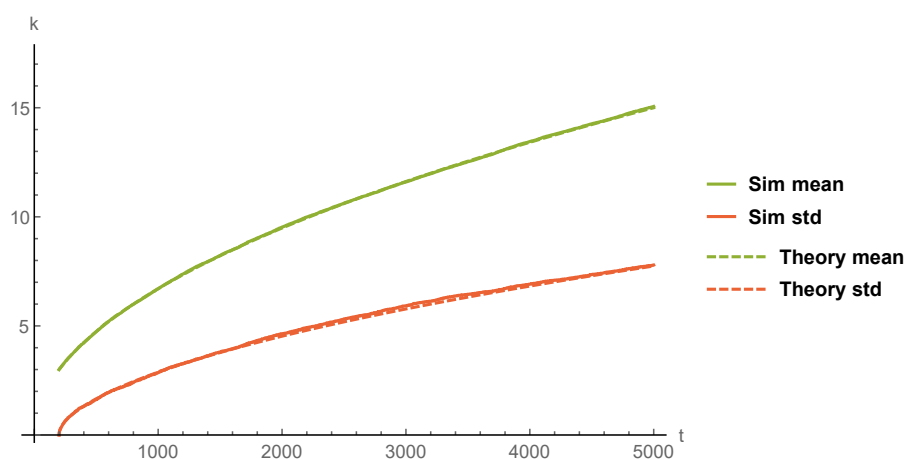
Figuur 2.4 toont het resultaat voor deze simulaties en vergelijkt ze met het theoretische resultaat van vergelijkingen 2.36 en 2.39. Wat meteen duidelijk is, is de goede overeenkomst tussen beide curves.

Dit is te verwachten voor de gemiddelde degree (blauwe curve), aangezien dit reeds het resultaat was van het BA-model. Daar was het gevolg van het eindige netwerk een afwijking van de voorspelde $\gamma = 3$ in de degree distributie (zie figuur 1.5). Hier wordt zichtbaar dat deze afwijking een gevolg is van nodes die gemiddeld een hogere degree hebben dan voorspeld wordt. Naast de verwachte degree, werd ook de variantie voorspeld (oranje curve). Ook voor de variantie (op de figuur eigenlijk de standaardafwijking) is er een goede overeenkomst, maar is dezelfde afwijking zichtbaar als bij de degree.

Dezelfde simulatie werd voor een later geboren node herhaald in figuur 2.5. De afwijking gezien in figuur 2.4 is daar niet (of weinig) zichtbaar. De geobserveerde fout is dus afhankelijk van de geobserveerde node: nodes die vroeg werden geboren hebben een grotere afwijking van het



Figuur 2.4: De verwachtingswaarde (blauw) en standaardafwijking (oranje) voor de degree van node 100 in een groeiend netwerk ($m = 3$) na 1000 simulaties. Ook het theoretische resultaat staat steeds geplot als onderbroken lijn.



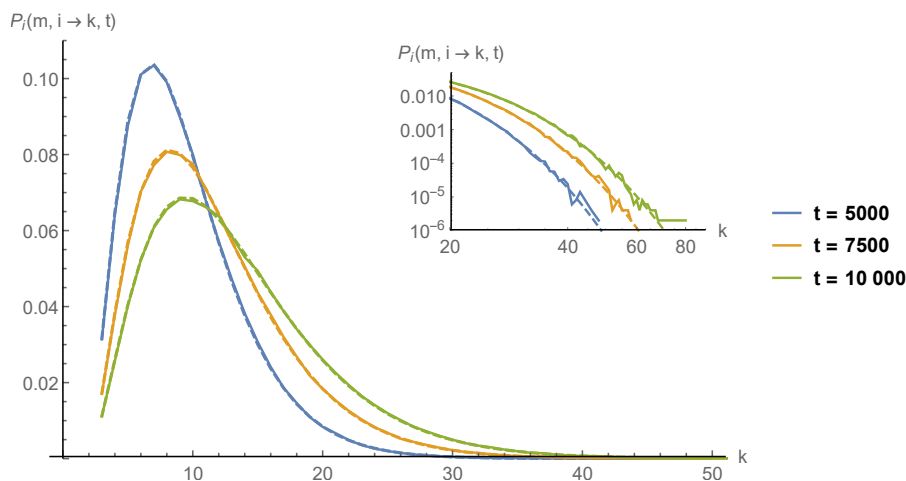
Figuur 2.5: Dezelfde simulatie als figuur 2.4, maar dit keer voor node 200. Dit maal is er (bijna) geen consistent groeiende afwijking zichtbaar bij zowel het gemiddelde als de standaardafwijking.

BA-model dan nodes die later toegevoegd werden. Met andere woorden, de eerste nodes van het netwerk tonen een afwijkende groei van het BA-model. Voor grotere netwerken wordt deze fractie nodes steeds kleiner, wat voor een betere overeenkomst zorgt tussen theorie en werkelijkheid. Grotere netwerken zullen dus een exponent γ hebben die dichter naar 3 nadert. De oorzaak van de afwijkende groei van oudere nodes kan de start van het netwerk zijn. Het begin van het groeiproces wordt namelijk niet behandeld door de master equation uit hoofdstuk 2. Wanneer het netwerk reeds gegroeid is, geldt de beschrijving (en de continue tijdsbenadering) dan steeds beter. De simulaties voor een groeiend netwerk werden steeds gestart met een volledig verbonden netwerk. De nodes beginnen dus allemaal met dezelfde degree.

2.3.2 Kansverdeling

Met dezelfde methode als die voor het achterhalen van de verwachtingswaarde, kan de volledige kansverdeling gezocht worden. In plaats van voor ieder tijdstip een uitmiddeling uit te voeren over alle simulaties, moet er dit keer gekeken worden naar de frequentie van ieder mogelijk resultaat. Hiervoor zijn voor een duidelijk resultaat veel meer simulaties nodig dan voordien.

Dit keer wordt ook een node gekozen die voldoende laat geboren is in de groei van het netwerk, zodat de node het BA-model beter volgt.



Figuur 2.6: De frequentie van iedere degree (volle lijn) bekomen na 500 000 simulaties, waarbij node 500 gevolgd werd in een groeiend schaalvrij netwerk met $m = 3$. Het theoretische resultaat van vergelijking 2.35 wordt voorgesteld door een onderbroken lijn. Curves met dezelfde kleur werden berekend voor hetzelfde tijdstip. De inzet toont dezelfde grafiek op log-log schaal.

Figuur 2.6 toont de resultaten van deze simulatie voor dezelfde node op drie verschillende tijdstippen. De inzet toont dezelfde grafiek op log-log schaal. Het is duidelijk dat de onderbroken lijn, die het theoretische resultaat van vergelijking 2.35 toont, een mooie overeenkomst heeft met de resultaten uit de simulatie. Zelfs voor grotere degrees blijft er een mooie consensus, totdat de telruis dominant wordt. Vooral op log-log schaal is dit duidelijk.

De combinatie van de goede overeenkomst voor de verwachtingswaarde, variantie én volledige kansverdeling toont aan dat de theoretische resultaten (vergelijkingen 2.35, 2.36 en 2.39) de groei van het netwerk correct beschrijven.

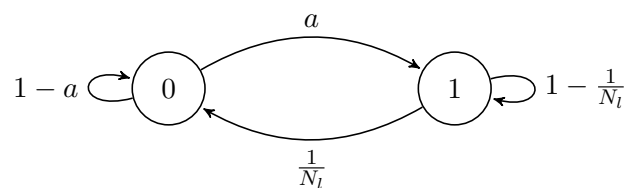
Hoofdstuk 3

Link herschikking

Het tot stand brengen van netwerken werd alvast beschreven in hoofdstuk 2. Daarbij werd geen rekening gehouden met links die na hun geboorte nog kunnen veranderen, links waren statisch. In realiteit hoeft dit niet zo te zijn, links kunnen verdwijnen of gevormd worden. Dit bemoeilijkt het beschrijven van het groeiproces, daarom bekijkt dit hoofdstuk een vereenvoudigde versie die groei buiten beschouwing laat. Ook de manier waarop links verdwijnen en weer verschijnen wordt vastgelegd om een theoretische beschrijving mogelijk te maken.

3.1 Willekeurige verplaatsing

Hier wordt een eenvoudig model bekeken: het willekeurig verplaatsen van één link per tijdstap, terwijl het aantal nodes N_n en aantal links N_l behouden blijft. De kans $p_{ij}(t)$ dat er een link bestaat tussen nodes i en j evolueert dan van een gegeven $p_{ij}(0)$ (de begintoestand) naar, hopelijk, een steady state. Het is te verwachten dat deze steady state een Bernoulli netwerk zal zijn, aangezien de links (na genoeg tijdstappen) willekeurig verdeeld zijn tussen de nodes.



