



Technische Universiteit Delft  
Faculteit Elektrotechniek, Wiskunde en Informatica  
Delft Institute of Applied Mathematics

**Een leider-volg model voor online sociale netwerken,  
gebaseerd op een SIR-model voor het Zikavirus.  
(Engelse titel: A leaders-followers model for online  
social networks, based on a SIR-model for the Zika  
virus.)**

Verslag ten behoeve van het  
Delft Institute of Applied Mathematics  
als onderdeel ter verkrijging

van de graad van

**BACHELOR OF SCIENCE  
in  
TECHNISCHE WISKUNDE**

door

**TIMO MAURITS VAN PRAAGH**

**Delft, Nederland  
September 2016**





BSc verslag TECHNISCHE WISKUNDE

**“De groei en afname van online sociale netwerken gemodelleerd met een SIR model gebaseerd op het Zikavirus.”**

**(Engelse titel: “The growth and decline of online social networks modelled with a Zika virus based model.”)**

TIMO MAURITS VAN PRAAGH

Technische Universiteit Delft

**Begeleider**

Prof.dr.ir. C. Vuik

**Overige commissieleden**

Dr. J. Dubbeldam

Drs. E.M. van Elderen

September, 2016

Delft



# Samenvatting

Online sociale netwerken zijn zeer populair tegenwoordig. Het gebruik en de invloed van de bedrijven achter deze netwerken hebben veel invloed op de maatschappij en hun macht is verstrekkend. Niet alleen op het voor de hand liggende sociale aspect, maar ook op het gebied van privacy en de economie. Het zijn immers zeer grote bedrijven waar zeer veel geld in om gaat en daarmee zijn de belangen enorm. Het gebruik modelleren van online sociale netwerken is dus interessant voor zowel bedrijven als ook individuen. In dit verslag worden verschillende vormen van SIR-model bekeken, alsmede uitbreidingen. Ook wordt een implementatie gegeven van een model voor het verloop van het Zikavirus. Hierbij worden de invloeden van twee verschillende populaties op elkaar en de effecten daarvan gemodelleerd. Het concept van twee verschillende groepen die effect op elkaar hebben wordt ook gebruikt om de populatie in twee groepen te verdelen bij het modelleren van het gebruik van online sociale netwerken. Er wordt onderscheidt gemaakt tussen leiders en volgers, ook wel ‘early adaptors’ en ‘followers’ genoemd. Ook wordt er rekening gehouden met het effect van leiders op volgers bij dit model. Het model wordt vergeleken met een bekende uitbreiding op het SIR-model, het irSIRSbd-model. Het zo genoemde leiders-volgersmodel presteert beter dan het SIR-model, en vergelijkbaar met het irSIRSbd-model, terwijl er geen vergelijkbare uitbreidingen aan zijn toegevoegd. De parameters voor de modellen zijn geoptimaliseerd in Matlab met `fminsearch` de kleinst mogelijke sum of squared errors te geven ten op zichte van de bekende zoekdata.



# Inhoudsopgave

<b>1</b>	<b>Inleiding</b>	<b>9</b>
<b>2</b>	<b>SIR modellen</b>	<b>11</b>
2.1	Differentiaalvergelijking modellen . . . . .	11
2.1.1	Immunitetscriterium . . . . .	11
2.1.2	Resistentiecriterium . . . . .	12
2.1.3	Stabiliteit . . . . .	12
2.1.4	Analytische benadering . . . . .	12
2.1.5	Uitbreiding irSIRS-model met geboorte- en sterftedata . . . . .	13
2.1.6	Zika-virus model . . . . .	14
2.2	Statistische modellen . . . . .	15
2.2.1	Van discreet naar continu . . . . .	15
2.3	Willekeurig netwerkgraafmodel . . . . .	16
2.3.1	Het Barabasi Albert model voor sociale netwerken . . . . .	16
2.3.2	Model . . . . .	17
2.3.3	Heidler's balanstheorie . . . . .	17
<b>3</b>	<b>Het leiders-volgersmodel</b>	<b>19</b>
<b>4</b>	<b>Resultaten</b>	<b>21</b>
4.1	Vergelijking analytische en numerieke benadering irSIRS model . . . . .	21
4.2	Nieuwe voorspelling voor Facebook . . . . .	23
4.3	Resultaten Zika-model . . . . .	25
4.4	Resultaten leiders-volgersmodel . . . . .	26
4.4.1	Leiders-volgers effect . . . . .	27
4.4.2	Leiders-volgersmodel vs SIR-model . . . . .	30
4.4.3	Leiders-volgersmodel vs irSIRbd-model . . . . .	32
<b>5</b>	<b>Conclusie en Discussie</b>	<b>33</b>
5.1	Conclusie . . . . .	33
5.2	Discussie . . . . .	33
	<b>Appendices</b>	<b>35</b>
<b>A</b>	<b>Uitwerking vergelijkingen sectie 2.1.4</b>	<b>37</b>
<b>B</b>	<b>Gebruikte Matlab-codes</b>	<b>39</b>
B.1	Matlabcode bij het SIR-model ter referentie . . . . .	39
B.2	Matlabcode bij het irSIRbd-model . . . . .	39
B.3	Matlabcode bij het leidervolg-model . . . . .	41

B.4	Matlabcode bij het Zikavirus-model . . . . .	42
B.5	Matlabcode bij vergelijking analytische benadering en numerieke benadering . . .	44
B.6	Matlabcode voor berekening SSE van SIR-model . . . . .	45
B.7	Matlabcode voor berekening SSE van irSIRSbd-model . . . . .	46
B.8	Matlabcode voor berekening SSE van leiders-volgersmodel . . . . .	46
B.9	Minimalisatiecode voor SSE van SIR-model . . . . .	50
B.10	Minimalisatiecode voor SSE van irSIRSbd-model . . . . .	51
B.11	Minimalisatiecode voor SSE van leiders-volgersmodel . . . . .	52
<b>C</b>	<b>Parameters</b>	<b>55</b>
C.1	Parameters bij voor facebook een jaar later . . . . .	55
C.2	Geoptimaliseerde parameterwaarden SIR-model . . . . .	55
C.3	Geoptimaliseerde parameterwaarden irSIRSdb-model . . . . .	55



# Hoofdstuk 1

## Inleiding

In 2009 werd in het vooraanstaande blad Nature het volgende artikel gepubliceerd: “Detecting influenza epidemics using search engine query data.” door J. Ginsberg et al [8]. Ginsberg, was een onderzoeker werkend bij Google inc. en deed onderzoek met de zoekdata die de zoekmachine van Google genereert. Aan de hand van zoektermen die gezocht worden kunnen griepedemiën eerder worden gedetecteerd dan door de gebruikelijke overheidsinstellingen die gemelde ziekenhuisbezoeken registreren. Door eerder deze epidemiën op te merken kunnen deze beter bestreden worden en ook de verspreiding beperkt worden. Dit wordt al in de praktijk gebracht en zo zien we met dit voorbeeld hoe zoekdata gebruikt kan worden om epidemiologische ziektes te bestrijden.

Deze zoekdata kan echter ook gebruikt worden om voorspellingen te doen van andere epidemieachtige verschijnselen. Zo laten Asatani et al [2] en Maring [10] zien hoe een relatief simpel wiskundig model voor een epidemie gebruikt kan worden om een voorspelling te doen van het gebruik van online sociale netwerken, als Facebook, Myspace, Twitter, Snapchat etc.

In dit verslag worden verschillende methoden besproken om de groei en afname in het gebruik van sociale netwerken te modelleren met (varianties op) het SIR-model. Eén van de varianties op dit model is gemaakt voor het Zika-virus, waarbij besmette mensen elkaar aansteken, maar ook door muggen die besmet raken door het steken van een mens en nu juist mensen besmetten door hen te steken. Dit idee voor een model wordt in dit verslag, voor zo ver bekend aan de auteur, voor het eerst gebruikt om early-adaptors en volgers te modelleren van sociale netwerken. Early adaptors van een online sociaal netwerk beïnvloeden andere early adaptors, maar ook volgers, ‘de grijze massa’. Volgers onderling beïnvloeden elkaar natuurlijk ook en zo groeit het gebruik van een online sociaal netwerk. De muggen uit het Zikamodel zijn de early adaptors en de grijze massa is de mensenpopulatie.

De data voor dit verslag zijn gevonden via Google Trends (<https://trends.google.com>), dit is een applicatie die de frequentie van zoektermen bijhoudt. Dit geeft een representatieve en vaak gebruikte dataset van de populariteit van online diensten. Wanneer een dienst nog niet bestaat, zal er ook niet naar gezocht worden op de zoekmachine en wanneer deze populairder is wordt deze vaker gezocht.



## Hoofdstuk 2

# SIR modellen

In dit hoofdstuk worden verschillende SIR-modellen besproken, opgedeeld in drie verschillende soorten: differentiaalvergelijking modellen, statistische modellen en netwerkgraaf modellen, elk apart besproken in een sectie.

### 2.1 Differentiaalvergelijking modellen

In het bachelor verslag van Sandra Maring [10], waar dit verslag op gebaseerd is en als een vervolg of uitbreiding kan worden gezien, worden verschillende modellen besproken waarmee epidemieën gemodelleerd kunnen worden. Het klassieke SIR-model is gebaseerd op drie differentiaal vergelijkingen:

$$\frac{dS}{dt} = -\frac{\beta IS}{N} \quad (2.1)$$

$$\frac{dI}{dt} = \frac{\beta IS}{N} - \frac{\gamma I}{N} \quad (2.2)$$

$$\frac{dR}{dt} = \frac{\gamma I}{N} \quad (2.3)$$

Waarbij traditioneel  $S$  het aantal vatbare mensen,  $I$  het aantal geïnfecteerde mensen en  $R$  het aantal genezen mensen is.  $\beta$  is de infectiegraad, of de ‘snelheid’ waarmee mensen geïnfecteerd raken.  $\beta$  is de infectiegraad per tijdseenheid  $t$ ,  $\gamma$  de genezingsratio. Een veel gebruikte tijdstap-grootte is één week. In dit model wordt aangenomen dat het totale aantale mensen,  $N$ , gelijk blijft. Dus dan krijgen we de volgende voorwaarde er bij:

$$\frac{dS}{dt} + \frac{dI}{dt} + \frac{dR}{dt} = 0 \quad (2.4)$$

Voor dit model is een analytische oplossing lastig te vinden. Wel is er een benadering te vinden van deze analytische oplossing, daarover later meer.

#### 2.1.1 Immunitetskriterium

De ongelijkheid  $S < \frac{\gamma}{\beta}$  is de voorwaarde die weergeeft dat uiteindelijk iedereen immuun wordt voor de ziekte. Door te stellen dat het rechterlid van de vergelijking (2.2) kleiner dan nul moet zijn, wordt deze verkregen.

### 2.1.2 Resistentiecriterium

Het resistentiecriterium is de voorwaarde dat iedereen ook uiteindelijk resistent wordt na een bepaalde tijd. Hiervoor moet het aantal resistente personen simpelweg altijd toenemen, ofwel  $\frac{dR}{dt} > 0$ . We zien dat dit geldt wanneer  $\gamma I > 0$ . Dit is altijd zo, dus aan deze voorwaarde wordt altijd voldaan.

### 2.1.3 Stabiliteit

Door  $\frac{dS}{dt}$ ,  $\frac{dI}{dt}$ ,  $\frac{dR}{dt}$  allemaal gelijk te stellen aan nul worden evenwichtspunten gevonden. Er worden de punten  $(S, 0, R)$  gevonden, met  $S$  en  $R$  niet negatief. Dit evenwichtspunt is wiskundig correct, maar stelt in het model niet zo veel voor. Immers, als er niemand ziek is, kan er ook niemand geïnfecteerd worden en verandert het aantal resistente personen ook niet. Met een Jacobimatrix wordt de stabiliteit bekeken en aan de hand daarvan wordt geconcludeerd dat als  $0 \leq S < \frac{\gamma}{\beta}$ , dan is het SIR model stabiel in  $I=0$ .

