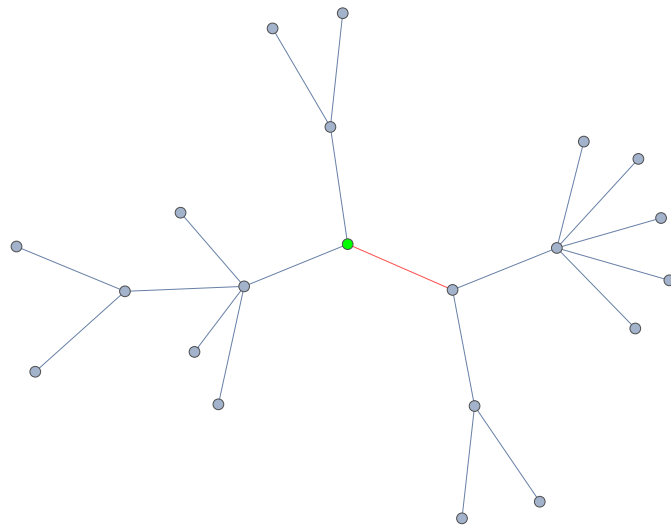


Wiskunde in de aanpak van bacteriële uitbraken



Dico Rozema
3864871

6 november 2015

Inhoudsopgave

1	Introductie	2
2	Theoretische achtergrond	2
2.1	De situatie	2
2.2	Grafen	3
2.3	Bomen	4
2.3.1	Gewortelde bomen	4
2.3.2	Ongewortelde bomen	4
2.3.3	Binaire en niet-binaire bomen	4
2.4	Phylogenetische bomen	5
3	Uitwerking	6
3.1	Maximale bijdrage	6
3.1.1	Het algoritme	7
3.2	Minimale maximale bijdrage	8
3.3	Simulatie	10
3.4	In de praktijk	11
3.4.1	<i>K. pneumoniae</i> -uitbraak NIH Clinical Center	14
4	Discussie	15
5	Conclusie	16
6	Bronnen	17

1 Introductie

In het nieuws hoort men steeds vaker over ‘ziekenhuisbacteriën’. Dit zijn bacteriën die in een ziekenhuis opduiken en daar bij verschillende mensen infecties veroorzaken. Deze ziekenhuisbacteriën komen in het nieuws omdat ze resistent zijn tegen antibiotica, waardoor ze lastig te bestrijden zijn.

De verwachting is dat zulke ziekenhuisbacteriën in de toekomst een steeds groter probleem zullen vormen. Naarmate antibiotica meer worden gebruikt, komen er steeds meer resistente stammen, en wordt het steeds lastiger om zo’n uitbraak aan te pakken. Het zal dus in de toekomst niet meer mogelijk zijn een uitbraak alleen aan te pakken door veel medicijnen toe te dienen, men zal strategischer te werk moeten gaan. Hiervoor is het belangrijk dat men de vijand kent. Er bestaan verschillende vormen van een uitbraak, en om een uitbraak snel aan te pakken is het belangrijk om er snel achter te komen met wat voor uitbraak men van doen heeft. Dit is waar deze scriptie over gaat: kan men wiskunde gebruiken om verschillende uitbraaksoorten te onderscheiden?

2 Theoretische achtergrond

2.1 De situatie

In deze scriptie zullen wij proberen om een wiskundige analyse op te stellen voor bacteriële uitbraken, zoals die zich bijvoorbeeld in een ziekenhuis voor kunnen doen. We onderscheiden hierin twee verschillende uitbraaksoorten. Ten eerste kan er sprake zijn van een puntbron. Dit wil zeggen dat alle infecties¹ éénzelfde bron hebben, bijvoorbeeld een besmette operatiekamer of een instrument dat niet goed gedesinfecteerd is. Ten tweede kan er sprake zijn van pure transmissie, wat wil zeggen dat de infectie van patiënt op patiënt wordt overgedragen. Uiteraard is het ook mogelijk om hierin tussenvormen te herkennen, waarin patiënten door een puntbron geïnfecteerd worden en vervolgens deze infectie weer aan anderen overdragen.

Het is op het eerste gezicht vaak niet mogelijk om deze twee vormen van uitbraak te herkennen. Immers, het enige wat men ziet is dat er een aantal mensen ziek wordt, waarbij uit verder onderzoek blijkt dat het hier om dezelfde soort bacterie gaat. Mocht de uitbraak zich erg wijd verspreiden, dan kan men waarschijnlijk wel stellen dat transmissie belangrijk is, echter zou men het liever niet zo ver laten komen. Daarom is het wenselijk om een methode te ontwikkelen die iets kan zeggen over het soort uitbraak dat zich voordoet, zodat de uitbraak efficiënt bestreden kan worden. Verschillende transmissieroutes vereisen immers verschillende methoden van aanpak.

Het is bij een bacteriekolonie mogelijk om het DNA uit te lezen, dit wordt ook wel *sequencing* genoemd. Wanneer men het DNA van meerdere bacteriekolonies sequencet, kan men vergelijken in hoeverre dit DNA gelijk is. Hierbij geldt dan dat bacteriekolonies met DNA dat meer op elkaar lijkt, waarschijnlijk

¹Met een infectie bedoelen wij een besmetting, niet een ontsteking.

een recentere gemeenschappelijke voorouder hebben dan diegenen die minder gelijkend DNA hebben. Bacteriën mengen hun DNA namelijk niet wanneer zij zich voortplanten, maar er vinden in de loop van de tijd wel mutaties plaats, waardoor twee gescheiden kolonies uiteindelijk steeds minder op elkaar zullen lijken.

Deze methode is ook toe te passen bij een bacteriële uitbraak. Wanneer een patiënt ziek wordt kan men de bacteriën die deze infectie veroorzaken sequencen, zodat men informatie krijgt over in hoeverre deze infectie lijkt op andere gedetecteerde infecties, met als uiteindelijk doel om een afstammingsboom te construeren van de verschillende infecties. Het is momenteel nog niet kostenefficiënt om dit standaard te doen, maar de verwachting is dat dit in de toekomst wel uit kan.

Wij gaan er hierbij wel van uit dat elke kolonie slechts één dominante stam heeft. Dit wil zeggen dat er per patiënt slechts één set DNA relevant is. Het kan in de praktijk voorkomen dat dit niet het geval is. Het wordt dan echter een stuk lastiger om verschillende uitbraakvormen uit elkaar te houden. Elke transmissieuitbraak zou dan namelijk ook te beschrijven zijn als een puntbronuitbraak waarbij de puntbron meerdere bacteriestammen huisvest. Uiteraard kunnen we wel iets zeggen over hoe waarschijnlijk dat is, maar dat gaat voor nu te ver. Voor het gemak zullen we er ook van uit gaan dat er bij elke transmissie ook mutatie plaatsvindt, zodat elke patiënt een andere bacteriestam draagt.