Figuur 3.1: Een (Markov-) keten die de overgangen en hun waarschijnlijkheden toont tussen de toestanden 0 (geen link aanwezig) en 1 (wel link aanwezig) voor de link $i \leftrightarrow j$. De kans a is een korte notatie voor $\frac{1}{\frac{N_n(N_n-1)}{2} - (N_l-1)}$.

Veronderstel dat de link $i \leftrightarrow j$ reeds bestaat op tijd t . De kans dat deze link wordt weggenomen na een tijdstap Δt is dan gelijk aan $\frac{1}{N_l}$, aangezien er per tijdstap één bestaande link willekeurig wordt weggenomen van haar originele plek. De kans dat de link niet wordt weggenomen is bijgevolg

$$1 - \frac{1}{N_l}.$$

Indien er geen link aanwezig is tussen i en j op dit tijdstip, zijn deze waarschijnlijkheden anders. De te verzetten link wordt op een willekeurige, lege plek geplaatst. Dit aantal plaatsen is gelijk

aan het totale aantal mogelijke links in het netwerk $\left(\frac{N_n(N_n-1)}{2}\right)$ min het aantal plaatsen dat al bezet is door een link $(N_l - 1)$. De kans dat er na een tijdstap Δt wél een link aanwezig is, is dan

$$\frac{1}{\frac{N_n(N_n-1)}{2} - (N_l - 1)} = a.$$

Figuur 3.1 toont een visualisatie van dit proces in een Markovketen.

Bovenstaande waarschijnlijkheden kunnen gebruikt worden om nogmaals een master equation op te stellen:

$$p_{ij}(t) = \left(1 - \frac{1}{N_l}\right) p_{ij}(t-1) + a [1 - p_{ij}(t-1)] \quad (3.1)$$

$$= \left(1 - \frac{1}{N_l} - a\right) p_{ij}(t-1) + a \quad (3.2)$$

Ook hier kan de genererende functiemethode gebruikt worden om een eenvoudige oplossing te verkrijgen. De genererende functie wordt dit keer geschreven als

$$G_{ij}(x) = \sum_{t=1}^{\infty} x^t p_{ij}(t). \quad (3.3)$$

In vergelijking 3.2 wordt de term in $p_{ij}(t-1)$ dan

$$\begin{aligned} \sum_{t=1}^{\infty} x^t p_{ij}(t-1) &= \sum_{t'=0}^{\infty} x^{t'+1} p_{ij}(t') \\ &= x \left(p_{ij}(0) + \sum_{t'=1}^{\infty} x^{t'} p_{ij}(t') \right) \\ &= x (p_{ij}(0) + G_{ij}(x)) \end{aligned}$$

De term in a wordt

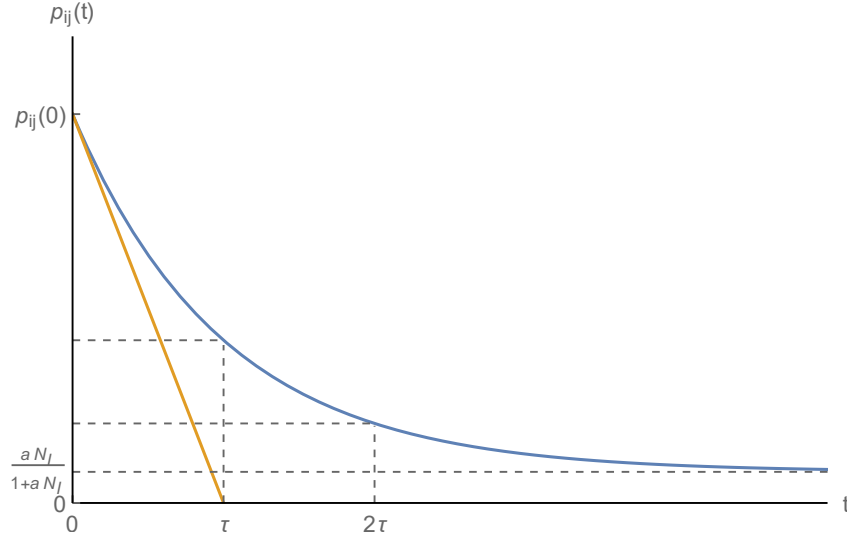
$$\begin{aligned} a \sum_{t=1}^{\infty} x^t &= a \left(\sum_{t=0}^{\infty} x^t - 1 \right) \\ &= a \left(\frac{1}{1-x} - 1 \right) \\ &= \frac{ax}{1-x} \end{aligned}$$

De volledige master equation geeft

$$G_{ij}(x) = \left(1 - \frac{1}{N_l} - a\right) x [G_{ij}(x) + p_{ij}(0)] + \frac{ax}{1-x}. \quad (3.4)$$

Aan de hand van de substitutie $b = 1 - \frac{1}{N_l} - a$ en wat algebra kan dit herschreven worden als

$$\begin{aligned} G_{ij}(x) &= p_{ij}(0) \frac{bx}{1-bx} + \frac{ax}{(1-x)(1-bx)} \\ &= p_{ij}(0) \frac{bx}{1-bx} + ax \left[\frac{1}{(1-b)(1-x)} + \frac{b}{(b-1)(1-bx)} \right] \\ &= \sum_{t=1}^{\infty} x^t \left[b^t \left(p_{ij}(0) - \frac{a}{1-b} \right) + \frac{a}{1-b} \right] \end{aligned}$$



Figuur 3.2: De visuele voorstelling van vergelijking 3.5. De figuur toont een duidelijk exponentieel verval, met tijdconstante τ die het snijpunt met de x-as geeft van de raaklijn aan het begin van het verval. De tijdconstante wordt gegeven door vergelijking 3.6. Dit exponentieel verval heeft als limietwaarde vergelijking 3.7.

De uitdrukking voor $p_{ij}(t)$ wordt dan uiteindelijk

$$p_{ij}(t) = \left(1 - \frac{1}{N_l} - a\right)^t \left(p_{ij}(0) - \frac{aN_l}{1 + aN_l}\right) + \frac{aN_l}{1 + aN_l}. \quad (3.5)$$

De vorm van bovenstaande uitdrukking geeft veel informatie prijs over hoe het algoritme de structuur van het netwerk beïnvloedt. De factor $1 - \frac{1}{N_l} - a$ is bijvoorbeeld steeds kleiner dan 1 en (voor $\frac{N_l}{N_n(N_n-1)} \ll 1$) groter dan 0. Naarmate de tijd vordert, vindt er dus een relaxatie plaats naar de tweede term $\frac{aN_l}{1 + aN_l}$. De tijdconstante van deze relaxatie kan gevonden worden door eerder vermelde voorfactor te schrijven als een exponentieel verval:

$$\begin{aligned} \left(1 - \frac{1}{N_l} - a\right)^t &= \exp \left[t \ln \left(1 - \frac{1}{N_l} - a\right) \right] \\ &= \exp \left[-\frac{t}{\tau} \right] \end{aligned}$$

Voor de tijdconstante τ betekent dit

$$\tau = -\frac{1}{\ln \left(1 - \frac{1}{N_l} - a\right)}. \quad (3.6)$$

Zoals eerder gezegd, zal voor grote t het netwerk evolueren naar een toestand waar de kans op een link tussen twee nodes gelijk is aan

$$p_{ij}(\infty) = \frac{aN_l}{1 + aN_l} \leq 1. \quad (3.7)$$

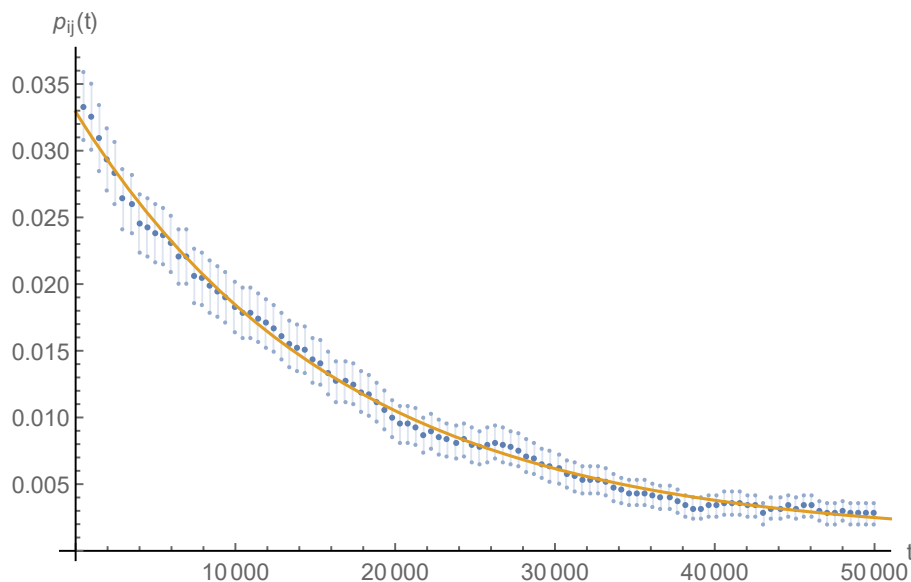
Deze kans is enkel afhankelijk van het aantal nodes en het aantal links in het netwerk en is dus gelijk voor alle nodes. Dit type netwerk is reeds bekend uit hoofdstuk 1: het Bernoulli netwerk. Het willekeurig verplaatsen van links zal dus, zoals voorspeld, *ieder* netwerk doen evolueren naar een Bernoulli netwerk. Dit gebeurt met een tijdconstante τ gegeven door vergelijking 3.6. Figuur 3.2 geeft het exponentieel verval van vergelijking 3.5 weer, samen met de betekenis van de tijdconstante τ .