Voor verdere uitwerkingen van de modellen kan gekeken worden naar het verslag van Sandra Maring [10]. Hierna wordt nog het irSIR model besproken, waar ook een analytische benadering als oplossing gevonden kan worden.

Het irSIR-model verschilt van het SIR-model door het feit dat er rekening gehouden wordt met het feit dat de verandering in resistentie beïnvloed wordt door het (relatieve) aantal mensen dat al resistent is. Dit is anders dan bij normale epidemiologische modellen, aangezien bij ziektes het aantal resistente mensen geen invloed heeft op hoe snel mensen genezen van een ziekte, maar bij een online sociaal netwerk dus juist wel.

### 2.1.4 Analytische benadering

Door Kazuyuki Fujii [9] wordt gesuggereerd dat er een betrekkelijk eenvoudige benadering te vinden is van de analytische oplossing van het irSIR-model. Hetzelfde model gebaseerd op vergelijkingen (2.1), (2.2) en (2.3) wordt herschreven tot:

$$\begin{aligned} \frac{dS}{dt \cdot S} &= -\frac{\beta I}{N} \\ \frac{dI}{dt \cdot I} &= \frac{\beta S}{N} - \frac{\nu R}{N} \\ \frac{dR}{dt \cdot R} &= \frac{\nu I}{N} \end{aligned} \tag{2.5}$$

Vervolgens herschrijven we:

$$S(t) = e^{s(t)}, \quad I(t) = e^{i(t)}, \quad R(t) = e^{r(t)} \tag{2.6}$$

We zien nu dat vergelijkingen (2.5) nu het volgende worden:

$$\begin{aligned} \dot{s} &= -\frac{\beta}{N} e^i \\ \dot{i} &= \frac{\beta}{N} e^s - \frac{\nu}{N} e^r \\ \dot{r} &= \frac{\nu}{N} e^i \end{aligned} \tag{2.7}$$

Wanneer we de simpele benadering  $e^x = 1 + x$  gebruiken dan wordt (2.7):

$$\begin{aligned}\dot{s} &= -\frac{\beta}{N}(1+i) \\ \dot{i} &= \frac{\beta}{N}(1+s) - \frac{\nu}{N}(1+r) \\ \dot{r} &= \frac{\nu}{N}(1+i)\end{aligned}\tag{2.8}$$

Dit kunnen we ook schrijven als een vector vergelijking:

$$\frac{d}{dt} \begin{pmatrix} s \\ i \\ r \end{pmatrix} = \frac{1}{N} \begin{pmatrix} -\beta \\ \beta - \nu \\ \nu \end{pmatrix} + \frac{1}{N} \begin{pmatrix} 0 & -\beta & 0 \\ \beta & 0 & -\nu \\ 0 & \nu & 0 \end{pmatrix} \begin{pmatrix} s \\ i \\ r \end{pmatrix}\tag{2.9}$$

Met de volgende substitutie krijgen we een vectorvergelijking waarvan we de oplossing weten en we die kunnen terug transformeren naar expliciete uitdrukkingen voor  $S(t)$ ,  $I(t)$ ,  $R(t)$ :

$$\begin{pmatrix} \bar{s} \\ \bar{i} \\ \bar{r} \end{pmatrix} = \begin{pmatrix} s \\ i \\ r \end{pmatrix} + \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}\tag{2.10}$$

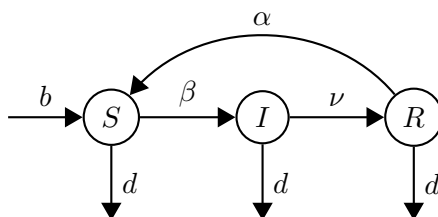
$$\Rightarrow \frac{d}{dt} \begin{pmatrix} \bar{s} \\ \bar{i} \\ \bar{r} \end{pmatrix} = \frac{1}{N} \begin{pmatrix} 0 & -\beta & 0 \\ \beta & 0 & -\nu \\ 0 & \nu & 0 \end{pmatrix} \begin{pmatrix} \bar{s} \\ \bar{i} \\ \bar{r} \end{pmatrix}\tag{2.11}$$

$$\Rightarrow \begin{pmatrix} \bar{s}(t) \\ \bar{i}(t) \\ \bar{r}(t) \end{pmatrix} = \exp \left( \frac{t}{N} \begin{pmatrix} 0 & -\beta & 0 \\ \beta & 0 & -\nu \\ 0 & \nu & 0 \end{pmatrix} \right) \begin{pmatrix} \bar{s}(0) \\ \bar{i}(0) \\ \bar{r}(0) \end{pmatrix}\tag{2.12}$$

Deze uitdrukking kan met terug transformeren naar een uitdrukking voor  $s(t)$ ,  $i(t)$  en  $r(t)$  en door daar de e-macht van te nemen verkrijgen we de expliciete uitdrukkingen voor  $S(t)$ ,  $I(t)$ ,  $R(t)$ . Voor verdere uitwerking hiervan zie Bijlage A.

### 2.1.5 Uitbreiding irSIRS-model met geboorte- en sterftedata

Het best presterende model uit het verslag van Maring [10] is het irSIRS-model met geboorte- en sterfte data. Dit model gebruikt de zelfde gedachte als het SIR-model, met de volgende uitbreidingen. Niet alleen kunnen resistente personen weer geïnfecteerd worden, ook beïnvloed het aantal resistente personen het aantal geïnfecteerden: wanneer meer mensen stoppen met het gebruik van een online sociaal netwerk, zullen anderen volgen. Daarnaast nemen we de geboorte- en sterftecijfers voor de gehele populatie mee. Dit model wordt in figuur 2.1 schematisch weergegeven.



Figuur 2.1: Illustratie van irSIRSbd-model

Aan de hand van het bovenstaande worden bijbehorende differentiaalvergelijkingen gevonden:

$$\frac{dS}{dt} = -\frac{\beta IS}{N} + \frac{\alpha R}{N} - \frac{dS}{N} + b \quad (2.13)$$

$$\frac{dI}{dt} = \frac{\beta IS}{N} - \frac{\nu I}{N} - \frac{dI}{N} \quad (2.14)$$

$$\frac{dR}{dt} = \frac{\nu I}{N} - \frac{\alpha R}{N} - \frac{dR}{N} \quad (2.15)$$

### 2.1.6 Zika-virus model

Een ander SIR-model met geboorte- en sterftcijfers, wat in het artikel "Mathematical Models for Fighting Zika Virus" [7] wordt besproken, brengt ons tot een bruikbaar concept voor online sociale netwerken. Er worden twee aparte populaties bekeken, mensen en muggen, die beiden een cyclus van vatbaar tot hersteld doorlopen en elkaar in verschillende fases kunnen beïnvloeden. Zo kunnen ingefecteerde muggen  $I_v$  vatbare mensen,  $S_h$ , bijten en hen infecteren, maar ook andersom: geïnfecteerde mensen,  $I_h$ , kunnen gebeten worden door vatbare muggen  $S_v$ , die vervolgens geïnfecteerd raken. In het geval van het Zikavirus geldt dat men eerst blootgesteld,  $E_h$  en  $E_v$ , wordt aan het virus en daarna pas besmettelijk is,  $I_h$  en  $I_v$ . In de context van online sociale netwerken is dit bijvoorbeeld te gebruiken als de splitsing van blootgesteld en daarna pas besmettelijk zou eventueel weggelaten kunnen worden, of gebruikt worden als een buffer waarbij elke gebruiker eerst iets uitprobeert en daarna pas actief in gebruik neemt. In het artikel "Comparing dengue and chikungunya emergence and endemic transmission in *A. aegypti* and *A. albopictus*" [6] worden deze fases gemodelleerd aan de hand van de volgende differentiaalvergelijkingen:

$$\begin{aligned} \frac{dS_h}{dt} &= \psi_n H_0 - \lambda_h S_h - \mu_h S_h \\ \frac{dE_h}{dt} &= \lambda_h S_h - \nu_h E_h - \mu_h E_h \\ \frac{dI_h}{dt} &= \nu_h E_h - \gamma_h I_h - \mu_h I_h \\ \frac{dR_h}{dt} &= \gamma_h I_h - \mu_h R_h \\ \frac{dS_v}{dt} &= h_v (N_v) N_v - \lambda_v S_v - \mu_v S_v \\ \frac{dE_v}{dt} &= \lambda_v S_v - \nu_v E_v - \mu_v E_v \\ \frac{dI_v}{dt} &= \nu_v E_v - \mu_v I_v \end{aligned} \quad (2.16)$$

Daarnaast zijn er twee populatiegroottes,  $N_h$  en  $N_v$  respectievelijk voor mensen en muggen. Hiervoor geldt:

$$N_h = S_h + E_h + I_h + R_h \quad (2.17)$$

$$N_v = S_v + E_v + I_v \quad (2.18)$$

Daarnaast wordt de infectieratio van muggen tot mensen gegeven door  $\lambda_h$  en andersom de infectieratio van mensen tot muggen gegeven door  $\lambda_v$ . Deze worden gedefinieerd door:

**Definitie 1** (infectie ratio's  $\lambda_i$ ). *We beginnen met het definiëren van  $\sigma_v$ , het maximale aantal keer dat een mug per tijdseenheid mensen kan bijten en  $\sigma_h$  met maximale aantal keer dan een*

mens gebeten kan worden per tijdseenheid. Het totale aantal keer mens-mug contact wordt gegeven door  $b = \frac{\sigma_v \sigma_h N_v N_h}{\sigma_v N_v + \sigma_h N_h}$ . Wanneer we deze breuk nog eens delen door  $N_h$  of  $N_v$  dan krijgen we de volgende vergelijkingen voor infectie ratio's mensen-mug en mug-mens:

$$\lambda_h := \frac{\sigma_v \sigma_h N_v}{\sigma_v N_v + \sigma_h N_h} \beta_{hv} \frac{I_v}{N_v}$$

$$\lambda_v := \frac{\sigma_v \sigma_h N_h}{\sigma_v N_v + \sigma_h N_h} \beta_{hv} \frac{I_h}{N_h}$$

Zo zien we dat  $\lambda_h$  afhangt van het aantal beten per mens per tijdseenheid  $b/N_h$ , de kans dat een mug geïnfecteerd is,  $I_v/N_v$  en de kans op transmissie van de ziekte door een beet van mug naar mens  $\beta_{vh}$ . Dit geldt op dezelfde wijze voor  $\lambda_v$ .

In het model wordt aangenomen dat de populatie van mensen constant blijft, dus dat geboortecijfer  $\psi$  gelijk is sterftcijfer  $\mu$  voor mensen, voor muggen wordt dit niet gedaan, om zo het effect te kunnen meten van het bestrijden van bijvoorbeeld levende muggen of larven of te wel de sterfte- en geboortecijfers van muggen. Ook is het goed om op te merken dat muggen niet in de genezen staat  $R_v$  komen, maar gedurende hun levensduur besmettelijk blijven.

Met deze differentiaal vergelijkingen kan men het verloop van de aantallen  $S_h, E_h, I_h, R_h$  en  $S_v, E_v, I_v$  modelleren.