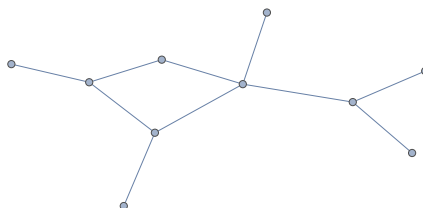
2.2 Grafen

Wij zullen dus veel gebruik maken van afstammings- oftewel phylogenetische bomen. We zullen deze later uitgebreider uitleggen, maar om dat te kunnen doen moeten we eerst definiëren wat een boom is, en om dat te doen, moeten we eerst weten wat een graaf is.

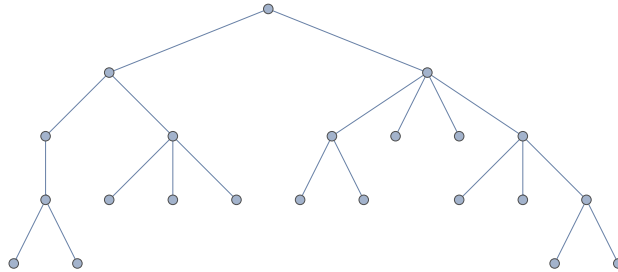
Een **graaf** is een verzameling **knopen** en **zijden**, waarbij een zijde altijd tussen twee knopen loopt. Wij zullen ons alleen bezig houden met **eenvoudige grafen**, waarbij er tussen elke twee knopen maar één zijde kan lopen, en waarbij er geen zijde mag lopen tussen een knoop en zichzelf. Twee knopen zijn **buren** wanneer er een zijde tussen deze knopen loopt. De **graad** van een knoop is het aantal buren dat hij heeft.

Een **pad** in een graaf is een rij van knopen $k_1, k_2, k_3, \dots, k_n$ waarbij er tussen elke k_i, k_{i+1} een zijde loopt, en waarbij niet tweemaal dezelfde zijde wordt gebruikt. Wanneer er tussen elke twee knopen in een graaf een pad mogelijk is, noemen we de graaf **samenhangend**. Wanneer er een pad loopt van een knoop naar zichzelf, heet dat een **cykel**.

Een **boom** is een samenhangende graaf zonder cyclen.



Figuur 1: Een voorbeeld van een graaf



Figuur 2: Een voorbeeld van een gewortelde boom

2.3 Bomen

2.3.1 Gewortelde bomen

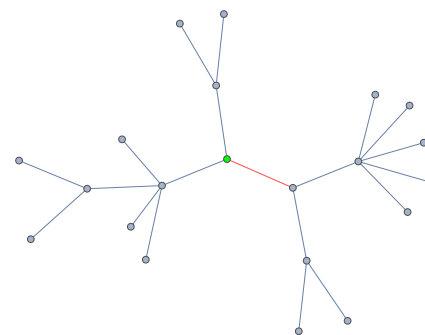
In een **gewortelde boom** is er een knoop r die wij de **stam** of de **wortel** noemen. De buren van r noemen wij de **kinderen** van r en r is hun **ouder**. De overige buren van de kinderen van r zijn *hun* kinderen. Een knoop zonder kinderen noemen wij een **blad**.

2.3.2 Ongewortelde bomen

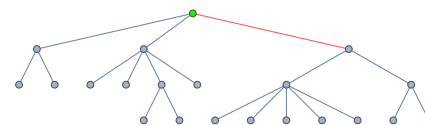
Soms gebeurt het dat we een boom hebben waarbij we niet weten wat de wortel is. Dit noemen wij een **ongewortelde boom**. Wanneer zo'n boom moet worden afgebeeld, wordt hij soms gewoon als graaf afgebeeld. Vaak wordt er echter een knoop als wortel gekozen, of maakt men een knoop aan op een van de zijden om als wortel te dienen. Bij ongewortelde bomen kunnen wij niet spreken van ouders of kinderen. Wel zeggen we dat alle knopen met graad 1 bladeren zijn. De andere knopen zijn dan **interne knopen**.

2.3.3 Binaire en niet-binaire bomen

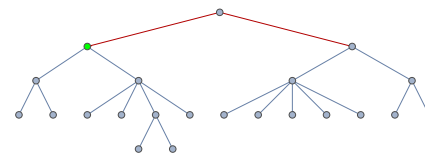
Een **binaire boom** is een boom waarin elke knoop maximaal twee



(a) Ongeworteld



(b) Geworteld op knoop



(c) Geworteld op zijde

Figuur 3: Een ongewortelde boom kan geworteld worden weergegeven door een knoop of een zijde als stam te kiezen.

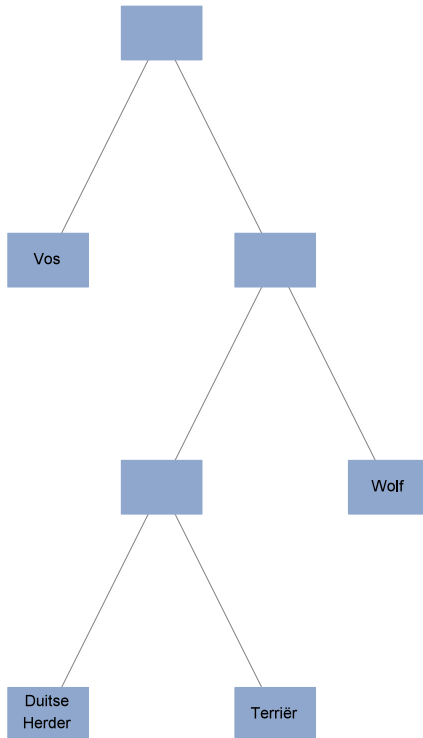
kinderen heeft. Deze definitie is echter alleen bruikbaar voor gewortelde bomen. Daarom zeggen wij dat een ongewortelde boom binair is wanneer elke knoop ten hoogste drie burens heeft.