3.2 Simulaties willekeurige verplaatsing

Deze sectie tracht de voorgaande resultaten te bevestigen en de verdere impact van het algoritme op andere eigenschappen van het netwerk, zoals de degree distributie en de clustering coëfficiënt. Voor deze laatste twee eigenschappen werd enkel de eindlimiet voorspeld, niet de tijdsevolutie. Vervolgens wordt een algoritme bekeken dat niet theoretisch behandeld werd, namelijk het introduceren van een bias in het verplaatsen van links.

3.2.1 De kans $p_{ij}(t)$

De kans op een link tussen twee nodes kan gevonden worden door op iedere tijdstap waarop het algoritme wordt toegepast, te kijken of desbetreffende link aanwezig is. Na dit veel te herhalen, wordt de kans op deze link gegeven door de fractie van de simulaties waar de link inderdaad aanwezig was. Aan de hand van deze methode kan het tijdsverloop, voorspeld in vergelijking 3.5, gevisualiseerd worden.



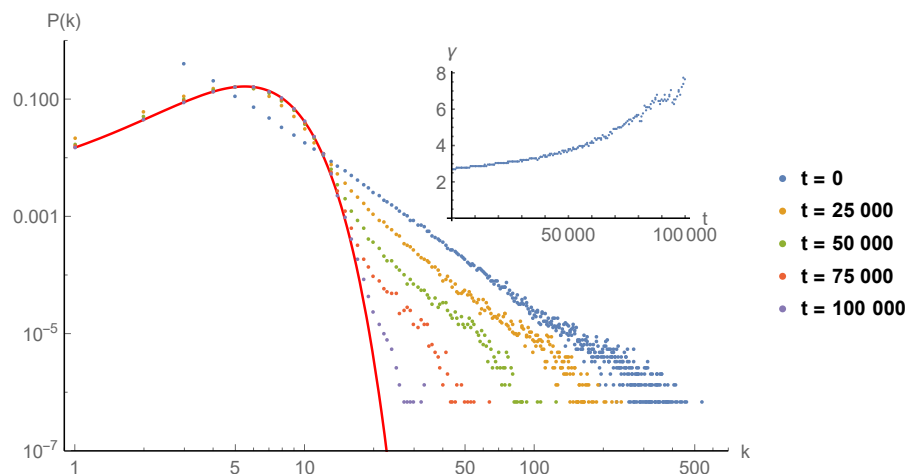
Figuur 3.3: De kans op een link tussen nodes 50 en 51 (blauw), beginnend van een schaalvrij netwerk van 5000 nodes met $m = 3$ voor 5040 simulaties. Iedere tijdstap werd een link willekeurig verplaatst, 50 000 tijdstappen lang. De standaardafwijking staat aangeduid als de lichtblauwe punten. De oranje curve duidt een fit aan van een exponentieel dalende functie met tijdconstante 16574.8 en limietwaarde $9.27291 \cdot 10^{-4}$.

Dit werd gedaan in figuur 3.3, waarbij de numerieke data (blauw met standaardafwijking: lichtblauw) gefit werd door een curve met de vorm van vergelijking 3.5 (oranje). Deze fit heeft als tijdconstante 16574.8 en als limietwaarde $9.27291 \cdot 10^{-4}$. Vooral de tijdconstante komt overeen met de voorspelling van 14981.5, de limietwaarde wijkt meer af van het theoretische resultaat ($6.0048 \cdot 10^{-4}$). Dit kan een gevolg zijn van het uitvoeren van te weinig simulaties, waardoor de foutenmarge groot is.

De fit heeft wel dezelfde vorm als de numerieke data én ligt steeds binnen de foutenmarge, wat alvast in de richting van een correcte voorspelling wijst. Natuurlijk kunnen meer (en langere) simulaties een duidelijker beeld geven van eventuele afwijkingen.

3.2.2 Tijdsverloop van de degree distributie

Aan de hand van een vergelijkbare simulatie kan ook het tijdsverloop van de degree distributie berekend worden door na iedere tijdstap de degree van alle nodes te berekenen en zo de degree distributie op te stellen. De begin- en eindtoestanden zijn gekend: zo zal een schaalvrij netwerk met een machtwet in de degree distributie evolueren naar een Bernoulli netwerk met een binomiale degree distributie.



Figuur 3.4: Enkele momentopnames van het verloop van de degree distributie tijdens het willekeurig verplaatsen van een link, beginnend van een schaalvrij netwerk van 10 000 nodes met $m = 3$. Dit is het resultaat na 150 simulaties. De rode lijn toont de verwachte limiet: een Bernoulli netwerk met $p = 0.000600$. De inzet toont de evolutie van de exponent γ van de ‘fat-tail’ ($k \geq 20$).

Figuur 3.4 toont het resultaat van dit algoritme. Er wordt begonnen van een schaalvrij netwerk (blauw), dat gaandeweg evolueert naar een Bernoulli netwerk (de rode curve is daarbij de theoretische voorspelling).

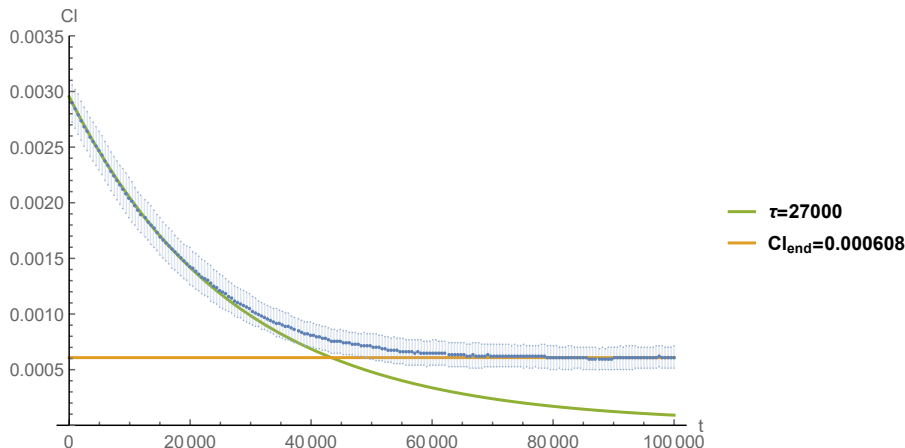
De evolutie verloopt aan verschillende snelheden voor nodes met lage en hoge degrees: nodes met lage degree evolueren zeer snel naar de verwachte degree distributie. Voor nodes met hoge degree duurt het veel langer voordat al hun links herschikt zijn volgens het Bernoulli netwerk. Hierdoor evolueert de ‘fat-tail’ van de machtwet slechts traag naar de binomiale verdeling. De exponent γ wordt groter, doordat nodes met hoge degree meer kans hebben dat één van hun links verplaatst wordt.

3.2.3 Clustering coëfficiënt

Uit dezelfde simulatie kan ook de clustering coëfficiënt bepaald worden. Deze wordt uitgemiddeld bij eenzelfde tijd over alle simulaties, waardoor er ook een standaardafwijking op kan worden bepaald. Figuur 3.5 toont deze data, samen met twee fits: één exponentiële fit voor de eerst 40 datapunten en een tweede horizontale fit voor de laatste 20 datapunten. Beide fits geven belangrijke informatie prijs over de evolutie van de clustering coëfficiënt.

De exponentiële fit met een tijdconstante van 27000 geeft aan dat het verloop van de clustering coëfficiënt snel afwijkt van het exponentieel verval, om te stabiliseren volgens de horizontale fit (met als waarde 0.000608). Deze limietwaarde ligt zeer dicht bij de eerder voorspelde theoretische waarde $p = 0.000600$, welke in een Bernoulli netwerk gelijk is aan de clustering coëfficiënt.