## 2.2 Statistische modellen

Het SIR-model kan ook gemodelleerd worden met een Markov-keten. Hierbij wordt in discrete tijdstappen,  $n$  naar  $n + 1$ , van een bepaalde staat  $(S_n, I_n, R_n)$  naar staat  $(S_{n+1}, I_{n+1}, R_{n+1})$ . Waarbij met kans  $i$  van  $S$  naar  $I$  gegaan kan worden en  $1 - i$  bij  $S$  gebleven wordt. Met kans  $r$  kan van  $I$  naar  $R$  gegaan worden en  $1 - r$  bij  $I$  gebleven worden. Met kans 1 wordt er gebleven bij  $R$ . We krijgen dan de volgende vergelijking:

$$\begin{bmatrix} S_n \\ I_n \\ R_n \end{bmatrix} = \begin{bmatrix} 1 - i & 0 & 0 \\ i & 1 - r & 0 \\ 0 & r & 1 \end{bmatrix} \begin{bmatrix} S_{n-1} \\ I_{n-1} \\ R_{n-1} \end{bmatrix} \quad (2.19)$$

waarbij de vector met startwaarden als volgt gedefinieerd wordt:

$$\begin{bmatrix} S_0 \\ I_0 \\ R_0 \end{bmatrix} = \begin{bmatrix} 1 - \epsilon \\ \epsilon \\ 0 \end{bmatrix} \quad (2.20)$$

voor een bepaalde  $0 \leq \epsilon < 1$ . Het gegeven model leent zich goed voor zowel snelle korte infecties als langdurige ziektes.

### 2.2.1 Van discreet naar continu

Om gemakkelijker op empirisch verzamelde gegevens, zoals gebruikers aantallen van sociale netwerken, analyses te kunnen uitvoeren leent zich continu model beter. In de paper van C. Bauchhage [5] wordt een continue kansdichtheidsfunctie  $f(t)$  gegeven voor vergelijking (2.19), gebaseerd op de convolutie van twee exponentiële functies. Dit geeft de volgende functie:

$$f(t) = \begin{cases} \frac{\alpha}{\alpha - \lambda} \lambda e^{-\lambda t} + \frac{\lambda}{\lambda - \alpha} \alpha e^{-\alpha t} & \text{if } \lambda \neq \alpha \\ \lambda^2 t e^{-\lambda t} & \text{if } \lambda = \alpha \end{cases} \quad (2.21)$$

Omdat  $f(t)$  een kansdichtheidsfunctie is leent het zich goed voor statistische methoden om het model te fitten. Er kan direct gebruik gemaakt worden van maximum likelihood methodes voor het fitten van de modelparameters. Veel gebruikte methoden zijn verdelingen gebaseerd op de Weibullverdeling of een LogNormale kansdichtheidsfunctie, een voordeel van de gevonden functie is echter dat de parameters  $\lambda$  en  $\alpha$  in een één op één relatie staat met infectieratio  $i$  en herstelratio  $r$ .

## 2.3 Willekeurig netwerkgraafmodel

In de paper “An alternative Explanation for the Rise and fall of Myspace” [4] en in de paper “Rise and decline process of online communities: Modeling social balance of participants” [2] worden sociale netwerken niet gemodelleerd met differentiaalvergelijkingen, maar door het bekijken van een grafen-netwerk. In het model van Chhabra [4] wordt er onderscheid gemaakt tussen zo genoemde sterke vriendschappen in een online sociaal netwerk en normale vriendschappen, aangezien iedereen maar met een aantal, (“Dunbars number” [2]), een hechtere vriendschap kan onderhouden. Vaak heeft iemand met (veel) meer mensen een connectie op een online sociaal netwerk dan alleen de mensen waarmee deze een hechte vriendschap onderhoudt. Echter heeft elke gebruiker altijd een beperkt aantal mensen waarmee deze een sterke vriendschap onderhoudt. Uit een empirisch onderzoek door Onnella et al [11], blijkt dat mensen die een sterke vriendschap met elkaar hebben ook vaak een grote overlap hebben in het aantal gemeenschappelijke vrienden, of in de termen van grafen: een grotere *neighbourhood overlap*. Als A en B twee punten in een netwerk zijn, of wel twee vrienden, dan definiëren we: neighbourhood overlap door:

**Definitie 2** (Neighbourhood overlap). *het aantal gemeenschappelijke vrienden van A en B gedeeld door het aantal vrienden wat met A en of met B is bevriend.*

In het model worden twee zaken bijgehouden:

1. De verandering in het aantal mensen die minimaal een bepaald aantal sterke vriendschappen onderhouden in een groeiend sociaal netwerk. Een vriendschap noemen we sterk wanneer de neighbourhood overlap een bepaalde drempelwaarde  $\epsilon$  overschreid. Het netwerk wordt gecreëerd door het Barabasi-Albert willekeurige grafen model, zie sectie 2.3.1
2. De verandering van de grootte van de grootste samenhangende subgraaf (ook wel de kern genoemd) van het gegenereerde netwerk wordt bijgehouden. Dit is interessant omdat de kern het meeste invloed heeft op het behouden van gebruikers van een sociaal netwerk. (verwijzing 10 uit artikel)

### 2.3.1 Het Barabasi Albert model voor sociale netwerken

Voor een willekeurig netwerkgraafmodel kan men het Barabasi-Albert model gebruiken om een willekeurig netwerk te genereren.

Het Barabasi-Albert model is een algoritme met onder andere de eigenschap dat een knoop in het netwerk met veel connecties waarschijnlijker is om een nieuwe knoop er aan verbonden te krijgen wanneer het netwerk groeit.

Het netwerk start met een verzameling  $m_0$  verbonden knooppunten en het algoritme voegt steeds een knoop toe aan het netwerk, verbonden met  $m \leq m_0$  bestaande knopen. De kans  $p_i$  dat een bestaande knoop  $i$  verbonden wordt met de nieuwe knoop is proportioneel aan het aantal knopen waar deze knoop nu al aan verbonden  $d_i$  is. We krijgen dan:  $p_i = \frac{d_i}{\sum_j d_j}$ . Met deze kans krijgt nu knoop  $i$  een verbinding met de knoop. Dit gebeurt voor alle bestaande knopen en zo wordt een netwerk gegenereerd.



### 2.3.2 Model

In het artikel wordt in de simulaties gestart met vaste waarden voor het aantal knopen in het netwerk en vaste drempelwaarden  $\epsilon$  voor sterke vriendschappen. De resultaten gaven grafieken die het aantal gebruikers die meer dan 25 sterke connecties hadden tegenover de totale grote van het netwerk. Hier is bij verschillende drempelwaarden en verschillende waarden voor  $m$ , het aantal knopen waarmee een nieuwe knoop werd verbonden, te zien dat eerst altijd het aantal mensen dat voldoet aan het criterium groeit, tot de graaf een kritieke grootte heeft bereikt. Daarna daalt het aantal gebruikers dat voldoet. Dit komt natuurlijk omdat het lastiger is om te voldoen aan de (relatieve) drempelwaarde  $\epsilon$  naarmate de graaf groter wordt. Ook voor de eerder genoemde kern, de grootste samenhangende graaf van sterke vriendschappen, geldt dat de grootte er van stijgt, tot een bepaalde kritieke grootte bereikt wordt en de grootte daarna afneemt. Ook hier ligt de oorzaak bij het voldoen aan  $\epsilon$ .

Uiteindelijk wordt er geconcludeerd dat MySpace moeite had met het behouden van zijn gebruikers, nadat het op beide vlakken zijn maximale grootte had bereikt, omdat voor veel gebruikers relatief het aantal sterke vriendschappen daalde en ze vervolgens het netwerk verlaten.

### 2.3.3 Heidler's balanstheorie

Een andere manier, besproken in het artikel van Kimitaka Asatani et al [2], om de groei en afname van online sociale netwerken te modelleren is gebaseerd op de balans theorie van Heidler [3]. Deze theorie bekijkt van drie objecten onderling waar een voorkeur voor of afkeer tegen is. In het geval van een sociaal netwerk waarin berichten geplaatst worden over bepaalde onderwerpen zijn die drie objecten: twee personen en een onderwerp of drie personen. De onderlinge 'likes' en 'dislikes' van deze objecten vormen stabiele en onstabiele relaties. Wanneer iemand tegenover een vriend én een onderwerp positief is en die vriend ook positief is over dat onderwerp, dan vormt dit bijvoorbeeld een stabiele driehoeksverhouding. Wanneer men echter positief is over een onderwerp en een vriend, maar die vriend niet positief tegenover dat onderwerp staat, dan levert dat een instabiele driehoeksverhouding. Uit deze stabiele en instabiele verhoudingen ontstaan weer acties die zorgen voor meer balans in alle relaties van een persoon. Zo zou iemand zijn houding kunnen veranderen ten opzichte van een onderwerp, er juist positiever op kunnen reageren of besluiten er niets van te zeggen. Ten slotte kan iemand ook niet meer geïnteresseerd raken in een bepaald onderwerp en zich interesseren in een ander onderwerp, of geïnteresseerd raken in een extra onderwerp. Het model voert steeds een 'update-strategie' uit die een gebruiker doet om te zorgen dat zijn totale verhoudingen beter in balans zijn indien mogelijk.

De groei van een netwerk ontstaat in dit model wanneer meer en zelfs nieuwe mensen zich binden aan bepaalde onderwerpen en hoe langer deze daarin geïnteresseerd blijven hoelanger het netwerk natuurlijk blijft bestaan of zelfs groeit. En wanneer uit de balans volgt dat minder mensen zich interesseren in minder onderwerpen, dan daalt het gebruik van het netwerk. Op deze manier wordt in dit model dus de groei en afname van het gebruik van online sociale netwerken gemodelleerd. Voor meer details kan men zich verdiepen in het artikel van Kimitaka Asatani et al [2].

We zien dat dit model verschilt van het eerder genoemde Barabasi model. Dit model neemt ook het gedrag van netwerkgebruikers mee en laat zien hoe dit gedrag effect heeft op de afname van het gebruik van het netwerk. Een bezwaar tegen het Barabasi en SIR-modellen zoals die van Canarella [1] of Maring [10] is dat men niet kijkt naar het achterliggende gedrag dat deze groei en afnames van gebruik veroorzaakt.



## Hoofdstuk 3

# Het leiders-volgersmodel

In dit hoofdstuk bekijken we het leiders-volgersmodel. Dit model is een voor zover bekend aan de auteur een nieuw model. Aan de hand van het Zikavirusmodel, eerder besproken in sectie 2.1.6, is dit model tot stand gekomen. Voor het leiders-volgersmodel wordt een gereduceerde vorm van het Zika-virus model gebruikt en aangepast zodat het toegepast kan worden voor het modelleren van online sociale netwerken. In dit hoofdstuk bekijken we net als bij het Zikavirusmodel een SIR-model met twee verschillende populaties, echter zijn het geen mensen en muggen, maar leiders en volgers in het gebruik van online sociale netwerken.

Zoals Rogers beschrijft in zijn boek ‘Diffusion of innovations’ [12] zijn er bij het gebruik van nieuwe technologieën altijd innovators (2,5% van de populatie), early adopters (13,5%), early majority (34%), late majority (34%) en laggards (16%). We zullen de eerste twee groepen van de populatie, innovators en early adopters zien als leiders in het gebruik van online sociale netwerken. De latere drie groepen zien we als volgers.