2.4 Phylogenetische bomen

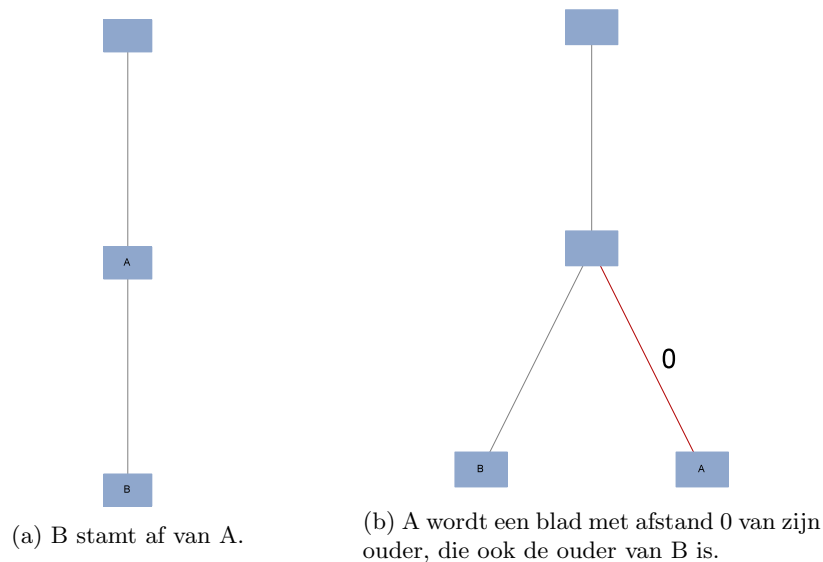
Een **phylogenetische boom** is een boom die een evolutionair proces weergeeft. Elk blad is hierbij een waargenomen soort of een waargenomen organisme, en wanneer twee knopen een ouder delen, hebben ze dus een gezamenlijke voorouder. Wanneer een phylogenetische boom geworteld is, is de wortel dus de gezamenlijke voorouder van alle waargenomen soorten die in de boom zijn meegenomen.

Er bestaan verschillende methoden om door het DNA van verschillende soorten te vergelijken, een phylogenetische boom te construeren met deze soorten als de bladeren. Wij zullen ons verder niet bezig houden met hoe dit gaat, maar het is dankzij deze mogelijkheid dat wij dit onderzoek op deze manier hebben kunnen doen. De meeste van deze methoden produceren ongewortelde bomen, de bomen waar wij mee werken zullen dan ook ongeworteld zijn.

Soms worden phylogenetische bomen weergegeven met getallen op de zijden. Zo'n getal geeft de **afstand** tussen twee knopen aan. De afstand is een manier om weer te geven in hoeverre twee stukken DNA op elkaar lijken. Er zijn verschillende definities voor afstand te geven, welke in verschillende toepassingen nuttig zijn. In alle gevallen geldt echter dat een afstand 0 aangeeft dat twee knopen identiek zijn. Dit kan bijvoorbeeld voorkomen wanneer men een niet-binaire boom tot een binaire boom omschrijft; men vervangt dan één knoop door meerdere knopen die van elkaar afstammen met afstand 0. Wij zullen het echter vooral tegenkomen door de eis dat alle waargenomen bacteriestammen bladeren moeten zijn. Hierdoor krijgen we dus bladeren met een afstand 0 tot hun ouder wanneer een waargenomen bacteriestam van een andere waargenomen bacteriestam afstamt.



Figuur 4: Een voorbeeld van een phylogenetische boom



Figuur 5: Alle waargenomen stammen zijn bladeren. Afstand 0 tussen twee knopen betekent dat deze identiek zijn.

3 Uitwerking

3.1 Maximale bijdrage

We zoeken een manier om, gegeven de phylogenetische boom van een bacteriële uitbraak, aan te geven in hoeverre deze uitbraak lijkt op een puntbronuitbraak of een transmissieuitbraak. Het is hierbij genoeg om aan te geven in hoeverre iets lijkt op een puntbronuitbraak, want wanneer er geen sprake is van een puntbron, is er transmissie. Er zijn twee vormen van puntbronnen te onderscheiden. Ten eerste zijn er niet-muterende puntbronnen, waarbij de dominante stam in de puntbron niet verandert. Daarnaast zijn er ook muterende puntbronnen, waarbij het DNA van de infectiebron steeds verandert.

Wanneer er een puntbron is die niet muteert, is dit makkelijk te herkennen. Immers, alle gevonden infecties stammen direct af van eenzelfde kolonie, en alle bladeren hebben dus dezelfde ouder. Als de puntbron wel muteert, wordt het een stuk lastiger. Immers, wanneer de puntbron muteert verschijnt dit als een nieuwe knoop in de phylogenetische boom. Dit betekent dat de puntbron niet meer wordt aangegeven door een enkele knoop, maar in plaats daarvan door een pad in de afstammingsboom.

Om aan te geven in hoeverre een uitbraak op een puntbronuitbraak lijkt, kunnen we kijken naar hoeveel van de gevonden infecties het gevolg zijn van directe transmissie vanaf de puntbron. Wanneer een gevonden infectie het gevolg kan zijn van directe transmissie van een puntbron zeggen we dat deze **toe te schrijven** is aan de puntbron. We weten echter niet of er überhaupt wel sprake

is van een puntbron, en wanneer er wel een puntbron is, weten we niet welke knoop of welk pad in de boom nu de puntbron aangeeft, onze boom is immers ongeworteld. Daarnaast moeten we nog goed definiëren wat we bedoelen met “directe transmissie”.

We lossen dit op door te kijken naar het maximale aantal infecties dat toe te schrijven is aan een puntbron. We weten niet of er een pad is dat een puntbron weergeeft, en als er een is weten we niet welk pad dit is. Daarom zoeken wij het pad waaraan het grootste aantal infecties toe te schrijven is. Hierbij beschouwen we een enkele knoop ook als een pad, dat dan maar één knoop bevat. Wanneer er sprake is van een pure puntbron, zullen alle gevonden infecties aan een pad toe te schrijven zijn. Wanneer er sprake is van pure transmissie, verwachten we dat er relatief weinig infecties aan een pad toe te schrijven zijn.

Nu moeten we definiëren wat we bedoelen als we zeggen dat een infectie aan een pad toe te schrijven is. We kunnen beginnen door te zeggen dat elk blad dat uit een pad ontspruit, aan dit pad toe te schrijven is, dit spreekt voor zich. Wat doen we echter wanneer er een grotere deelboom aan het pad vastzit? Dit suggereert dat, nadat de puntbron iemand heeft geïnfecteerd, deze persoon de infectie naar anderen heeft verspreidt. Het is echter mogelijk, en zelfs zeker als we er vanuit gaan dat elke patiënt gevonden wordt, dat de oorspronkelijk door de puntbron besmette patiënt één van de aangetroffen infecties in de deelboom is, eventueel meerdere malen gemuteerd in de patiënt. Daarom beschouwen we in elke uit het pad ontsproten deelboom één blad als toe te schrijven aan het pad. Welk blad dit is, maakt niet uit. Immers, we zijn geïnteresseerd in het *aantal* aan het pad toe te schrijven bladeren, niet in *welke* bladeren toe te schrijven zijn.