Het afwijkende gedrag van de clustering coëfficiënt hangt samen met het eerder besproken verloop van de degree distributie. In figuur 3.4 werd duidelijk gemaakt dat nodes met een lage degree (welke zeer talrijk zijn in een schaalvrij netwerk) snel de binomiale verdeling opbouwen voor kleine degrees. Deze nodes lijken het grootste deel van de clustering coëfficiënt te bepalen. Vanaf zij binomiaal verdeeld zijn, begint de clustering coëfficiënt uit te vlakken. De minder talrijke nodes met hoge degree hebben bijgevolg slechts een kleine impact op de waarde van de clustering coëfficiënt.



Figuur 3.5: De evolutie van de clustering coëfficiënt voor de simulaties uit figuur 3.4. De groene fit aan de eerst 40 datapunten toont een exponentieel verval, de horizontale fit (oranje) aan de laatste 20 datapunten toont de limietwaarde.

3.3 Simulatie met bias

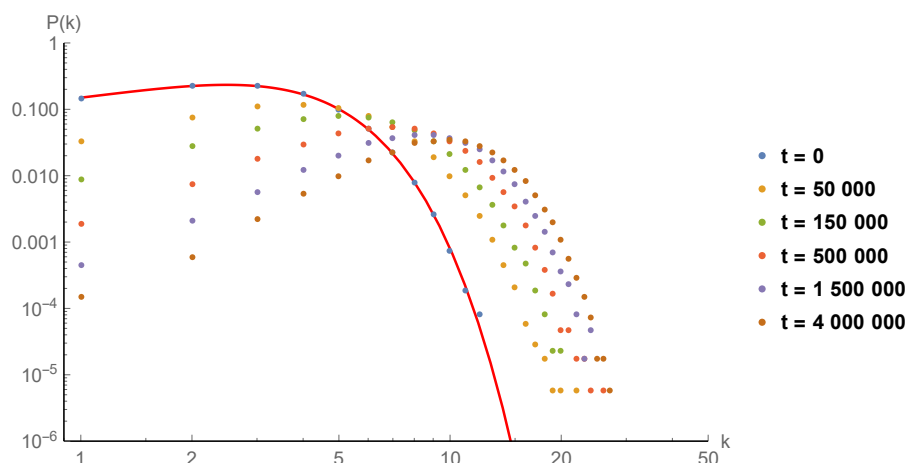
Het kiezen van de te verplaatsen link hoeft niet per se volledig willekeurig te gebeuren. De waarschijnlijkheid waarmee de te verplaatsen link gekozen wordt, kan bijvoorbeeld afhangen van de degrees van de nodes die de link verbindt. De nieuwe plek van de gekozen link, kan op een andere manier afhangen van de degree's van de te verbinden nodes (of van andere eigenschappen). Dit opent dus een waaier aan verschillende mogelijkheden waarop het netwerk gemanipuleerd kan worden.

In wat volgt wordt één van deze mogelijkheden gesimuleerd: links die tussen nodes met hoge degree liggen worden met prioriteit gekozen. Deze links worden daarna verplaatst naar een lege plek die ook weer een hogere waarschijnlijkheid heeft om tussen twee nodes met reeds hogere degree te liggen. Dit proces zal verder preferentiële verplaatsing genoemd worden.

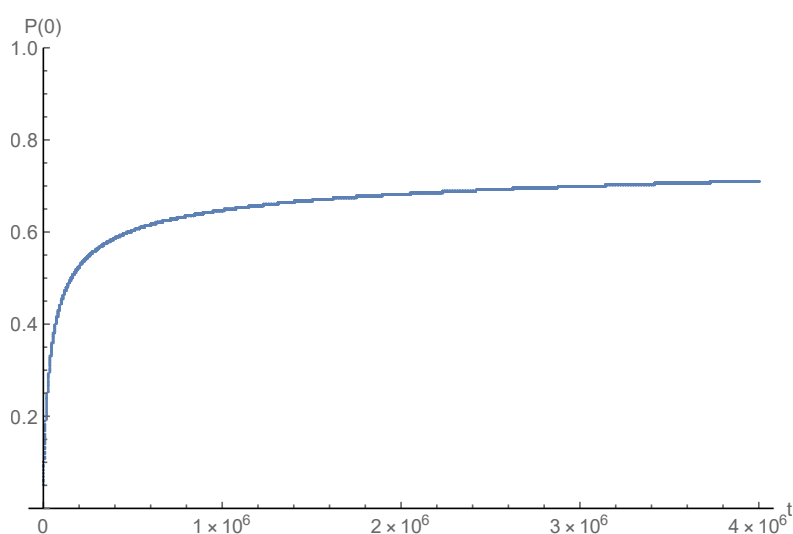
De resultaten van 35 dergelijke simulaties worden weergegeven in figuur 3.6, waarbij begonnen werd vanaf een Bernoulli netwerk van 5000 nodes en $p = 0.0006$ (verduidelijkt door de rode curve). De degree distributie lijkt naar hogere degrees te evolueren, wat te verwachten valt, aangezien nodes met reeds hoge degree links blijven accumuleren. Er volgt echter geen schaalvrij netwerk (zolang de simulatie liep), de degree distributie krijgt een gespitste vorm die smaller wordt naarmate de tijd vordert.

Dit proces blijft vertragen naarmate de tijd vordert, wat duidelijk wordt uit de tijdsaanduidingen bij de verschillende datapunten. Dit kan een gevolg zijn van de preferentiële verplaatsing die zichzelf tegenwerkt: links die net verplaatst werden naar plekken tussen nodes met hoge degree, hebben een hoge waarschijnlijkheid om opnieuw verplaatst te worden, terwijl de rest van het netwerk onveranderd blijft.

Hoewel de degree distributie meer gewicht lijkt te geven aan hogere degrees, blijft de gemiddelde degree gelijk aan 3. Dit is het gevolg van een groeiende fractie nodes met degree 0, wat niet zichtbaar is op log-log schaal. Figuur 3.7 toont deze groeiende fractie doorheen de tijd. Deze geïsoleerde nodes zijn het resultaat van verplaatste links die oorspronkelijk een node met lage degree en een node met hoge degree verbonden. De node met lage degree verliest zo al haar links, waardoor ze geïsoleerd wordt en geen kans meer heeft om links te accumuleren.



Figuur 3.6: Het verloop van de degree distributie bij preferentiële verplaatsing, startend van een Bernoulli netwerk van 5000 nodes en $p = 0.0006$. De rode curve verduidelijkt de degree distributie van de begintoestand. Dit is het resultaat na 35 simulaties.

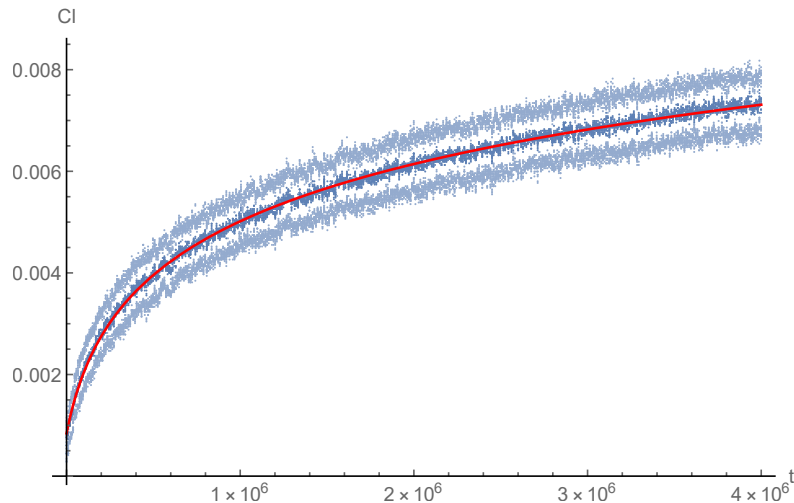


Figuur 3.7: De fractie aan nodes met degree 0 in de simulatie van figuur 3.6. Deze fractie zorgt ervoor dat de gemiddelde degree in het netwerk constant blijft.

Ten slotte kan de clustering coëfficiënt ook voor dit proces op ieder tijdstip bepaald worden (figuur 3.8). Daarbij wordt een duidelijke groei zichtbaar die (volgens de rode fit) logaritmisches verloopt:

$$Cl(t) = 0.00174379 \ln[t + 200.841] - 0.00840501 \quad (3.8)$$

Er is dus een duidelijk verschil met de willekeurige verplaatsing, waar de clustering coëfficiënt weldegelijk de voorspelde waarde aannam. Hier lijkt er geen duidelijke eindtoestand zichtbaar. Daarom zouden langere simulaties van preferentiële verplaatsing duidelijkheid moeten scheppen over de verdere evolutie van het netwerk, haar eigenschappen en eventuele evenwichtstoestanden.