De twee populaties, leiders en volgers, bestaan elk uit de bekende vatbare personen  $S_l$  en  $S_v$ , geïnfecteerde personen  $I_l$  en  $I_v$  en uiteindelijk resistente ex-gebruikers  $R_l$  en  $R_v$ . In dit model nemen we niet alleen de infectiegraad en de genezingsgraden mee, maar ook het effect van de leiders op de volgers. Zo beïnvloeden de leiders de volgers om ook het sociale netwerk te gebruiken  $\lambda_I$  en wanneer zij stoppen met het gebruik zullen ook de volgers langzamerhand het online sociale netwerk minder gaan gebruiken  $\lambda_R$ . Deze  $\lambda_I$  en  $\lambda_R$  hangen beide af van de verhouding geïnfecteerde leiders en genezen leiders ten opzichte van de totale populatie. We definiëren deze lambda's als volgt:

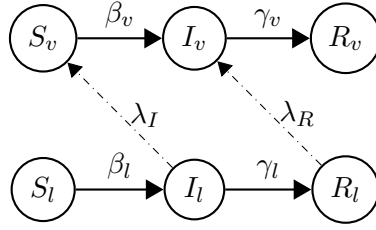
**Definitie 3** ( $\lambda_I$  en  $\lambda_R$ ).

$$\lambda_I := \sigma_I \frac{I_v}{N_v}$$

$$\lambda_R := \sigma_r \frac{I_l}{N_l}$$

met  $\sigma_I$  en  $\sigma_r$  parameters die het effect van de verhouding geïnfecteerde en resistente leiders schalen.

In figuur 3.1 zijn de verhoudingen schematisch weergegeven.



Figuur 3.1: Illustratie van het leider-volgmmodel

Aan de hand van de volgende differentiaal vergelijkingen kan vervolgens het leidervolg-model worden gemodelleerd.

$$\frac{dS_v}{dt} = \frac{-\beta_v I_v S_v}{N_v} - \lambda_I S_v \quad (3.1a)$$

$$\frac{dI_v}{dt} = \frac{\beta_v I_v S_v}{N_v} + \lambda_I S_v - \gamma_v I_v - \lambda_R I_v \quad (3.1b)$$

$$\frac{dR_v}{dt} = \gamma_v I_v + \lambda_R I_v \quad (3.1c)$$

$$\frac{dS_l}{dt} = \frac{-\beta_l I_l S_l}{N_l} \quad (3.1d)$$

$$\frac{dI_l}{dt} = \frac{\beta_l I_l S_l}{N_l} - \gamma_v I_v \quad (3.1e)$$

$$\frac{dR_l}{dt} = \gamma_v I_v \quad (3.1f)$$

In vergelijking (3.1a) en (3.1b) staat de term  $\lambda_I S_v$ . Deze term geeft de invloed weer van de geïnfekteerde leiders op de vatbare volgers. Wanneer deze verhouding toeneemt, zal het aantal vatbare volgers sneller afnemen en het aantal geïnfekteerde volgers sneller toenemen, vandaar het tekenverschil in beide vergelijkingen voor deze term.

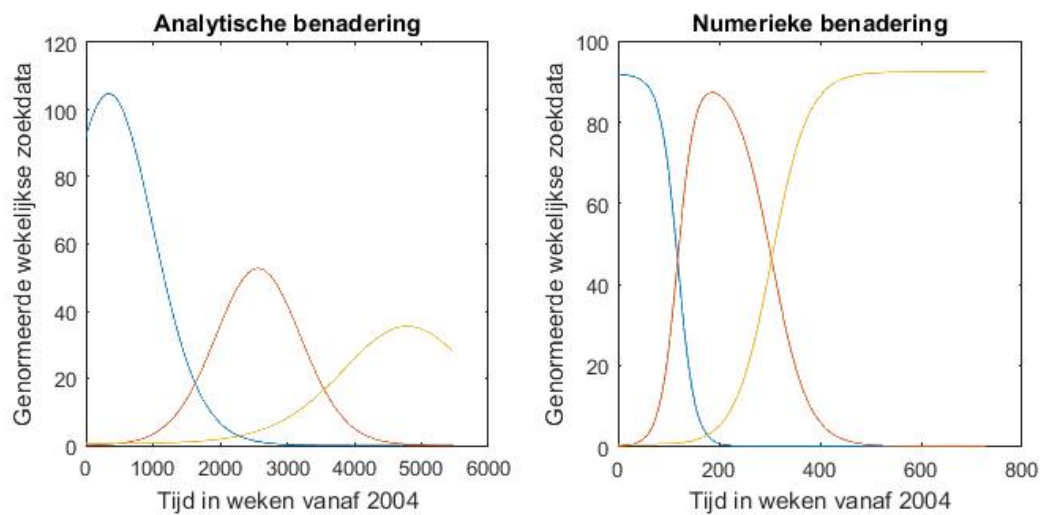
In vergelijking (3.1b) en (3.1c) komt de term  $\lambda_R I_v$  voor. Deze term modelleert de invloed van resistente leiders op geïnfekteerde volgers om ook resistent te worden. Wanneer er relatief meer resistente leiders zijn zal het aantal geïnfekteerde volgers sneller afnemen en het aantal resistente volgers weer toenemen.

## Hoofdstuk 4

# Resultaten

### 4.1 Vergelijking analytische en numerieke benadering irSIRS model

Aan de hand van in sectie 2.1.4 beschreven analytische benadering van het model vergelijken we de resultaten met de numerieke benadering van het irSIRS model. Voor beide modellen zijn dezelfde parameters en startwaarden gebruikt.



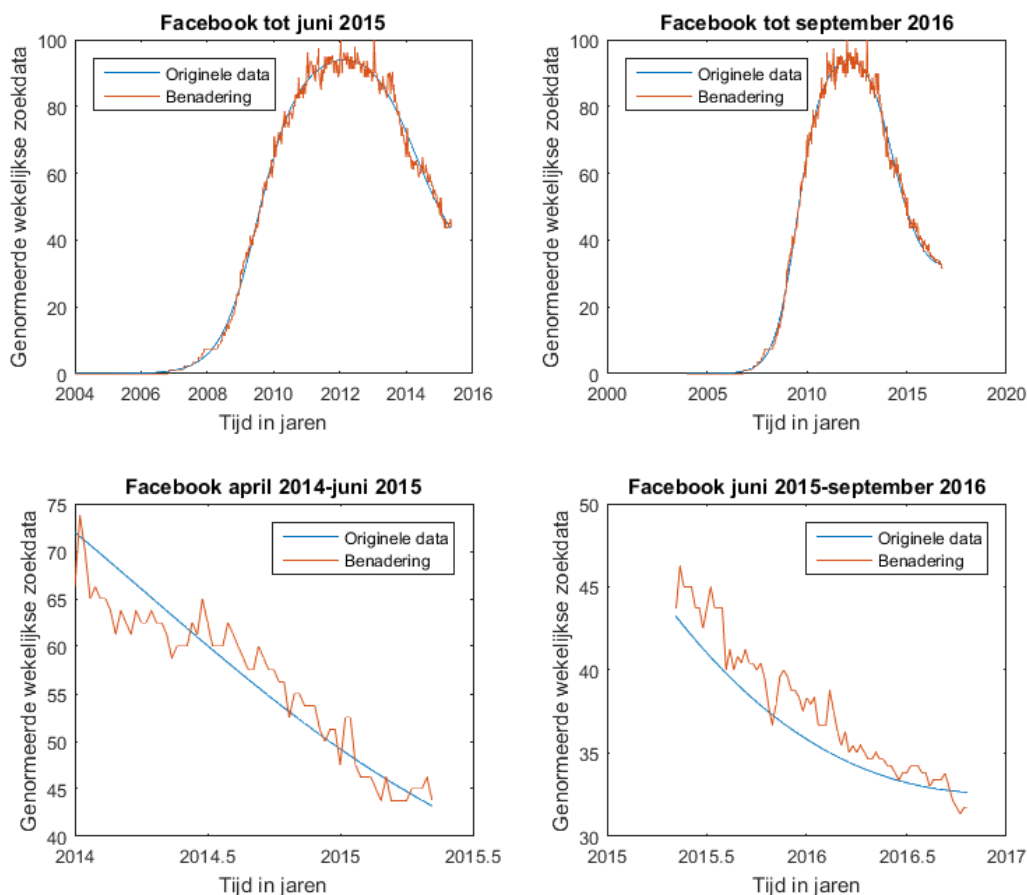
Figuur 4.1: plot van irSIR model analytische benadering en het numerieke model

We zien dat de analytische benadering wel vormen aanneemt die we verwachten bij een SIR-model, echter zien we ook direct dat de analytische benadering met ongeveer een factor 10 er naast zit op de tijdschaal. Hiervoor is geen verklaring gevonden. De berekeningen om tot

deze vergelijkingen te komen zijn gecontroleerd met Maple. Het verschil, los van de tijdschaal, in de vorm van de grafieken zou te wijten kunnen zijn aan de redelijk grote afbreekfout die ontstaat bij de benadering. Bij de benadering van  $e^x$  met het eerste orde Taylor polynoom  $1 + x$  wordt natuurlijk veel ruimte gegeven voor fouten. De benadering is een gebruikelijke benadering voor kleine waarden voor  $x$ . We doen hier echter een voorspelling over meerdere jaren en dus honderden weken, dit zou dus ook kunnen meespelen in de bruikbaarheid van het model, los van de fout van factor 10. Er zou nog gekeken kunnen worden hoe dit model presteert op een hele kleine tijdschaal. Ook zou een tweede orde benadering een beter resultaat kunnen geven.

## 4.2 Nieuwe voorspelling voor Facebook

Bij het implementeren van het irSIRSbd-model dat Cannarella [1] beschrijft, zoals Maring [10] dat gedaan heeft, gebruikmakend van dezelfde geoptimaliseerde parameters uit het verslag van Maring [10], vinden we de volgende resultaten tot en met augustus 2016.



Figuur 4.2: Vier afbeeldingen van de resultaten gebruikmakend van het model van Maring [10]. Linksboven het eerder gevonden resultaat, rechtsboven in de nieuwe resultaten tot en met 31 augustus 2016, rechtsonder hoe het model presteert sinds juni 2015 (ongeveer 70 weken) en links hoe het model de 70 weken ervoor presteerde.

We zien in Figuur 3 linksboven hoe het model geoptimaliseerd werd voor data tot en met juni 2015. In het model rechtsboven worden dezelfde parameters (zie Bijlage C) gebruikt en lijkt niet erg verschillend te presteren ruim een jaar later. Wanneer we echter kijken naar juni 2015 tot september 2016 zien we een verschil met de 70 weken ervoor, het lijkt alsof de voorspelling er consequent onder zit. Wanneer we echter naar de gemiddelde absolute fout (deze gebruiken we in verband met het verschil in precieze tijdspanne in plaats van de sum of squared errors), zien we dat deze kleiner is dan de 70 weken er voor.