Het aantal bladeren dat maximaal aan een pad toe te schrijven is noemen wij de maximale bijdrage van een puntbron, of de **maximale bijdrage** in het kort. We noteren deze maximale bijdrage met het symbool C_{max} .

3.1.1 Het algoritme

Wij hebben nu de maximale bijdrage gedefiniëerd, dit is een eerste stap naar een maat om aan te geven in hoeverre een uitbraak op een puntbronuitbraak lijkt. We moeten echter wel weten hoe we deze maximale bijdrage nu vinden. Indien nodig zouden we elk mogelijk pad kunnen vergelijken, dit is echter niet erg efficiënt, het aantal mogelijke paden kan namelijk potentieel erg hoog zijn bij grotere bomen.

Er is echter een betere methode. Elke deelboom heeft een maximaal bijdragend deelpad, en een maximale bijdrage voor de deelboom zelf. Wanneer er dan twee of meer deelbomen bij elkaar komen, hoeft men alleen maar het grootste maximaal bijdragend deelpad te kiezen om uit te breiden tot het maximaal bijdragend deelpad van de nieuwe deelboom. Voor de maximale bijdrage van de nieuwe deelboom hoeft men alleen maar de grootste en op één na grootste maximaal bijdragende deelpaden met elkaar te verbinden en te kijken of dit een grotere maximale bijdrage levert dan de grootste maximale bijdrage van de oorspronkelijke deelbomen, waarna men uiteraard de grootste maximale bijdrage

als maximale bijdrage van de nieuwe deelboom kiest. Hiermee kan men door gaan tot men de oorspronkelijke boom weer terug heeft.

We kunnen dit algoritme als volgt weergeven. Bemerkt dat we terminologie voor gewortelde bomen (ouder, kind) gebruiken, terwijl we met ongewortelde bomen werken. Men kan immers willekeurig een wortel kiezen, hoewel deze wortel dan geen betekenis heeft.

1. Elke knoop i heeft n_i kinderen, waarvan l_i bladeren zijn en k_i niet, dus $n_i = l_i + k_i$.
2. We schrijven voor elke knoop de **onderwaarde** ν_i op, beginnende bij de bladeren en dan naar boven werkend.

$$\nu_i = l_i + \max_{x \in \text{kinderen}(i)} \nu_x + \max(0, k_i - 1) \quad (1)$$

Hierbij is $\text{kinderen}(i)$ de verzameling van kinderen van knoop i .

3. Nu schrijven we voor elke knoop i de **bijdraagwaarde** b_i op.

$$b_i = l_i + \max_{x \in \text{kinderen}(i)} \nu_x + \underset{q}{2\text{ndmax}}_{x \in \text{kinderen}(i)} \nu_x + \max(0, k_i - 2) + |\text{ouders}(i)| \quad (2)$$

Hierbij bedoelen we met $\underset{q}{2\text{ndmax}}$ de op één na grootste r_q voor alle mogelijke q , of 0 als deze niet bestaat. Daarnaast is $\text{ouders}(i)$ de verzameling van ouders van i . Deze bevat ten hoogste één element, maar kan ook leeg zijn.

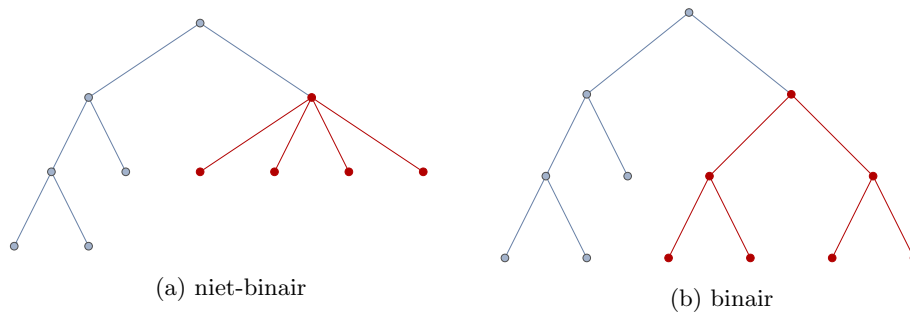
4. De maximale bijdrage is nu $C_{max} = \max_i b_i$.

3.2 Minimale maximale bijdrage

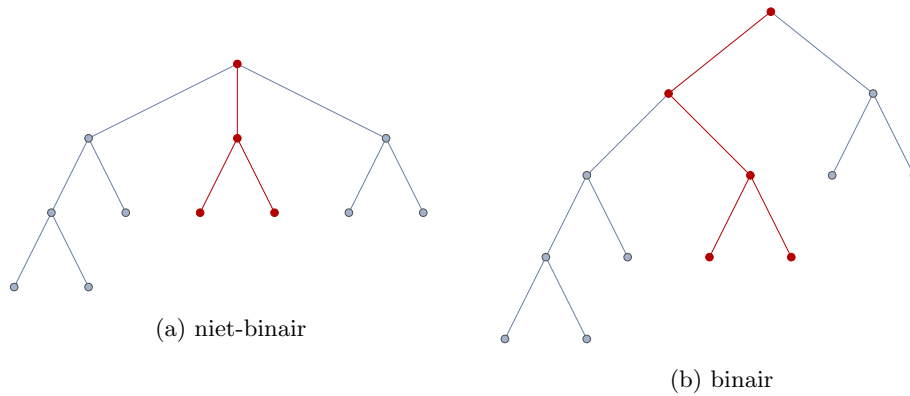
We hebben nu het concept van maximale bijdrage gedefiniëerd, we weten hoe we deze moeten uitrekenen, en we weten dat de maximale bijdrage relatief klein is wanneer er sprake is van pure transmissie. Echter, wat is “relatief klein”? We weten dat er, ook bij pure transmissie, altijd een maximale bijdrage is. Om te kunnen zeggen wanneer de maximale bijdrage van een boom met n bladeren “relatief klein” is, moeten we dus weten wat de **minimale maximale bijdrage** $C_{max,min}(n)$ is voor een boom met n bladeren.

Om te beginnen merken we op dat de minimale maximale bijdrage plaatsvindt bij een binaire boom. Immers, bij een niet-binaire boom kan men het niet-binaire gedeelte vervangen door een binair gedeelte waarbij het aantal bladeren behouden blijft maar de maximale bijdrage maximaal even groot blijft.