Figuur 3.8: De evolutie van de clustering coëfficiënt (blauw) in de simulatie van figuur 3.6. De standaardafwijking op de clustering coëfficiënt staat aangeduid in een lichtblauwe kleur. De rode curve toont een logaritmische fit van de data, gegeven door vergelijking 3.8.

3.4 Link verwijdering bij toevoeging van nodes

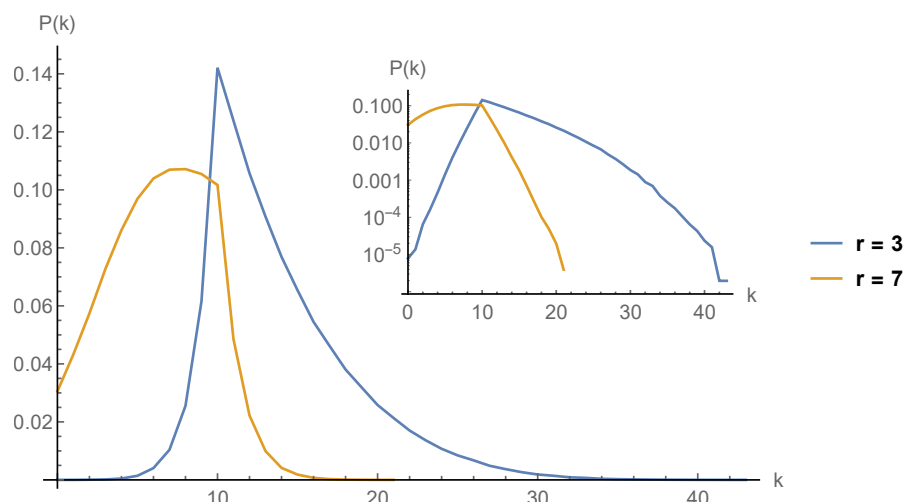
In de vorige secties van dit hoofdstuk evolueerde een netwerk door links te verwijderen en weer toe te voegen volgens voorgeschreven regels. Dit verwijderen en toevoegen werd zo ingesteld dat er geen verandering plaatsvond in het aantal links of nodes. In deze secties worden *nodes* toegevoegd, in plaats van links. Het verwijderingsproces blijft via links. Het idee is dat de (preferentiële of random) groei van het netwerk ‘verstoord’ wordt door links te verwijderen. Deze verwijdering kan ook weer willekeurig (random) gebeuren, of met een grotere kans voor links die nodes met hoge degree verbinden (preferentieel).

Om de groei van het netwerk te verzekeren en niet met een verzameling geïsoleerde nodes te eindigen, wordt op iedere tijdstap de nieuwe node met meer links aangehecht dan er links worden verwijderd. Meer bepaald wordt het aantal links per nieuwe node steeds ingesteld op $m = 10$, met het aantal verwijderde links zowel $r = 3$ als $r = 7$. Er wordt voor ieder scenario een uitmiddeling gedaan over 1000 simulaties.

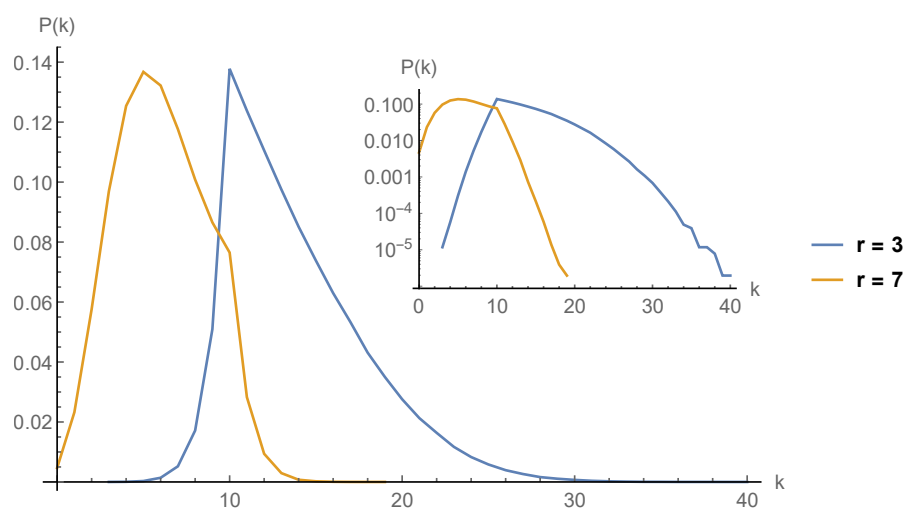
3.4.1 Random groei

Figuur 3.9 toont de degree distributie voor een willekeurige verwijdering van 3 en 7 links in een random groeiend netwerk. Het resultaat is voldoende gestabiliseerd na 500 tijdstappen. Er is voor beide gevallen een duidelijke ‘knik’ zichtbaar bij $k = 10$. Dit is niet verrassend, aangezien de nodes met degree kleiner dan $m = 10$ enkel gegenereerd kunnen worden door het verwijderen van links. Voor een kleine hoeveelheid verwijderde links lijkt de degree distributie < 10 exponentieel te stijgen, terwijl de degree distributie > 10 exponentieel daalt bij het grotere aantal verwijderde links.

Het scenario waarbij de links preferentieel verwijderd worden tijdens de random groei van het netwerk, staat weergegeven in figuur 3.10. Deze degree distributie toont veel gelijkenissen met de random verwijdering, al zijn er enkele verschillen. Zo is de degree distributie voor kleine degrees sterker gepiekt bij de verwijdering van 7 links per tijdstap. Voor beide hoeveelheden van verwijderde links is de fractie geïsoleerde nodes kleiner dan in het geval van random verwijdering hierboven.



Figuur 3.9: De degree distributie na 500 tijdstappen bij het verwijderen van 3 (blauw) en 7 (oranje) links per stap. De inzet toont dezelfde figuur op log-schaal.



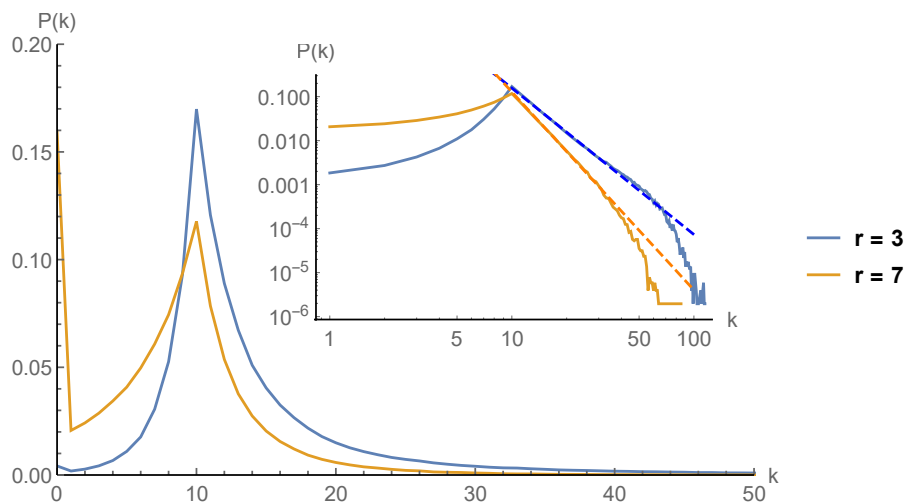
Figuur 3.10: De degree distributie na 500 tijdstappen van random groei en preferentiële verwijdering van 3 (blauw) en 7 (oranje) links. De inzet toont alweer dezelfde figuur op log-schaal.

3.4.2 Preferentiële groei

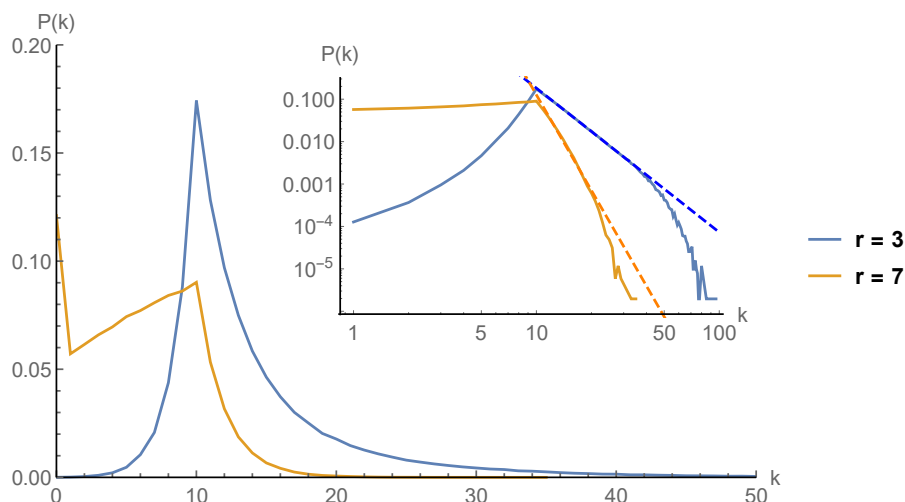
Ook in het geval van een preferentieel groeiend netwerk en random verwijdering is dezelfde kenmerkende knik aanwezig bij $k = 10$ (figuur 3.11). Wat meteen opvalt, is de grote fractie aan nodes met degree 0, bij de verwijdering van 7 links per tijdstap. De grootte van de grootste samenhangende component in het netwerk is hier nog steeds ongeveer 400 nodes, zodat deze nog altijd representatief is voor het netwerk.