Model	MAE
Facebook tot juni 2015	1.432325823523053
Facebook tot september 2016	1.447807723103996
Facebook april 2014-juni 2015	2.441138182291401
Facebook juni 2015-september 2016	1.554616746306263

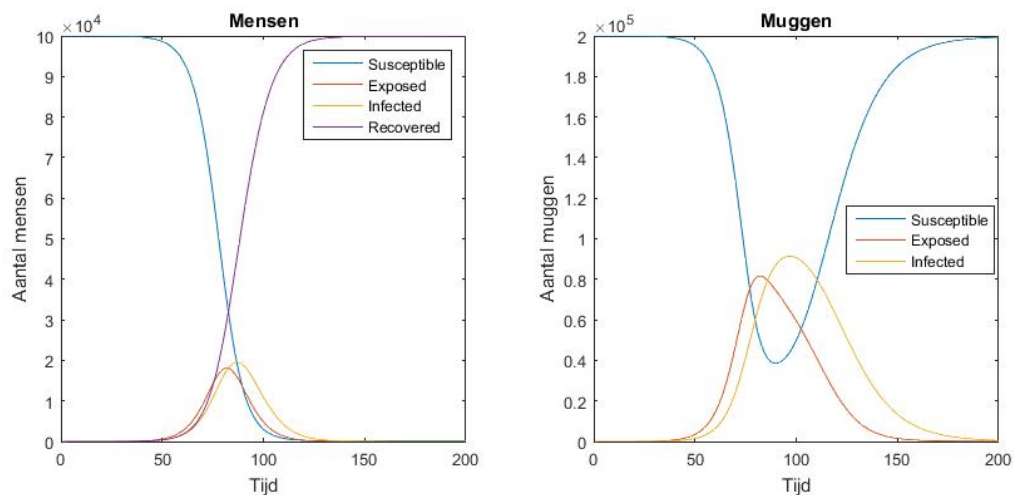
Tabel 4.1: Gemiddelde absolute fouten bij Figuur 4.2

We zien hier dat de gemiddelde fout groter is geworden sinds juni 2015. Erg groot lijkt dit niet te zijn, 1.08%. Wellicht is een kortere tijdspanne niet representatief voor het presteren van het algehele model, zoals we zien bij de gemiddelde absolute fouten van Facebook juni 2015-september 2016 en Facebook april 2014-juni 2015. In sectie 4.4 zien we hoe het opnieuw geoptimaliseerde model presteert tegenover het leiders-volgersmodel.



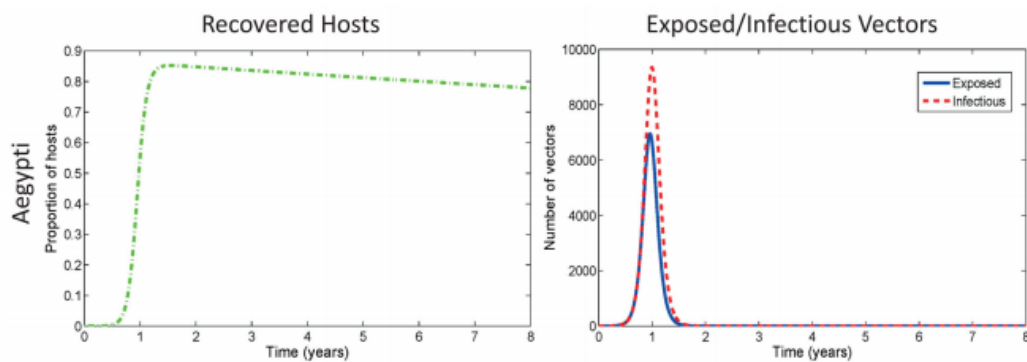
### 4.3 Resultaten Zika-model

Bij het implementeren van het model voor het Zikavirus, zoals in 2.1.5 beschreven, worden vergelijkbare resultaten gevonden als in het referentie artikel [6].

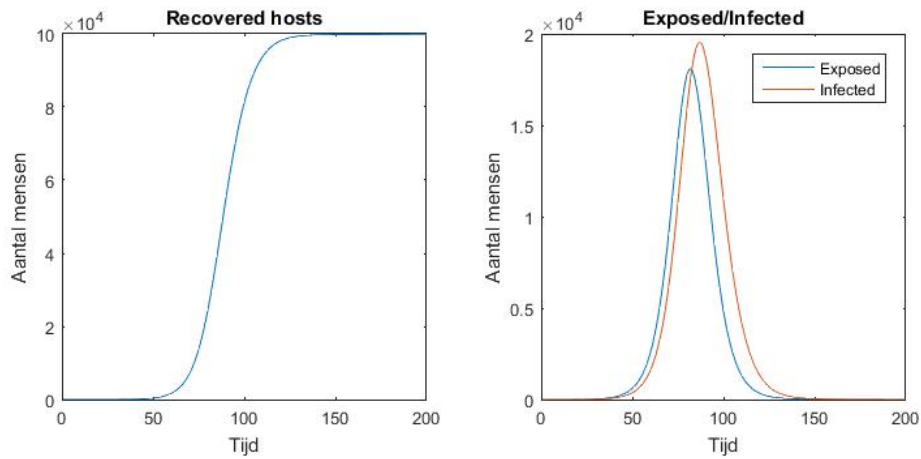


Figuur 4.3: Resultaat overzicht verdeling totale mensen en muggen populaties van het model.

Het aantal vatbare mensen en muggen worden weergegeven aan de hand van de blauwe lijn. We zien hier dat het aantal vatbare muggen weer terug stijgt naar de startwaarde, dit komt omdat de populatie van muggen snel weer groeit, echter niet meer geïnfecteerd raakt, omdat de epidemie voor mensen al voorbij is en er dus geen overdracht meer plaats vindt. In de volgende afbeelding zien we de resultaten die door Carrie A. Manore et al [6] zijn gevonden en hoe dit vergelijkt met de resultaten van het beschreven model wanneer we dezelfde plot genereren.



Figuur 4.4: Afbeelding uit het artikel van Carrie A. Manore et al [6] die resultaten van het Zikamodel weergeven.

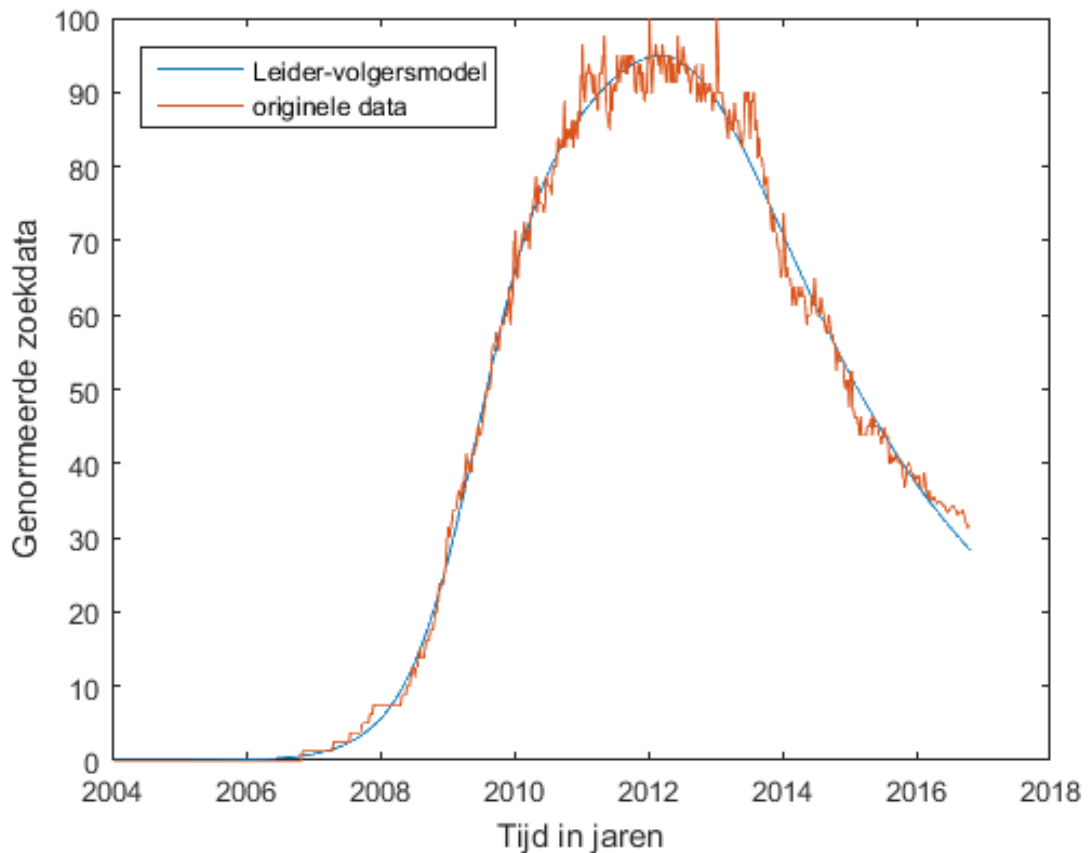


Figuur 4.5: Resultaat van het Zikamodel beschreven in 2.1.5.

We zien vergelijkbare waarden voor zowel de geïnfecteerde mensen, als herstelde mensen. Kleine verschillen zouden kunnen worden verweten aan de gebruikte numerieke methode hier en de modelleermethode in het artikel.

#### 4.4 Resultaten leiders-volgersmodel

Wanneer we het leiders-volgersmodel uit hoofdstuk 3 implementeren in Matlab en de parameters variëren met `fminsearch` om de sum of squared errors (SSE) te minimaliseren, zoals ook gebeurt in het verslag van Maring [10]. Dan verkrijgen we het resultaat in figuur 4.6 wanneer we het naast de originele data van Facebook leggen.



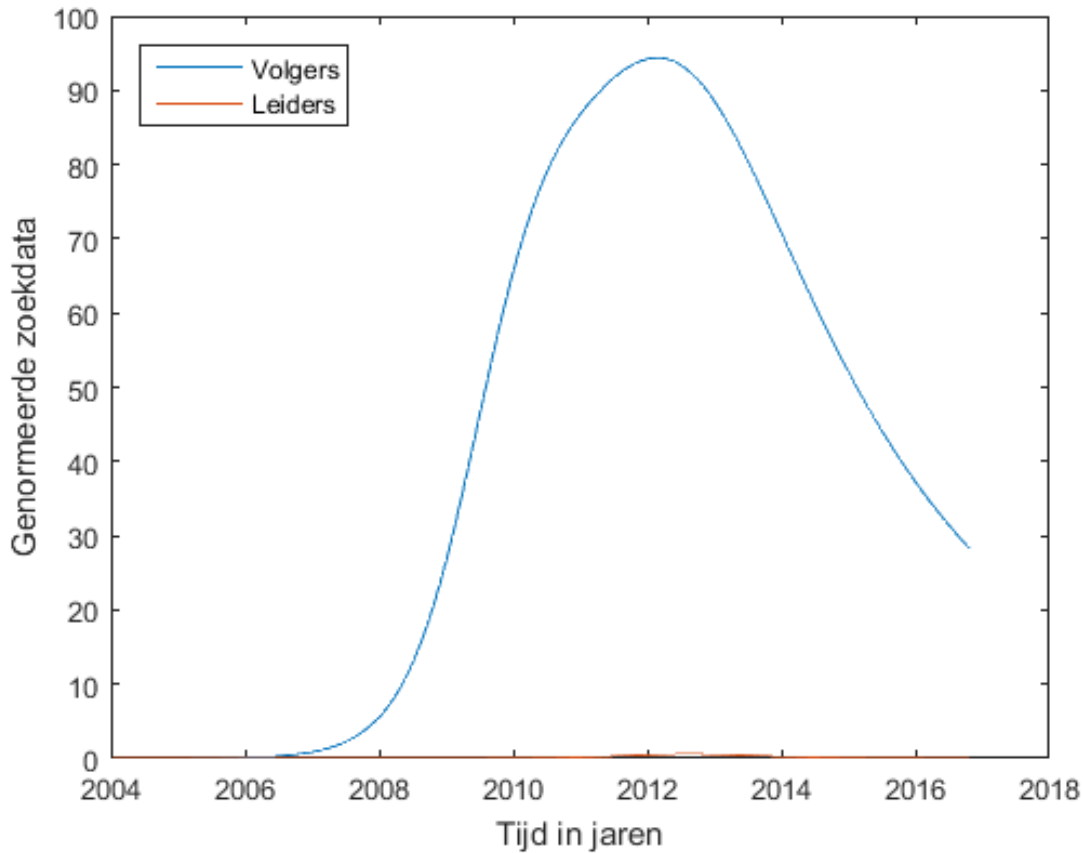
Figuur 4.6: Resultaat na het optimaliseren met `fminsearch` van het leiders-volgersmodel.