Wij zullen ons nu eerst richten op bomen met een aantal bladeren $n = 2^k$ voor een zeker geheel getal k . Hierbij is namelijk makkelijk te zien welke soort boom de minimale maximale bijdrage levert. Dit is namelijk de volledig gebalanceerde binaire boom, dat wil zeggen de binaire boom die bestaat uit twee samengevoegde helften die elk zelf weer de volledig gebalanceerde binaire boom



Figuur 6: Een niet-binair gedeelte wordt vervangen door een binair gedeelte, maximale bijdrage gaat omlaag

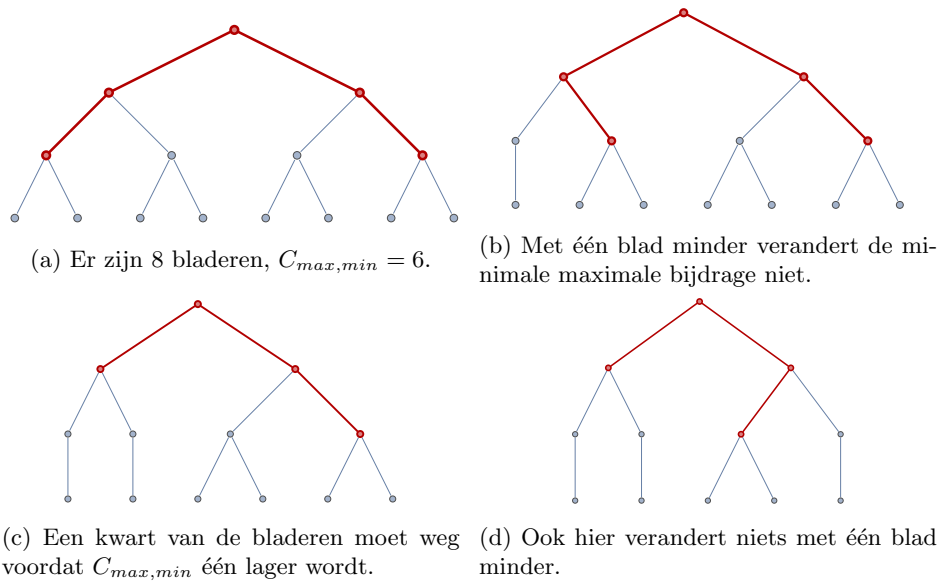


Figuur 7: Een niet-binair gedeelte wordt vervangen door een binair gedeelte, maximale bijdrage blijft gelijk

van $2^{k-1} = n/2$ bladeren vormen. In deze boom wordt het pad van maximale bijdrage gevormt door elk zo lang mogelijk pad dat door de wortel loopt. Al deze paden zijn equivalent en geven een maximale bijdrage van $2k$.

De minimale maximale bijdrage voor 2^k bladeren is dus $2k$. Nu willen we graag de minimale maximale bijdrage weten voor bomen met $n \neq 2^k$ bladeren. We doen dit door met 2^k bladeren te beginnen en dan bladeren weg te halen. Indien $k > 2$ is het door het weghalen van één blad niet mogelijk om een kleinere maximale bijdrage te maken. De minimale maximale bijdrage wordt één lager als in één helft van de boom de helft van de bladeren worden weggehaald, dus als een kwart van alle bladeren weg is. Om de minimale maximale bijdrage met nog één te verlagen moet ook in de andere helft van de boom de helft van de bladeren worden weggehaald, dat is dus de helft van alle bladeren. Dit is consistent met wat we hadden, want dat betekent dat bij $2^k/2 = 2^{k-1}$ bladeren de minimale maximale bijdrage $2k - 2 = 2(k - 1)$ is.

Om de minimale maximale bijdrage bij n bladeren uit te rekenen nemen we



Figuur 8: We kunnen het verloop van de minimale maximale bijdrage zien door bladeren weg te halen.

dus de volgende stappen:

1. We noemen $k = \lceil \log_2(n) \rceil$.
2. Als $n < 2^k - 2^{k-2}$ dan $C_{max,min}(n) = 2k - 1$, anders $C_{max,min}(n) = 2k$.

Hierbij is $\lceil x \rceil$ de *ceiling* van x , oftewel x afgerond naar boven.

Eén van de dingen die we hieruit kunnen opmaken is dat, voor $n \leq 5$, $C_{max,min}(n) = n$. In andere woorden, om iets te kunnen zeggen over of een uitbraak wel of niet een puntbronuitbraak is, moeten er ten minste zes infecties gevonden zijn.

3.3 Simulatie

We hebben nu dus de maximale bijdrage C_{max} , alsmede een minimale waarde $C_{max,min}(n)$ en een maximale waarde n voor de maximale bijdrage in een boom met n bladeren. Een logische manier om dit te verwerken zou dan zijn om te kijken naar de **bijdrageverhouding** $R = (C_{max} - C_{max,min}(n))/(n - C_{max,min}(n))$. Dit geeft een uitdrukking voor de mate waarop een uitbraak op een puntbronuitbraak lijkt, door middel van een getal van 0 tot 1.

We weten hiermee echter nog steeds niet hoe laag deze waarde moet zijn voordat we zeker weten dat het om pure transmissie gaat. We weten dat een pure puntbronuitbraak een waarde van 1 geeft, maar is een waarde van 1 genoeg om te zeggen dat het om een puntbronuitbraak gaat?

Om hierachter te komen kunnen we de computer gebruiken om transmissieuitbraken te simuleren. Als we dan na elk nieuw blad de maximale bijdrage uitrekenen, en dit vaak genoeg herhalen, kunnen we een indruk krijgen van wat een normale R is voor een transmissieuitbraak.

Om hier enig inzicht in te krijgen hebben wij vijftigduizend uitbraken gesimuleerd. We zijn er hierbij van uit gegaan dat de tijd totdat een besmetting wordt ontdekt en de tijd totdat een besmetting tot een nieuwe besmetting leidt beide exponentieel verdeeld zijn. Wanneer dit het geval is, is de tijd totdat één van beide gebeurt ook exponentieel verdeeld. We kunnen een uitbraak dus al als boom simuleren, waarbij we elke *tick* in elke knoop iets laten gebeuren (transmissie of ontdekking), en dan vervolgens een exponentieel verdeelde tijd bij de huidige tijd in elke knoop optellen. Uiteindelijk kunnen we dan in alle bladeren kijken op welke tijd ze zijn ontdekt, en dan kunnen we dus de boom op verschillende momenten bekijken, en dus met de resultaten van één boom R uitrekenen voor verschillende waarden van n .