Vervolgens is voor beide hoeveelheden van verwijderde links een machtwet voor degrees > 10 plausibel. Om dit te kunnen onderscheiden (en de exponent γ betrouwbaar te bepalen), zijn er echter simulaties nodig bij grotere netwerken.

Dezelfde conclusies gelden voor preferentiële verwijdering in een preferentieel groeiend netwerk (figuur 3.12). De verwijdering van 3 links heeft enkel een kleinere fractie met degree < 10 . Bij verwijdering van 7 links is er een groter verschil met random verwijdering: er zijn minder nodes met degree 0 en er is een meer uniforme verdeling van nodes met degree < 10 .



Figuur 3.11: Het resultaat na 500 stappen van preferentiële groei ($m = 10$) en random verwijdering van 3 (blauw) en 7 (oranje) nodes. De inzet is dit keer op log-log schaal. De onderbroken lijnen tonen een fit met $\gamma = 3.33$ voor $r = 3$ en $\gamma = 4.50$ voor $r = 7$.



Figuur 3.12: De degree distributie voor preferentiële groei ($m = 10$) én verwijdering van 3 (blauw) en 7 (oranje) links. De grafiek wordt in de inzet op log-log schaal getoond. De twee fits (onderbroken lijnen) geven $\gamma = 3.4$ voor $r = 3$ en $\gamma = 7.48$ voor $r = 7$.

3.5 Conclusies link herschikking

Hoofdstuk 3 behandelde dus het effect van de verplaatsing of verwijdering van links op een netwerk. In sectie 3.1 werd een theorie opgebouwd rond het willekeurig verplaatsen van links, zonder het aantal nodes en links te veranderen. Sectie 3.2 bevestigde deze theoretische resultaten beginnend van een schaalvrij netwerk en beschreef verder de evolutie van de degree distributie en clustering coëfficiënt, waar toch afwijkend gedrag werd geobserveerd.

Dezelfde simulatie werd herhaald in sectie 3.3 met dit keer een bias voor nodes met hoge degree in het verplaatsingsproces en toegepast op een Bernoulli netwerk. Ook hier werd gekeken naar degree distributie en clustering coëfficiënt.

Tot slot werd in sectie 3.4 het netwerk toegelaten te groeien, terwijl er links verwijderd werden. Terwijl netto groei werd verzekerd, werden er 4 combinaties van groei- en verwijderingsprocessen toegepast. Nadat het netwerk een steady-state leek bereikt had, werd de degree distributie voor de 4 algoritmes beschreven. Het verwijderen van links tijdens de groei blijkt in staat om de macht γ van BA netwerken te veranderen naar een andere waarde dan $\gamma = 3$, wat nuttig kan zijn bij het simuleren van sociale of economische netwerken, met typisch $2 < \gamma < 3$.

Hoofdstuk 4

Conclusie

Het eerste hoofdstuk van deze thesis werd gewijd aan het introduceren van de nodige kennis omtrent netwerk theorie, in het bijzonder het Barabási-Albert model voor schaalvrije netwerken. Dit model geeft een resultaat voor de verwachte degree van een node en voorspelt bovendien de verwachte degree distributie voor een netwerk dat groeit volgens preferentiële aanhechting. Deze resultaten zijn het gevolg van twee aannames: mean-field theorie en de continue tijdsbenadering.

De mean-field benadering werd in hoofdstuk 2 achterwege gelaten in een poging de kans te bepalen dat een node een bepaalde degree heeft op dat tijdstip. Op die manier konden bijvoorbeeld de fluctuaties rond de verwachte degree bepaald worden.

Deze kansverdeling werd succesvol gevonden voor zowel preferentieel, als niet-preferentieel groeiende netwerken. Dit resultaat werd bekomen aan de hand van het opstellen van een master equation die werd herleid tot een (partiële) differentiaalvergelijking in de genererende functie. Deze genererende functie gaf dan uiteindelijk de gewenste kansverdeling als de coëfficiënten van haar machtreeks. Voor het niet-preferentieel groeiende netwerk werd de verwachte Poissonverdeling teruggevonden, voor het schaalvrij netwerk werd de tot nu toe onbekende kansverdeling gevonden. Bij beide groeimethoden werden verwachtingswaarde en variantie bepaald. Ook werd bevestigd dat de kansverdelingen correct genormeerd zijn. Voor het schaalvrij netwerk werd als verwachtingswaarde het resultaat van het Barabási-Albert model teruggevonden. Bovendien werd voor het niet-preferentieel groeiende netwerk een kansverdeling afgeleid voor de wachttijd. De verwachtingswaarde van deze kansverdeling kwam echter niet overeen met het resultaat van het mean-field model.

De theoretische resultaten voor het schaalvrij netwerk werden vervolgens bevestigd aan de hand van simulaties. Hierbij werden zowel de resultaten voor verwachtingswaarde, variantie en de volledige kansverdeling gecontroleerd en bevestigd. Eventuele afwijkingen van het theoretische resultaat konden toegekend worden aan de eerste fractie nodes die een grotere degree hadden dan voorspeld. Latere nodes, die in grote netwerken de bulk van de nodes vormen, zorgden voor een mooie overeenkomst met de eerder afgeleide vergelijkingen.

Hoofdstuk 3 spitste zich toe op het beschrijven van een herschikking van links. Daarbij werd het aantal links en nodes in een netwerk constant gehouden, maar werden de links wel willekeurig verplaatst. Door een master equation op te stellen voor de kans op een link tussen twee nodes, kon via de genererende functiemethode een theoretische beschrijving worden geformuleerd. Het schrijven van dit theoretisch resultaat als een exponentieel verval maakte het mogelijk een tijdconstante en limietwaarde te formuleren. De limietwaarde voorspelde dan, zoals verwacht, een Bernoulli netwerk als eindtoestand.

Beginnend van een schaalvrij netwerk werden deze resultaten weer gecontroleerd aan de hand van simulaties. Het verloop van de kans op een link tussen twee nodes werd bevestigd aan de hand van de vorm van de curve, al waren er meer (en langere) simulaties nodig om de standaardafwijking te verkleinen en een sluitend resultaat te geven.

Het verloop van de degree distributie werd bekeken, waarbij vastgesteld werd dat deze inderdaad evolueert naar de verwachte binomiaal verdeling van een Bernoulli netwerk. Bovendien werd er opgemerkt dat nodes met lage degree zich zeer snel schikken naar de verwachte verdeling. De 'fat-tail' van het schaalvrije netwerk (nodes met zeer grote degree) heeft meer tijd nodig om te evolueren naar de eindtoestand.

De clustering coëfficiënt gedraagt zich afwijkend van de voorspelling. Doordat deze voornamelijk bepaald wordt door de talrijke nodes met lage degree, welke zeer snel de eindtoestand bereiken, nadert de clustering coëfficiënt haar limietwaarde sneller dan voorspeld. De uiteindelijke waarde kwam wel overeen met de theoretisch berekende waarde voor het Bernoulli netwerk.

Vervolgens werd een soortgelijk link-verplaatsend algoritme toegepast op een Bernoulli netwerk, waarbij links tussen nodes met hoge degree een hogere kans hadden om verplaatst te worden. De verplaatsing gebeurde daarna ook bij voorkeur naar een lege plek tussen twee nodes met reeds hoge degree. De evolutie van het netwerk verliep uitzonderlijk traag. Vermoedelijk had het algoritme een hoge kans om steeds dezelfde links te verplaatsen, zodat de rest van het netwerk weinig veranderde.

De degree distributie werd wel steeds meer gepiekt en schoof (traag) op naar hogere degrees. Ongeacht de groei van de fractie nodes met hoge degree, bleef de gemiddelde degree over het netwerk constant. Een stijgende fractie nodes met degree 0 bleek daarvoor de oorzaak.

De clustering coëfficiënt leek logaritmisch te stijgen, hoewel door de trage voortgang van het algoritme niet verder kon worden gezocht naar een eventuele limietwaarde.