De parameters die gevarieerd worden zijn  $\beta_1, \beta_v, \gamma_1, \gamma_v, n, n_v$  en  $r_v_0$  en  $r_l_0$ .  $\beta_1, \beta_v, \gamma_1, \gamma_v$  spreken voorzich waarom ze gevarieerd worden.  $n$  en  $n_v$  zijn respectievelijk de totale populatie en het aantal geïnfecteerde leiders als startwaarde. Anders dan in Maring [10] worden de startwaardes van vatbaren, geïnfecteerden en resistenten niet gevarieerd nu. Dit omdat we bij dit model er van uit gaan dat de verhouding leiders volgers vast ligt en er van uitgaan dat als iemand al gebruikt maakt van een online sociaal netwerk op het start moment deze altijd een early adaptor zou moeten zijn, hij gebruikt het netwerk immers als een van de eersten. Wel worden het aantal resistente gebruikers van beide groepen gevarieerd. Dit omdat het zou kunnen zijn dat van beide groepen er altijd mensen zouden kunnen zijn die bij voorbaat al resistent zijn en dus het netwerk nooit zullen gebruiken. Deze resistente gebruikers, specifiek resistente leiders, kunnen wel nog het aantal geïnfecteerde volgers beïnvloeden volgens het model.

De gevonden sum of squared errors is bij figuur 4.6:  $3.458 \cdot 10^3$ .

#### 4.4.1 Leiders-volgers effect

Om te kunnen bekijken wat het effect is van leiders en volgers, bekijken we de resultaat uit figuur 4.7.



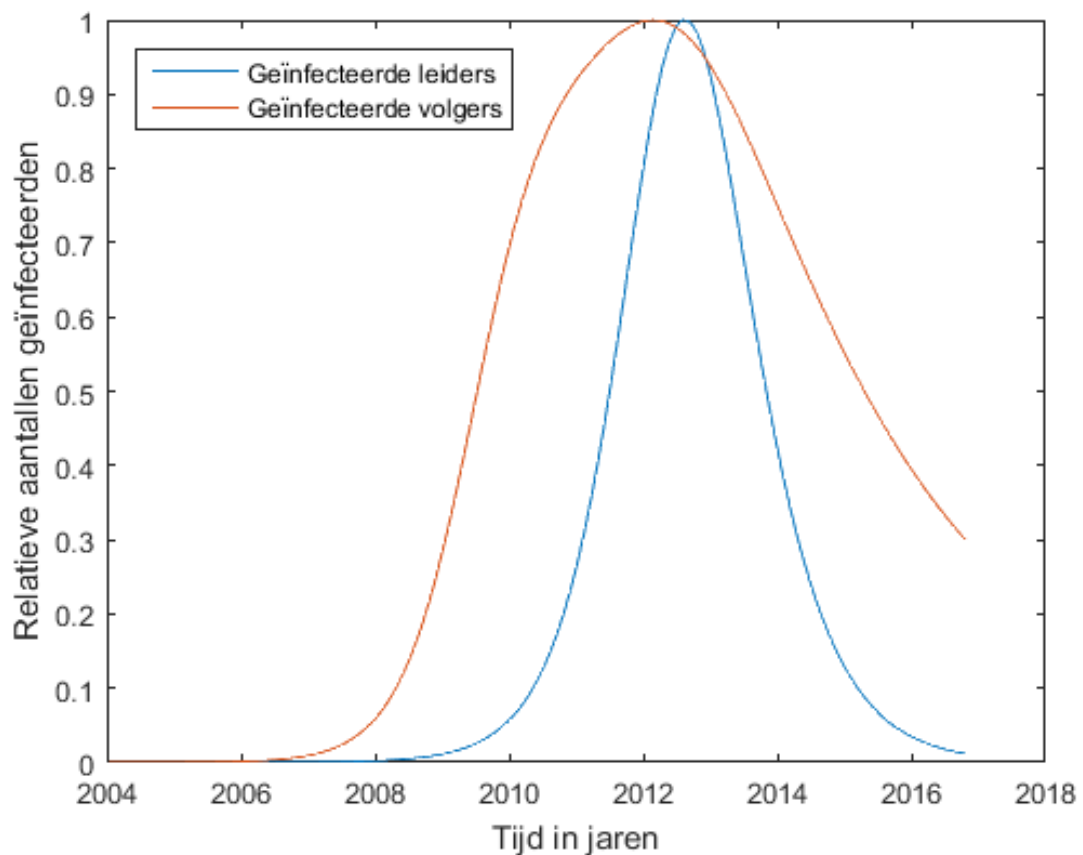
Figuur 4.7: Uitsplitsing van geïnfecteerde leiders en volgers in het leiders-volgersmodel

Parameter	Geoptimaliseerde waarde
$\gamma_v$	0.044329576642123
$\gamma_l$	0.044329576642123
$\beta_v$	0.090546963019920
$\beta_l$	0.094315283723067
$\sigma_I$	4.244972568047686
$\sigma_R$	0.015326573347482
$N$	80.994104287546040
$N_v$	$1.104302509381040 \cdot 10^{-5}$
$r_{v0}$	-2.130130486141346
$r_{l0}$	$1.524628910365468 \cdot 10^2$

Tabel 4.2: Geoptimaliseerde parameterwaarde voor zoekdata Facebook

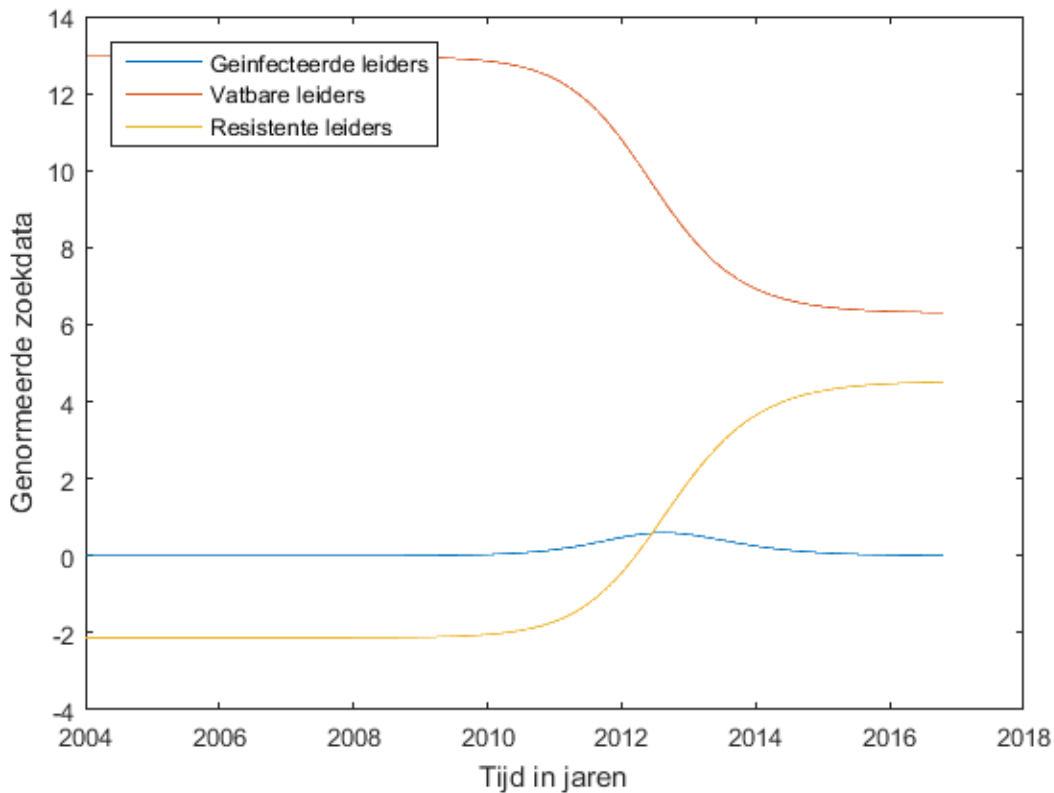
We zien, er zijn maar weinig geïnfecteerde leiders in het geoptimaliseerde model. En wanneer we kijken naar de parameter die de verandering van vatbaar naar geïnfecteerd beïnvloed, dan zien we ook dat deze voor de leiders 0.044 is en voor de volgers bijna twintig keer zo groot: 0.87. De schalingsparameter  $\sigma_I$  is daarentegen ongeveer 4, dit zegt ons dat wanneer er geïnfecteerde leiders zijn ze wel ten opzichte van het totale aantal leiders veel effect hebben op de infectiegraad van volgers. De schalingsparameter  $\sigma_R$  is veel kleiner en beschrijft hoe het relatieve aantal resistente leiders effect heeft op geïnfecteerde volgers.

In figuur 4.8 zien we de relatieve aantallen geïnfecteerde leiders en volgers. We zien als eerste dat het aantal geïnfecteerde leiders later piekt dan het aantal geïnfecteerde volgers. Dit resultaat lijkt vreemd als we verwachten dat leiders eerder het meeste een online sociaal netwerk gebruiken dan volgers. Dit zou verklaard kunnen als bijvoorbeeld de invloed resistente leiders meer meespeelt voor dit model dan het aantal geïnfecteerde leiders.



Figuur 4.8: Relatieve aantallen van geïnfecteerde leiders en volgers ten opzichte van hun maxima

Daarmee komen we op het volgende punt op te bekijken hoe überhaupt de verdeling van de geoptimaliseerde vatbare, geïnfecteerde en resistente leiders ligt.



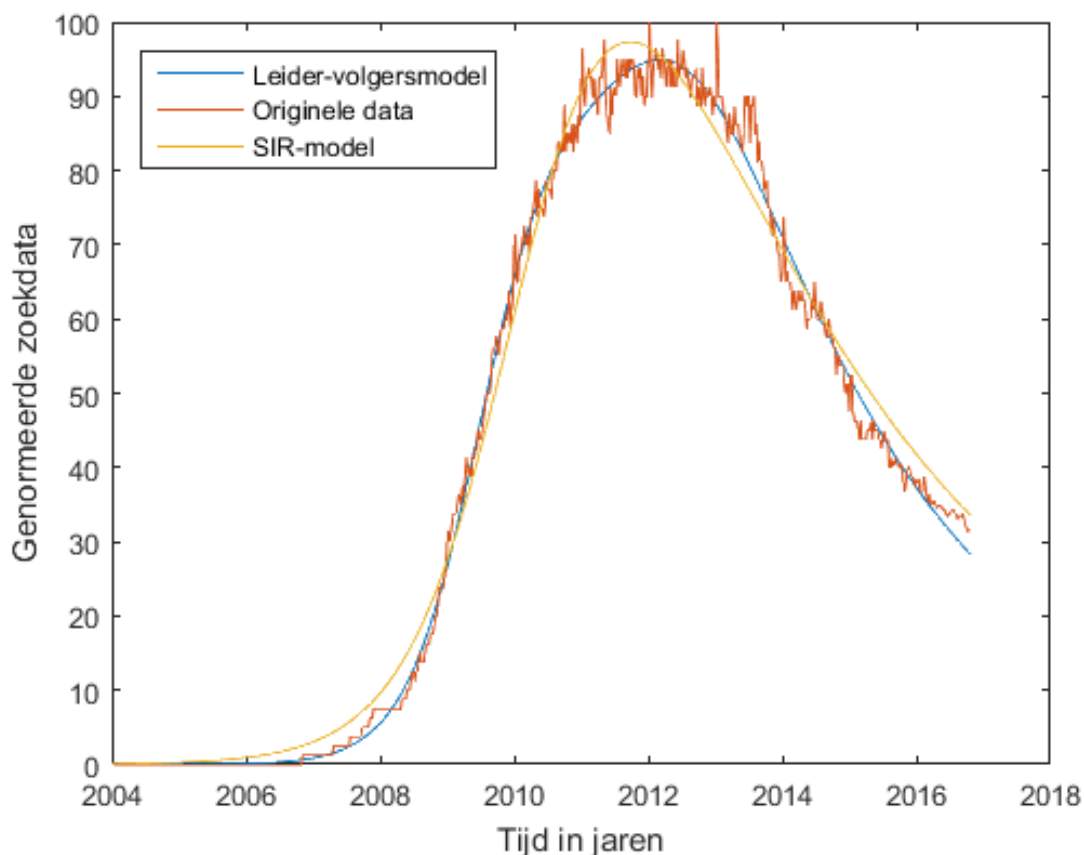
Figuur 4.9: Uitsplitsing van leiders in vatbaren, geïnfecteden en resistenten in het leiders-volgersmodel

In figuur 4.9 zien we de verdeling van  $S_l$ ,  $I_l$  en  $R_l$  in het model. We zien meteen een aantal opvallende resultaten. We zien dat het aantal resistente leiders negatief is in het model. Dit lijkt vreemd en is wellicht onrealistisch, omdat je natuurlijk nooit een negatief aantal mensen kunt hebben. Wanneer we echter kijken in welke periode dit plaats vindt, zien we dat dit aan het begin, het stijgende stuk van het model is. Hier fit het leiders-volgersmodel ook beter dan het SIR-model. Het lijkt er op dat deze negatieve waarden nodig zijn om het aantal geïnfectede volgers zodanig te beïnvloeden dat deze beter passen bij de data.