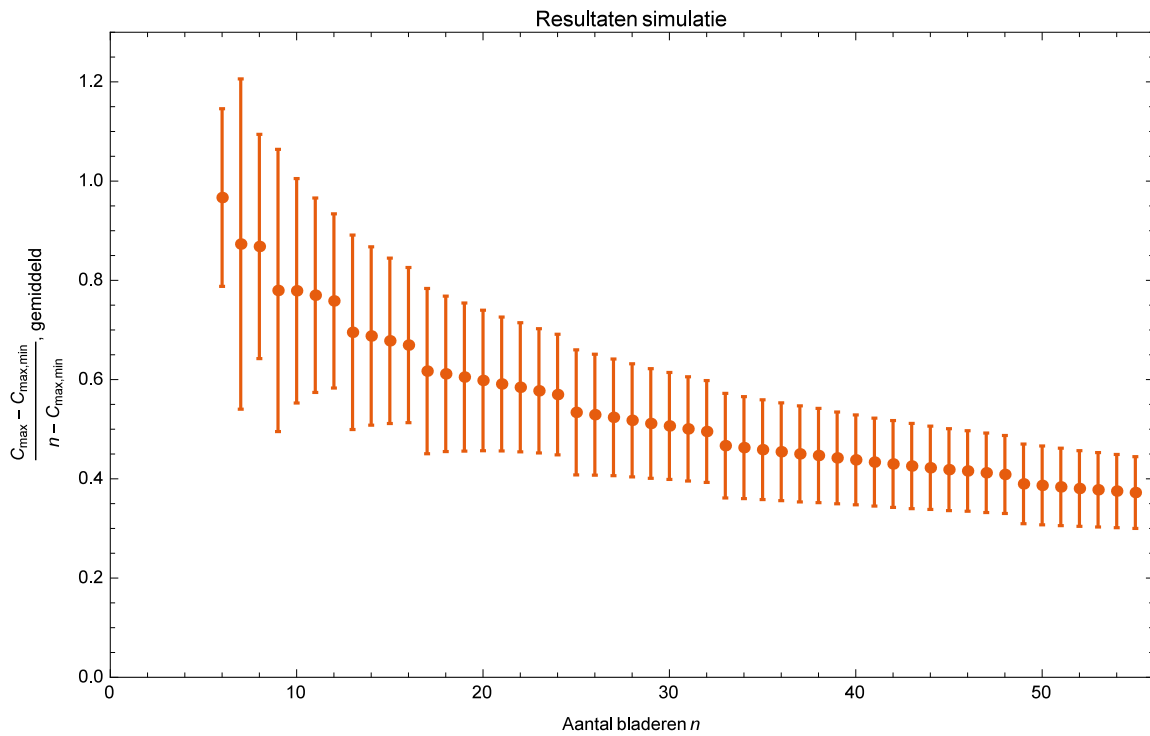
De broncode van onze simulatie (geschreven voor Python 3.4) is te vinden op <https://github.com/rakkar16/TransmissieBoomSim>. Het resultaat is te zien in Figuur 10. Omdat we vooral geïnteresseerd zijn in de gevallen met kleine n , we zouden de uitbraak immers graag stoppen voordat deze uit de hand loopt, hebben we in Figuur 9 de resultaten voor de eerste vijftig bruikbare waarden van n geplot. Dit is n van 6 tot 55, want voor $n \leq 5$ is R niet gedefiniëerd.

In Figuur 9a zijn de gemiddelden van R geplot voor verschillende n , met de standaarddeviaties als *error bars*. In Figuur 9b zijn over hetzelfde domein de boxplots van de verschillende gevonden waarden van R geplot. Hierbij zit dus de middelste 50% van de resultaten in de box. In Figuur 10 staan weer gemiddelden en standaarddeviaties.

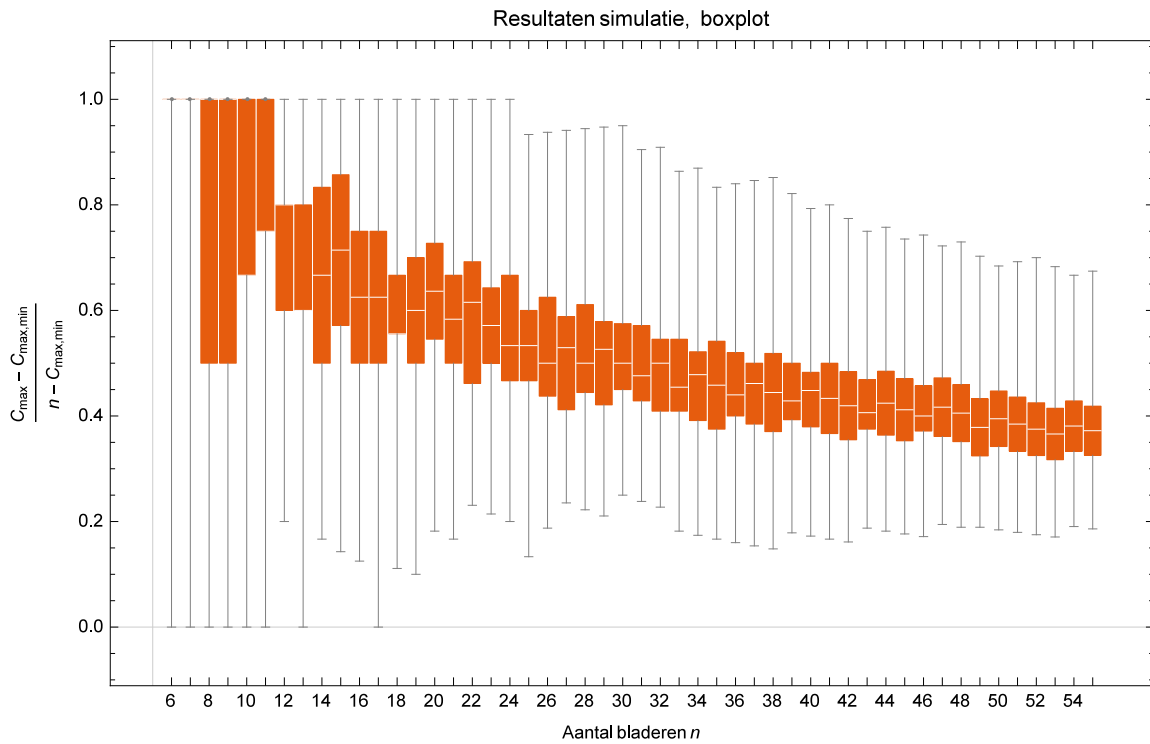
Wat we hierin zien is dat, bij niet al te grote uitbraken, een hoge R niet perse hoeft te betekenen dat er een puntbronuitbraak is. We zien namelijk pas bij 25 bladeren voor het eerst geen bomen met $R = 1$ meer. Bij kleine uitbraken kan het dus lastig zijn om te zeggen wat voor uitbraak er is. Bij grote uitbraken wordt dit makkelijker. We zien dat bij hoge n de R meestal wel een duidelijk lage waarde aanneemt. In deze simulatie is voor $n \geq 12$ consistent driekwart van de R -waarden onder de 0.9. Voor $n \geq 16$ is dit zelfs onder de 0.8. Voor $n = 6$ en $n = 7$ heeft echter nog minstens driekwart van de transmissieuitbraken een $R = 1$.