Ten slotte werden de twee eerder beschouwde groeiprocessen (preferentieel en random) aangepast, zodat op iedere tijdstap een vast aantal links random of preferentieel verwijderd werd. Het verwijderde aantal links werd zo gekozen dat het netwerk nog wel kon groeien. De degree distributies van deze vier simulaties toonden allen een 'knik' bij het aantal links waarmee een nieuwe node zich aanhechtte. Voor verdere, sluitende, conclusies zijn meer simulaties nodig waarbij grotere netwerken gebruikt worden en meer eigenschappen bekeken worden (zoals clustering coëfficiënt, diameter, gemiddelde padlengte ...).

Bijlagen

Bijlage A

Genererende functiemethode

Deze appendix schetst het gebruik van de genererende functie zoals Herbert S. Wilf dat doet in zijn boek *generatingfunctionology* [18]. Dit boek geeft een uitgebreide basis over de genererende functie en toont hoe zij als brug gebruikt kan worden tussen discrete en continue wiskunde. Wat volgt, is een kleine samenvatting van de hier relevante delen van het boek¹.

A.1 Formele machtreeks

Sommige problemen, zoals telproblemen, zijn discreet van aard. Dat wil zeggen, de oplossing is geen continue functie, maar een rij getallen $a_0, a_1, a_2 \dots$. Een algemene uitdrukking voor een getal a_n uit deze rij is dan zeer aantrekkelijk en vereenvoudigt de berekening van een willekeurig getal uit de rij.

Zulke algemene uitdrukkingen zijn echter niet altijd meteen af te leiden uit, bijvoorbeeld, een gegeven recursierelatie. Wat misschien wel mogelijk is, is een uitdrukking voor de machtreeks, waarvan de coëfficiënten de gezochte rij zijn. Deze machtreeks wordt gedefinieerd als de *genererende functie*:

$$G(x) = \sum_{n=0}^{\infty} a_n x^n \quad (\text{A.1})$$

Om het in Wilf's eigen woorden te zeggen: "A generating function is a clothesline on which we hang up a sequence of numbers for display."

De genererende functie moet daarbij vaak als *formele* machtreeks gezien worden: een algebraïsch object, gedefinieerd door haar coëfficiënten $\{a_n\}_0^{\infty}$. Dit zonder de kennis of de reeks effectief convergeert naar een functie. Zo is bijvoorbeeld de reeks

$$f = 1 + x + 2x^2 + \dots + n!x^n + \dots$$

goed gedefinieerd als formele machtreeks, terwijl ze niet convergeert voor alle $x \neq 0$. Zo kan er dus zonder de eis van convergentie met de genererende functie gewerkt worden.

Bijgevolg moeten verschillende operaties opnieuw gedefinieerd worden voor de formele machtreeks:

¹De tweede editie van het boek is, voor educatief gebruik, gratis te downloaden op <https://www.math.upenn.edu/~wilf/DownldGF.html>.

- **Som (en verschil):**

$$\sum_n a_n x^n \pm \sum_n b_n x^n = \sum_n (a_n \pm b_n) x^n \quad (\text{A.2})$$

- **Product:** Aan de hand van de Cauchy productregel geeft dit

$$\sum_n a_n x^n \sum_n b_n x^n = \sum_n \left(\sum_k a_k b_{n-k} \right) x^n \quad (\text{A.3})$$

- **Reciproke:** Indien $a_0 \neq 0$ voor de machtreeks f , geldt er

$$\frac{1}{f} = \sum_n b_n x^n, \quad (\text{A.4})$$

met $b_0 = \frac{1}{a_0}$ en voor alle $b_{n \geq 1}$:

$$b_n = -\frac{1}{a_0} \sum_{k \geq 1} a_k b_{n-k}.$$

- **Afgeleide:** De afgeleide wordt in dit geval niet gedefinieerd aan de hand van een limiet, aangezien er geen gebruik kan worden gemaakt van de analytische eigenschappen van de genererende functie. Daarom:

$$\frac{df}{dx} = \sum_n n a_n x^{n-1} \quad (\text{A.5})$$

Deze definitie van de afgeleide heeft dezelfde rekenregels als de analytische versie, maar is gedefinieerd voor formele machtreeksen.

A.2 Oplossen van recursierelaties

Zoals in deze thesis ook duidelijk wordt, komen genererende functies van pas bij het oplossen van recursierelaties. Wilf geeft zelfs een stappenplan tot het vinden van de genererende functie, gegeven een recursie voor a_n :

1. Bepaal de waarden die de index n kan aannemen, d.w.z. waarvoor de recursie geldt.
2. Definieer de genererende functie en geef ze een naam (bijvoorbeeld $G(x) = \sum_n a_n x^n$).
3. Vermenigvuldig beide zijden van de recursierelatie met x^n en sommeer over alle toegelaten waarden van n .
4. Schrijf beide zijden van de vergelijking in termen van de genererende functie $G(x)$.
5. Los deze vergelijking op naar de genererende functie.

Indien er een exacte uitdrukking voor de coëfficiënten gezocht wordt, kan er nog een stap worden toegevoegd:

6. Probeer de uitdrukking voor de genererende functie te schrijven als een machtreeks. De uitdrukking voor de coëfficiënten geeft dan een formule voor a_n .

Als voorbeeld kan een algemene uitdrukking voor de rij van Fibonnaci worden afgeleid. De recursierelatie wordt hierbij voor $n \geq 1$ gegeven door

$$F_{n+1} = F_n + F_{n-1}, \quad (\text{A.6})$$

met $F_0 = 0$ en $F_1 = 1$.

De index n is bijgevolg een natuurlijk getal, beginnend vanaf 0. De genererende functie kan dan worden gedefinieerd als

$$F(x) = \sum_{n=0}^{\infty} F_n x^n. \quad (\text{A.7})$$

Na het uitvoeren van stap 3, wordt de recursierelatie

$$\sum_{n=0}^{\infty} F_{n+1} x^n = \sum_{n=0}^{\infty} (F_n + F_{n-1}) x^n.$$

Aangezien

$$\begin{aligned} \sum_{n=0}^{\infty} F_{n+1} x^n &= \frac{F(x) - x}{x} \\ \sum_{n=0}^{\infty} F_{n-1} x^n &= xF(x), \end{aligned}$$

wordt de vergelijking voor de genererende functie

$$\frac{F(x) - x}{x} = (1 + x)F(x), \quad (\text{A.8})$$

met als oplossing

$$F(x) = \frac{x}{1 - x - x^2}. \quad (\text{A.9})$$

De noemer van deze oplossing kan ontbonden worden in $1 - x - x^2 = (1 - r_+x)(1 - r_-x)$, met $r_{\pm} = \frac{1 \pm \sqrt{5}}{2}$. Aan de hand van splitsen in partieelbreuken wordt de genererende functie dan

$$\begin{aligned} F(x) &= \frac{x}{(1 - r_+x)(1 - r_-x)} \\ &= \frac{1}{r_+ - r_-} \left(\frac{1}{1 - r_+x} - \frac{1}{1 - r_-x} \right) \\ &= \frac{1}{\sqrt{5}} \sum_{n=0}^{\infty} (r_+^n - r_-^n) x^n. \end{aligned}$$

Dit geeft als uitdrukking voor het n -de Fibonnaci getal

$$F_n = \frac{1}{\sqrt{5}} (r_+^n - r_-^n). \quad (\text{A.10})$$

A.3 Bepalen van momenten

Indien de coëfficiënten $\{a_n\}_0^{\infty}$ een kansverdeling definiëren, kan de genererende functie gebruikt worden om, bijvoorbeeld, het gemiddelde (ofwel het eerste moment) uit te rekenen. Het gemiddelde is gedefinieerd als

$$\mathbb{E}[a_n] = \sum_n n a_n, \quad (\text{A.11})$$

indien de coëfficiënten genormaliseerd zijn ($\sum_n a_n = 1$). Definieer de genererende functie nu als $G(x) = \sum_n a_n x^n$, dan is het eenvoudig in te zien dat, wegens de normering, $G(1) = 1$. Aan de hand van de definitie van de afgeleide van een formele machtreeks (vergelijking A.5), wordt het gemiddelde:

$$\mathbb{E}[a_n] = \left. \frac{dG(x)}{dx} \right|_{x=1} \quad (\text{A.12})$$

De standaardafwijking $\sigma^2[a_n]$ hangt van het tweede moment van de kansverdeling af:

$$\sigma^2[a_n] = \mathbb{E}[a_n^2] - (\mathbb{E}[a_n])^2 \quad (\text{A.13})$$

Met als tweede moment

$$\mathbb{E}[a_n^2] = \sum_n n^2 a_n. \quad (\text{A.14})$$

Aangezien

$$x \frac{dG(x)}{dx} = \sum_n n a_n x^n,$$

kan het tweede moment ook berekend worden aan de hand van

$$\mathbb{E}[a_n^2] = \left. \frac{d}{dx} \left[x \frac{dG(x)}{dx} \right] \right|_{x=1} \quad (\text{A.15})$$

Bijlage B

Mathematica implementatie

De implementaties werden geschreven in *Mathematica 11.0*. Hieronder de belangrijkste functies waarmee de simulaties tot stand werden gebracht.