Er zou nog geoptimaliseerd kunnen worden onder de voorwaarde dat bijvoorbeeld de fysieke aantallen geen negatieve waarden aan kunnen worden. Dit is echter niet gebeurd voor dit verslag en zou in de toekomst als verbetering van het model gedaan kunnen worden.

#### 4.4.2 Leiders-volgersmodel vs SIR-model

Wanneer we het leiders-volgersmodel vergelijken met het basis SIR model, dan verkrijgen we het volgende resultaat:



Figuur 4.10: Vergelijking leiders-volgersmodel en het SIR-model.

We zien, het leiders-volgersmodel presteert een stuk beter dan het opnieuw geoptimaliseerde SIR-model. Voor het SIR-model is opdezelfde manier als bij Maring [10] geoptimaliseerd, alleen dan met de Google trends data tot september 2016. We zien hier dat het meenemen van leiders en volgers een beter resultaat oplevert. Ook wanneer we de SSE's vergelijken van beide modellen zien we dat deze meer dan twee keer zo goed presteert.

Model	SSE
Leiders-volgersmodel	$3.458 \cdot 10^3$
Basis SIR-model (geoptimaliseerd voor data tot sept. 2016)	$8.467 \cdot 10^3$

Tabel 4.3: Sum of squared errors bij figuur 4.10

We zien dat het anders structureren van het leiders-volgersmodel dan het SIR-model, met meerdere groepen een zeer sterk effect heeft op het presteren van het model. Deze uitbreiding verschilt van de toevoegingen aan het model, zoals de geboorte- en sterftcijfers, dat niet alleen effecten toevoegt maar het model structuur anders is en blijkbaar daarmee beter kan passen bij de groeidata. Dit geeft de gedachte dat een eenvoudig SIR-model, waarbij de bevolking in slechts drie groepen wordt opgedeeld niet optimaal is. Als laatste kan nog worden opgemerkt dat bij een SIR-model het niet uitmaakt hoeveel resistenten er zijn, omdat dat aantal geen invloed heeft op het model. dit wordt zowel bij toevoegingen op het SIR-model als bij dit model wel verbeterd en zal daarom ook wezenlijk verschillen van het basis model.

### 4.4.3 Leiders-volgersmodel vs irSIRbd-model

Wanneer we het leider-volgmodel vergelijken met het irSIRbd-model, dan zien we dat het irSIRbd-model beter presteert van het leider-volgmodel. Met een SSE van kleiner dan  $2.88 \cdot 10^3$  zien we dat dit ook, het verschil is echter niet zo groot als bij het SIR-model. Beide zijn modellen waarbij bij het optimaliseren van parameters er veel parameters worden gevarieerd: acht parameters bij het irSIRbd-model en tien bij het leiders-volgersmodel worden er gevarieerd. Het irSIRbd-model presteert dus met acht parameters beter dan het leiders-volgersmodel met tien parameters. We kunnen hier uit concluderen dat het irSIRbd-model relatief beter presteert.

Daarnaast moet worden opgemerkt dat qua toevoegingen dit leider-volgersmodel het simpelste model is wat afgeleid kan worden uit het Zika-virus model. Aan dit leiders-volgersmodel zouden vergelijkbare parameters als geboorte- en sterftcijfers kunnen worden toegevoegd, alsmede parameters die de invloed van nieuwe verbanden weergeven, zoals het effect van resistenten op vatbaren. Deze toevoegingen die al in het irSIRbd-model zitten zouden dus aan het leiders-volgersmodel kunnen worden toegevoegd, waardoor deze uiteindelijk nog beter zou kunnen presteren en eventueel ook meer realistische resultaten kunnen geven (zie negatieve waarden voor  $R_t$ ).



# Hoofdstuk 5

## Conclusie en Discussie

### 5.1 Conclusie

We hebben gezien dat de gemaakte analytische benadering van het model geen bruikbaar resultaat heeft opgeleverd. Suggesties om verbeteringen te vinden, zouden een betere analytische benaderingsmethode, zoals de tweede orde term van het Taylorpolynoom er bij te nemen, zoals ook door Kazuyuki Fujii [9] gesuggereerd wordt. Dit zal echter de analytische berekeningen veel moeilijker maken.

We hebben gezien dat het irSIRSbd-model nog steeds een goede benadering geeft voor de zoekdata van Facebook. De gemiddelde absolute fout is met iets meer dan 1% toegenomen. Een zeer kleine toename, zeker wanneer men in acht neemt dat de parameters daarvoor niet opnieuw zijn geoptimaliseerd.

De implementatie van het Zika-model leverde vergelijkbare resultaten als in het referentie-artikel [6] en gaf een goede basis om het leiders-volgersmodel af te leiden.

Met het leiders-volgersmodel kunnen goede resultaten behaald worden om de zoekdata van Facebook te modelleren. Zoals eerder genoemd levert de andere structuur van het model, ten opzichte van het SIR-model zeker toegevoegde waarde. Echter, het is de vraag of de implementatie die we nu aan die nieuwe groepen geven, leiders en volgers, wel de juiste zijn, wellicht is er een andere populatieverdeling die beter modelleert wat deze twee verschillende groepen weergeven. Wanneer we het over leiders en volgers hebben zouden we verwachten dat leiders als eerste gebruikmaken van het netwerk, we zien echter dat dezen net na de piek van de volgers en met een veel kleinere piek gebruikmaken van het online sociale netwerk. Het grote effect van leidende ex-gebruikers ( $R_l$ ) lijkt een grote verbetering te geven, dan wanneer dit niet wordt meegegeven, zoals in het SIR-model.

### 5.2 Discussie

In het verslag van Maring [10] werd al geconcludeerd dat een model met meer parameters beter presteerde en dat is ook hier weer het geval. En ook hier geldt dat de rekentijd van de optimalisatie erg groot werd bij het leiders-volgersmodel. Daarnaast geldt voor `fminsearch` ook nog dat we niets weten over convergentie van te voren en dat er altijd nog sprake kan zijn van een lokaal minimum, als er al een minimum is gevonden. Dit maakt het altijd onzeker of men wel echt de optimale parameters binnen de grenzen van de mogelijkheden heeft gevonden. De berekeningen zijn een aantal keer opnieuw gedaan, net zo lang tot er minimale verschillen ontstonden.

Zoals hierboven genoemd, de nieuwe structuur van het leidersvolgers-model is gebaseerd op een model voor het Zika-virus. Deze twee toepassingsgebieden liggen ver van elkaar en hoewel er

goed te valideren is of het idee van het Zika-model goed is, is het niet zeker of leiders en volgers wel écht effect hebben op de groei en afname van het gebruik van online sociale netwerken.

# Appendices



## Bijlage A

# Uitwerking vergelijkingen sectie 2.1.4

Het terug transformeren van de vergelijking zo als gedaan door Kazuyuki Fujii[9] bij de volgende uitdrukking:

$$\begin{pmatrix} \bar{s}(t) \\ \bar{i}(t) \\ \bar{r}(t) \end{pmatrix} = \exp\left(\frac{t}{N} \begin{pmatrix} 0 & -\beta & 0 \\ \beta & 0 & -\nu \\ 0 & \nu & 0 \end{pmatrix}\right) \begin{pmatrix} \bar{s}(0) \\ \bar{i}(0) \\ \bar{r}(0) \end{pmatrix} \quad (\text{A.1})$$

We transformeren terug naar  $(s(t), i(t), r(t))$ :

$$\begin{pmatrix} s(t) \\ i(t) \\ r(t) \end{pmatrix} = \left[ \exp\left(\frac{t}{N} \begin{pmatrix} 0 & -\beta & 0 \\ \beta & 0 & -\nu \\ 0 & \nu & 0 \end{pmatrix}\right) - \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & & 1 \end{pmatrix} \right] \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} + \exp\left(\frac{t}{N} \begin{pmatrix} 0 & -\beta & 0 \\ \beta & 0 & -\nu \\ 0 & \nu & 0 \end{pmatrix}\right) \begin{pmatrix} s(0) \\ i(0) \\ r(0) \end{pmatrix} \quad (\text{A.2})$$

Wanneer we de matrix exponent uitrekenen levert dit het volgende:

$$\left[ \begin{array}{ccc} \frac{1}{\beta^2 + \nu^2} \left( \nu^2 + \beta^2 \cos\left(\frac{t \sqrt{\beta^2 + \nu^2}}{N}\right) \right) & -\frac{\beta}{\sqrt{\beta^2 + \nu^2}} \sin\left(\frac{t \sqrt{\beta^2 + \nu^2}}{N}\right) & \frac{\beta \nu}{\beta^2 + \nu^2} \left( 1 - \cos\left(\frac{t \sqrt{\beta^2 + \nu^2}}{N}\right) \right) \\ \frac{\beta \nu}{\beta^2 + \nu^2} \sin\left(\frac{t \sqrt{\beta^2 + \nu^2}}{N}\right) & \cos\left(\frac{t \sqrt{\beta^2 + \nu^2}}{N}\right) & -\frac{\nu}{\sqrt{\beta^2 + \nu^2}} \sin\left(\frac{t \sqrt{\beta^2 + \nu^2}}{N}\right) \\ \frac{\beta \nu}{\beta^2 + \nu^2} \left( 1 - \cos\left(\frac{t \sqrt{\beta^2 + \nu^2}}{N}\right) \right) & \frac{\nu}{\sqrt{\beta^2 + \nu^2}} \sin\left(\frac{t \sqrt{\beta^2 + \nu^2}}{N}\right) & \frac{1}{\beta^2 + \nu^2} \left( \beta^2 + \nu^2 \cos\left(\frac{t \sqrt{\beta^2 + \nu^2}}{N}\right) \right) \end{array} \right] \quad (\text{A.3})$$

Deze matrix vullen we nu in in vergelijkg(A.2) en om de leesbaarheid te bevorderen, vervangen we  $\sqrt{\beta^2 + \nu^2}$  door  $\phi$ . We verkrijgen nu de uitdrukkingen voor  $s(t), i(t)$  en  $r(t)$ .