3.4 In de praktijk

Graag zouden we zien hoe goed onze methode in de praktijk werkt, maar *sequencing* wordt nog niet veel gebruikt in de bestrijding van uitbraken. Er zijn echter wel een aantal gevallen waarin *sequencing* is gebruikt bij een uitbraak, hetzij om een bijzonder ernstige uitbraak aan te pakken, hetzij om een uitbraak achteraf te analyseren. Sommige van de hieruit komende resultaten zijn ook gepubliceerd, wat ons in staat stelt om onze methode de testen.

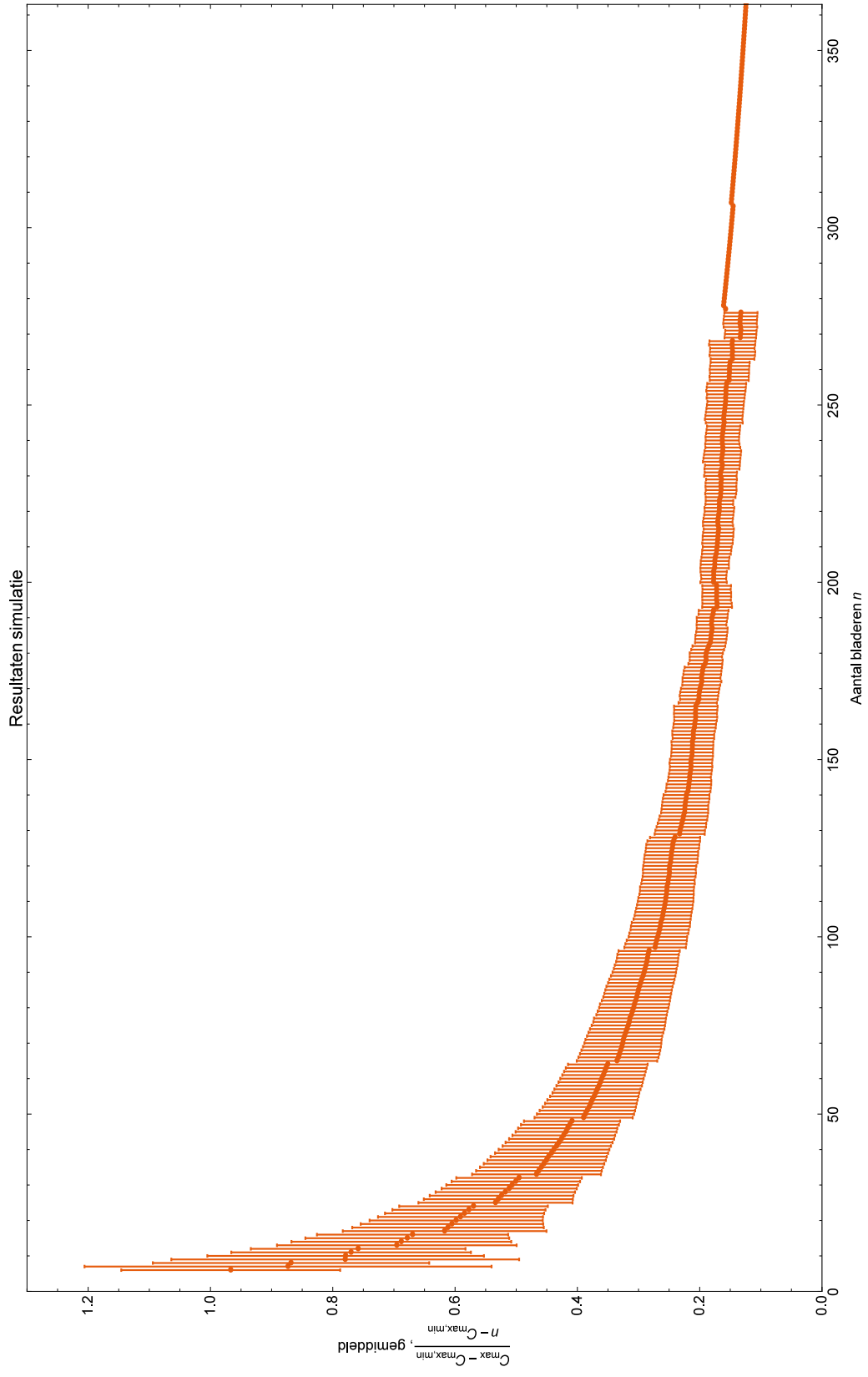


(a) Plot van de gemiddelden en standaarddeviaties voor n van 6 tot 55.



(b) Boxplot voor n van 6 tot 55. De *whiskers* geven de minima en maxima aan.

Figuur 9: Plots van de resultaten voor de eerste vijftig bruikbare waarden van n van vijftigduizend gesimuleerde uitbraken.



Figuur 10: De resultaten van vijftigduizend gesimuleerde uitbraken. De punten geven de gemiddeldes aan, de balken de standaarddeviaties. Daar waar geen foutbalken zijn, was er slechts één datapunt.

4 Discussie

Wij hebben een model gemaakt om bacteriële uitbraken te analyseren, om daarmee een poging te doen te zeggen wat voor uitbraak dit is. We hebben hier zeker wat inzicht in gekregen. Echter, graag hadden we gewild dat we aan de hand van een boom hadden kunnen zeggen wat de kans is dat het om een puntbron- of transmissieuitbraak gaat. Zover zijn wij niet gekomen, en hier zou dus nog meer onderzoek naar moeten worden gedaan. Daarnaast hebben wij enkele aannames gedaan, en hier moet rekening mee worden gehouden.