B.1 Random groei

```
In[1]:= addNodeRandom[network_, m_] := Block[
  {vertices = VertexList@network,
   nodes,
   vertex
  },
  (* Kies m willekeurig nodes in het netwerk *)
  nodes = RandomSample[vertices][[1 ;; m]];
  vertex = Last@vertices + 1;
  If[Length@nodes > 0,
    EdgeAdd[
      VertexAdd[network, vertex],
      Table[vertex <-> nodes[[i]], {i, Length@nodes}
    ],
    network]
];
```

B.2 Preferentiële groei

```
In[2]:= addNodePreferential[network_, m_] := Block[
  {edges = Flatten[List @@@ EdgeList@network], (*
   Lijst met indices van nodes die even vaak voor komen als hun degree*)
   nodes,
   vertex,
   choice
  },
  vertex = Last@VertexList@network + 1;
  (*Kies willekeurig nodes uit de eerdere lijst nodes en verwijder ze
```

```

vervolgens om dubbel trekken te vermijden*)
nodes = Flatten@Last@Reap@
  Do[{
    choice = RandomChoice@edges,
    Sow[choice],
    edges = DeleteCases[edges, choice]
  },
  {m}];
If[Length@nodes > 0,
  EdgeAdd[
    VertexAdd[network, vertex],
    Table[vertex <-> nodes[[i]], {i, Length@nodes}]]
  ],
network]
];

```

B.3 Random link verwijdering

```

In[3]:= deleteLinkRandom[network_] := Block[
  {vertices = VertexList@network,
  edges,
  node
  },
  node = RandomChoice[vertices];
  edges = EdgeList[network, node <-> _];
  If[Length@edges != 0, EdgeDelete[network, RandomChoice[edges]],
  network]
];

```

B.4 Preferentiële link verwijdering

```

In[4]:= deleteLinkPreferential[network_] := Block[
  {edges = EdgeList@network},
  If[Length@edges != 0, EdgeDelete[network, RandomChoice[edges]],
  network]
];

```

B.5 Voorbeeld simulatie

```

In[5]:= simulateRP[n_, m_, r_] := Block[
  {lastgraph,
  degreeList = ConstantArray[1, n]},
  lastgraph =
  CompleteGraph[m]; (*Begin vanaf een volledig verbonden netwerk*)
  degreeList[[1]] = VertexDegree[lastgraph];
  Do[

```



```

lastgraph =
  addNodeRandom[Nest[deleteLinkPreferential, lastgraph, r],
    m]; (*Verwijder links, voeg daarna de node toe*)
degreeList[[i]] =
  VertexDegree@lastgraph, (*Bereken alle degrees in het netwerk*)
  {i, 1, n}];
{lastgraph, degreeList}
];

```

B.6 Willekeurige verplaatsing

```

In[6]:= relocateR[network_] := Block[
  {edgelist = EdgeList@network,
  choice,
  nodes = VertexList@network,
  node1,
  node2},
  choice = RandomChoice@edgelist; (* Kies willekeurige link
  om te verwijderen *)
  node1 = RandomChoice@nodes; (* Twee willekeurige nodes
  die niet verbonden zijn *)
  node2 =
  RandomChoice@
  Complement[nodes,
  Flatten[{node1, AdjacencyList[network, node1]}]];
  EdgeAdd[EdgeDelete[network, choice], node1 <-> node2]
];

```

B.7 Verplaatsing met bias

```

In[7]:= relocateP[network_] := Block[
  {edgelist = EdgeList@network,
  nodes,
  neighbours,
  weights,
  choice1,
  choice2,
  node1,
  node2},
  nodes = Flatten[List @@@ edgelist];
  choice1 =
  RandomChoice@nodes; (*Kies node met bias voor hoge orde *)
  neighbours = AdjacencyList[network, choice1];
  weights = (VertexDegree[network, #]) & /@ neighbours;
  choice2 = RandomChoice[weights -> neighbours]; (* Kies een een buur
  van choice1 met bias voor hoge orde *)
  node1 = RandomChoice@nodes;
  node2 =

```

```

RandomChoice@
Complement[nodes,
  Flatten@{node1, AdjacencyList[network, node1]}];
EdgeAdd[EdgeDelete[network, choice1 <-> choice2], node1 <-> node2]
];

```

B.8 Verplaatsing simulatie

```

In[8]:= simRelocation[network_, n_, mode_String] := Block[
  {degreedist = ConstantArray[0, Ceiling[n/500] + 1],
  clustering = ConstantArray[0, Ceiling[n/500] + 1],
  temp = network,
  func = Switch[mode,
    "P", relocateP,
    _, relocateR],
  j = 1},
  (* Bereken de degree distributie aan de hand van Tally en een
normalizering *)
  degreedist[[
    j]] = (Transpose@{#[[1]] // N, Normalize#[[2]], Total] // N}) &@
    Transpose@Tally@VertexDegree@temp;
  (* Bereken de clustering coefficient *)
  clustering[[j]] = GlobalClusteringCoefficient[temp] // N;
  Do[
    temp = func[temp];
    If[Mod[i, 500] == 0,
      j = j + 1;
      degreedist[[
        j]] = (Transpose@{#[[1]] // N,
        Normalize#[[2]], Total] // N}) &@
        Transpose@Tally@VertexDegree@temp;
      clustering[[j]] = GlobalClusteringCoefficient[temp] // N,
    {i, n}];
  {temp, degreedist, clustering}
];

```

Bibliografie

- [1] Alexander A. Alemi e.a. „You can run, you can hide: The epidemiology and statistical mechanics of zombies”. In: *Physical Review E* 92.5 (2015), p. 052801.
- [2] Albert-László Barabási en Réka Albert. „Emergence of Scaling in Random Networks”. In: *Science* 286 (1999), p. 509–512.
- [3] J. L. Coolidge. „The Story of the Binomial Theorem”. In: *American Mathematical Monthly* 56.3 (1949), p. 147–157.
- [4] Paul Erdős en Alfréd Rényi. „On Random Graphs I.” In: *Publicationes Mathematicae* 6 (1959), p. 290–297.
- [5] Michalis Faloutsos, Petros Faloutsos en Christos Faloutsos. „On Power-Law Relationships of the Internet Topology”. In: *ACM SIGCOMM* 29.4 (1999), p. 251–262.
- [6] Agata Fronczak, Piotr Franczak en Janusz A. Holyst. „Average path length in random networks”. In: *eprint arXiv:cond-mat/0212230* (2002). eprint: [cond-mat/0212230](https://arxiv.org/abs/cond-mat/0212230).
- [7] E. N. Gilbert. „Random Graphs”. In: *Annals of Mathematical Statistics* 30.4 (1959), p. 1141–1144.
- [8] Matthew O. Jackson. *Social and Economic Networks*. Princeton University Press, 2008.
- [9] J. Leskovec, J. Kleinberg en C. Faloutsos. „Graph Evolution: Densification and Shrinking Diameters”. In: *ACM TKDD* 1.1 (2007).
- [10] Stanley Milgram. „The Small-World Problem”. In: *Psychology Today* 1.1 (1967), p. 61–67.
- [11] M. E. J. Newman. „Scientific collaboration networks. I. Network construction and fundamental results”. In: *Phys. Rev. E* 64.1 (2001), p. 016131.
- [12] M. E. J. Newman. „The structure and function of complex networks”. In: *SIAM Rev.* 45.2 (2003), p. 167–256.
- [13] Derek J. de Solla Price. „A General Theory of Bibliometric and Other Cumulative Advantage Processes”. In: *ASIS&T* 27.5-6 (1976), p. 292–306.
- [14] Derek J. de Solla Price. „Networks of Scientific Papers”. In: *Science* 149.3683 (1965), p. 510–515.
- [15] *Modeling the Sars outbreak in Hong Kong with small world or scale free networks*. NOLTA 2004. 2004, p. 581–584.
- [16] Duncan J. Watts en Steven H. Strogatz. „Collective dynamics of 'small-world' networks”. In: *Nature* 393 (1998), p. 440–442.
- [17] Eric W. Weisstein. *Incomplete Gamma Function*. MathWorld—A Wolfram Web Resource. 2016. URL: <http://mathworld.wolfram.com/IncompleteGammaFunction.html>.
- [18] Herbert S. Wilf. *generatingfunctionology*. Academic Press, Inc., 1994.