Ten eerste zijn wij er van uit gegaan dat er altijd sprake is van ofwel een puntbronuitbraak of een transmissieuitbraak. Wij hebben nog wel ruimte gelaten voor puntbronuitbraken die zich vervolgens via transmissie nog verder verspreiden, maar wij hebben geen rekening gehouden met nog andere situaties. De reden hiertoe is dat deze situaties zich zelden zullen voordoen in een ziekenhuis, maar als men deze methode wilt gebruiken buiten een ziekenhuissituatie, moet hier rekening mee worden gehouden.

Zo is het mogelijk dat een puntbron wordt gesplitst in meerdere bronnen, bijvoorbeeld wanneer de bron een geïnfecteerde lading voedsel is. Wanneer deze bron niet muteert, zal dit ook weer op een puntbron neerkomen, maar wanneer er wel mutatie optreedt, zal de resulterende boom niet op een puntbronboom lijken. In sommige gevallen is dit echter nog op te lossen omdat er dus meerdere uitbraken op meerdere locaties zijn, waarbij elke afzonderlijke uitbraak wel puntbronachtig is.

Ook is het mogelijk dat een geïnfecteerde patiënt de oorzaak is van een nieuwe puntbron. Denk bijvoorbeeld aan cholera, wat meestal wordt veroorzaakt doordat drinkwater besmet is door uitwerpselen van een besmette persoon. Het drinkwater dient hier als puntbron, maar wanneer de infectie zich verspreidt, zullen nieuwe patiënten ook voor nieuwe puntbronnen zorgen. Dit zal waarschijnlijk tot een transmissieachtige boom leiden, maar in tegenstelling tot pure transmissie kan men de uitbraak niet stoppen met quarantainemaatregelen.

Binnen de ziekenhuissituatie is er nog de mogelijkheid dat de uitbraak niet binnen het ziekenhuis plaatsvindt, maar dat er meerdere patiënten van buiten komen die door dezelfde soort bacterie zijn besmet. Meestal zal men dit wel in de gaten hebben, hetzij doordat er bekend is dat er een uitbraak buiten het ziekenhuis heeft plaatsgevonden, hetzij doordat de stammen zozeer van elkaar verschillen dat het onwaarschijnlijk is dat ze een recente gezamenlijke voorouder hebben. Dit betekent echter wel dat men niet naïef in elk geval onze methode kan toepassen.

Wij zijn er ook van uit gegaan dat we toegang hebben tot een correcte boom. Wij hebben ons niet beziggehouden met hoe zo'n boom geconstrueerd dient te worden, maar men heeft meestal geen garantie dat de resulterende boom ook correct is. Het zou nog de moeite waard kunnen zijn om onderzoek te doen naar in hoeverre de verschillende bomen die men uit een dataset kan construeren verschillende resultaten opleveren.

Wij zijn er steeds van uit gegaan dat elke patiënt een unieke infectie had.

Mocht dit echter niet zo blijken te zijn, dan is dit geen probleem. Voeg de patiënten met identieke infecties samen tot één blad, behalve bij het uitrekenen van de maximale bijdrage, waarbij dit blad als meerdere bladeren moet worden geteld.

Lastiger is dat wij er van uit zijn gegaan dat elke patiënt en elke puntbron slechts één dominante stam bij zich draagt. Wanneer dit niet het geval is, kan men feitelijk binnen de puntbron al een transmissieproces hebben, waardoor in principe elk blad in elke mogelijke boom nog aan een puntbron toe te kennen is. Het zou wellicht nog mogelijk zijn om statistische methodes hiervoor toe te passen, het is immers niet waarschijnlijk dat een gigantische transmissieboom binnen een puntbron ontstaat. Hiervoor is echter meer onderzoek vereist.

Verder hebben wij geen rekening gehouden met de mogelijkheid dat er *horizontal gene transfer* plaatsvindt. Dit wil zeggen dat een bacterie niet alleen zelf muteert, maar ook DNA van andere bacteriën kan krijgen. Dit zal echter vooral het opstellen van een boom van de uitbraak lastiger maken, en wij gaan er van uit dat we de boom al hebben.

Ten slotte moeten we niet vergeten dat we hier vooral op de muterende puntbron hebben gefocust, omdat dit de meest ingewikkelde situatie is. Een niet-muterende puntbron is veel makkelijker te herkennen, men gebruikt dan de maximale graad in plaats van de maximale bijdrage, welke bij transmissie zelden in de buurt van n komt.

5 Conclusie

Teneinde aan de hand van een phylogenetische boom een onderscheid te kunnen maken tussen puntbronuitbraken en transmissieuitbraken, hebben wij de maximale bijdrage C_{max} en de bijdrageverhouding R ingevoerd. Wanneer $R < 1$ kunnen wij met zekerheid zeggen dat er geen sprake is van een pure puntbronuitbraak. Voor kleine uitbraken is pure transmissie echter niet altijd makkelijk te herkennen. Dit gaat zelfs zo ver dat zelfs voor middelgrote uitbraken ($n \approx 20$), $R = 1$ nog kan overeenkomen met pure transmissie.

In de meeste gevallen is er echter wel sprake van een duidelijk verschil, mits n niet al te klein is. (vanaf ongeveer $n \geq 10$) Wanneer er sprake is van transmissie, ook al is dit geen pure transmissie, zal er meestal gelden dat $R < 1$, en de wetenschap dat er sprake is van transmissie kan al helpen bij het bestrijden van een uitbraak.

Deze bijdrageverhouding is geïntroduceerd opdat ook muterende puntbronnen te herkennen zijn. Niet-muterende puntbronnen zijn makkelijker en betrouwbaarder te herkennen door te kijken naar de maximale graad in een boom, in plaats van naar C_{max} .

De maximale bijdrage vertelt ons dus niet alles over een uitbraak, en kan ons niet altijd helpen, maar het kan zeker nuttig zijn bij het analyseren en bestrijden van uitbraken, omdat het wel iets kan zeggen over het soort uitbraak waar men mee te maken heeft, nog zonder iets te weten over de situatie waarin deze uitbraak plaatsvindt.

6 Bronnen

- Warnow, Tandy. *An Introduction to Computational Phylogenetics*. Nog niet gepubliceerd.
- Ross, Sheldon M. *Introduction to probability models*. Academic press, 2003.
- Snitkin, Evan S., et al. “Tracking a hospital outbreak of carbapenem-resistant *Klebsiella pneumoniae* with whole-genome sequencing.” *Science translational medicine* 4.148 (2012): 148ra116-148ra116.