$$s(t) = \frac{1}{\phi^2} \left( \nu^2 + \cos\left(\frac{t\phi}{N}\right) \beta^2 \right) - 1 - \frac{\beta}{\phi} \sin\left(\frac{t\phi}{N}\right) + \frac{\beta\nu}{\phi^2} \left( 1 - \cos\left(\frac{t\phi}{N}\right) \right) \quad (\text{A.4})$$

$$+ \frac{s(0)}{\phi^2} \left( \nu^2 + \cos\left(\frac{t\phi}{N}\right) \beta^2 \right) - \frac{\beta i(0)}{\phi} \sin\left(\frac{t\phi}{N}\right) + \frac{r(0)\beta\nu}{\phi^2} \left( 1 - \cos\left(\frac{t\phi}{N}\right) \right)$$

$$i(t) = \frac{\beta\nu}{\phi^2} \sin\left(\frac{t\phi}{N}\right) + \cos\left(\frac{t\phi}{N}\right) - 1 - \frac{\nu}{\phi} \sin\left(\frac{t\phi}{N}\right) + \frac{\beta s(0)}{\phi} \sin\left(\frac{t\phi}{N}\right) \quad (\text{A.5})$$

$$+ \cos\left(\frac{t\phi}{N}\right) i(0) - \frac{\nu r(0)}{\phi} \sin\left(\frac{t\phi}{N}\right)$$

$$r(t) = \frac{\beta\nu}{\phi^2} \left( 1 - \cos\left(\frac{t\phi}{N}\right) \right) + \frac{\nu}{\phi} \sin\left(\frac{t\phi}{N}\right) + \frac{1}{\phi^2} \left( \beta^2 + \nu^2 \cos\left(\frac{t\phi}{N}\right) \right) \quad (\text{A.6})$$

$$- 1 + \frac{s(0)\beta\nu}{\phi^2} \left( 1 - \cos\left(\frac{t\phi}{N}\right) \right) + \frac{\nu i(0)}{\phi} \sin\left(\frac{t\phi}{N}\right) + \frac{r(0)}{\phi^2} \left( \beta^2 + \nu^2 \cos\left(\frac{t\phi}{N}\right) \right)$$

## Bijlage B

# Gebruikte Matlab-codes

Bij sommige modellen is de Matlabcode uit het verslag van Maring [10] gebruikt of licht aangepast zodat representatieve vergelijkingen gemaakt kunnen worden en waar van toepassing op haar werk voortgeborduurd kon worden.

### B.1 Matlabcode bij het SIR-model ter referentie

```
clc;
clear all;

%parameters
beta = 4.92*10^-2;
gamma = 5.39;
N = 322;
dt = 1;
t_size = 571;

%initialisatie S, I, R
S = zeros(t_size,1);
I = zeros(t_size,1);
R = zeros(t_size,1);

S(1)=0.996*N;
I(1)= 4.1*10^-3*N;
R(1)=0*N;

for i = 2:dt:t_size
    S(i) = S(i-1) - (beta*I(i-1)*S(i-1))*dt/N;
    I(i) = I(i-1) + (beta*I(i-1)*S(i-1))*dt/N - gamma*I(i-1)*dt/N;
    R(i) = R(i-1) + gamma*I(i-1)*dt/N;
end

figure
plot(1:t_size,S,1:t_size,I,1:t_size,R)
```

### B.2 Matlabcode bij het irSIRbd-model

```
clc;
clear all;
```

```

original = xlsread('FacebookWorldJuni.xlsx', 'E2:E667');
% parameter waarden
%%% Facebook
alpha = 1.2392444494510886;
beta = 0.040700260071227;
nu = 0.045411746122411;
s0 = 10.256377170091664;
i0 = 0.089634046479910;
r0 = 57.557789765325481;
b = 0.546388510009792;
d = 0.435866288773971;
t_size = 666;

% Initialiseren %
S = zeros(t_size,1);
I = zeros(t_size,1);
R = zeros(t_size,1);
N = zeros(t_size,1);
% Beginwaarden %
N(1) = s0+i0+r0;
S(1) = s0;
I(1) = i0;
R(1) = r0;

for i=2:t_size
S(i) = S(i-1) + (-beta*S(i-1)*I(i-1) + alpha*R(i-1) - d*S(i-1))/N(i-1) +b;
I(i) = I(i-1) + (beta*S(i-1)*I(i-1) - nu*I(i-1)*R(i-1) - d*I(i-1))/N(i-1);
R(i) = R(i-1) + (nu*I(i-1)*R(i-1) - alpha*R(i-1) - d*R(i-1))/N(i-1);
N(i) = S(i) + I(i) + R(i);
end
week = transpose(1:size(original));
jaar = 2004+week/52;

week2 = transpose(1:590);
jaar2 = 2004+week2/52;

week3 = transpose(590:666);
jaar3 = 2004+week3/52;

week4 = transpose(520:590);
jaar4 = 2004+week4/52;

figure
subplot(2,2,1)
plot(jaar2,I(1:590),jaar2,original(1:590))
title('Facebook tot juni 2015')
legend('Originele data', 'Benadering', 'Location', 'northwest')
xlabel('Tijd in jaren');
ylabel('Genormeerde wekelijkse zoekdata');

subplot(2,2,2)
plot(jaar,I,jaar,original)
title('Facebook tot september 2016')
legend('Originele data', 'Benadering', 'Location', 'northwest')
xlabel('Tijd in jaren');
ylabel('Genormeerde wekelijkse zoekdata');

subplot(2,2,4)
plot(jaar3,I(590:666),jaar3,original(590:666))

```



```

title('Facebook juni 2015–september 2016')
legend('Originele data', 'Benadering', 'Location', 'northeast')
xlabel('Tijd in jaren');
ylabel('Genormeerde wekelijkse zoekdata');

subplot(2,2,3)
plot(jaar4,I(520:590),jaar4,original(520:590))
title('Facebook april 2014–juni 2015')
legend('Originele data', 'Benadering', 'Location', 'northeast')
xlabel('Tijd in jaren');
ylabel('Genormeerde wekelijkse zoekdata');

%fouten:
error1 = I(1:590)-original(1:590);
mae1 = sum(abs(error1))/length(error1);

error2 = I(1:666)-original(1:666);
mae2 = sum(abs(error2))/length(error2);

error3 = I(520:590)-original(520:590);
mae3 = sum(abs(error3))/length(error3);

error4 = I(590:666)-original(590:666);
mae4 = sum(abs(error4))/length(error4);

```

### B.3 Matlabecode bij het leidervolg-model

```

clc;
clear all;
%original = xlsread('FacebookWorldJuni.xlsx', 'E2:E667');

%parameters
gamma_l = 5.39;           %recovery rate infected→resistant
gamma_v = 5.39;           %
beta_v = 4.92*10^-2;      % infectie rate vatbaar→infected
beta_l = 4.92*10^-2;
sigma_I = 0.1*beta_v;     % kans dat geïnfecteerde leider een vatbaar persoon aansteekt.
sigma_R = 0.1*gamma_v;    % kans dat een resistente leider een geïnfecteerde motiveerd om te stoppen
dt = 1;                   % timestep size
t_size = 666;             % time frame data

%initialisatie S_h, I_h, R_h
S_l = zeros(t_size,1);
I_l = zeros(t_size,1);
R_l = zeros(t_size,1);

%initialisatie S_v, E_v, I_v
S_v = zeros(t_size,1);
I_v = zeros(t_size,1);
R_v = zeros(t_size,1);

%initialisatie hulpwaarden
N_l = zeros(t_size,1);
N_v = zeros(t_size,1);
lambda_I = zeros(t_size,1);
lambda_R = zeros(t_size,1);

```

```

%start waardes leiders
N_l(1) = 322*0.16;
S_l(1) = 0.996*N_l(1);
I_l(1) = N_l(1)-0.996*N_l(1);
R_l(1) = 0;

%start waardes volgers
N_v(1) = 332*0.84;
S_v(1) = N_v(1);
I_v(1) = 0;
R_v(1) = 0;

lambda_I(1) = sigma_I*(I_l(1)/N_l(1));
lambda_R(1) = sigma_R*(R_l(1)/N_l(1));

for i = 2:dt:t_size
    S_v(i) = S_v(i-1) + ( -beta_v*I_v(i-1)*S_v(i-1)/N_v(i-1) - lambda_I(i-1)*S_v(i-1)
) *dt;
    I_v(i) = I_v(i-1) + ( beta_v*I_v(i-1)*S_v(i-1)/N_v(i-1) + lambda_I(i-1)*S_v(i-1) - gamma_v*I_v(i-1)
    R_v(i) = R_v(i-1) + ( gamma_v*I_v(i-1)/N_v(i-1) + lambda_R(i-1)*I_v(i-1)
) *dt;

    S_l(i) = S_l(i-1) + ( -beta_l * I_l(i-1)*S_l(i-1)/N_l(i-1)
) *dt;
    I_l(i) = I_l(i-1) + ( beta_l * I_l(i-1)*S_l(i-1)/N_l(i-1) - gamma_l*I_l(i-1)/N_l(i-1)
) *dt;
    R_l(i) = R_l(i-1) + ( gamma_l*I_l(i-1)/N_l(i-1)
) *dt;

    N_l(i) = S_l(i) + I_l(i-1) + R_l(i-1);
    N_v(i) = S_v(i) + I_v(i-1) + R_v(i-1);
    lambda_I(i) = (sigma_I*(I_l(i)/N_l(i)));
    lambda_R(i) = (sigma_R*(R_l(i)/N_l(i)));
end
S = S_v + S_l;
I = I_v + I_l;
R = R_v + R_l;
figure
plot(1:t_size, S_v, 1:t_size, I_v, 1:t_size, R_v)
figure
plot(1:t_size, S_l, 1:t_size, I_l, 1:t_size, R_l)
figure
plot(1:t_size, I)

```

## B.4 Matlabcode bij het Zikavirus-model

```

clc;
clear all;

%parameters
mu_h = 1/(70*365); % death rate human
gamma_h = 1/6; % recovery rate human infected—>resistant
psi_h = mu_h; % birthrate humans
sigma_h = 19; % max # mosquito bites humans can sustain
v_h = 1/5; % transition speed exposed—>infected human

```

```

beta_hv = 0.33;           % transmission probability human→mosquito
beta_vh = 0.33;           % transmission probability mosquito→human
psi_v = 0.3;              % mosquito birthrate natural state
sigma_v = 0.5;            % biterate mosquito to human
v_v = 1/10;               % transition speed exposed→infected mosquito
mu_v = 1/14;              % death rate mosquito

H_0 = 100000;             % number of humans;
K_v = 2*H_0;              % carrying capacity mosquito

r_v = psi_v - mu_v;       % growrate mosquito

dt = 1;                   % timestep size
t_size = 200;             % time frame data

%initialisatie S_h, E_h, I_h, R_h
S_h = zeros(t_size,1);
E_h = zeros(t_size,1);
I_h = zeros(t_size,1);
R_h = zeros(t_size,1);

%initialisatie S_v, E_v, I_v
S_v = zeros(t_size,1);
E_v = zeros(t_size,1);
I_v = zeros(t_size,1);

%initialisatie hulpwaarden
N_h = zeros(t_size,1);
N_v = zeros(t_size,1);
H_v = zeros(t_size,1);
lambda_h = zeros(t_size,1);
lambda_v = zeros(t_size,1);

%start waardes mensen
S_h(1) = H_0-1;
E_h(1) = 0;
I_h(1) = 1;
R_h(1) = 0;

%start waardes muggen
S_v(1) = K_v;
E_v(1) = 0;
I_v(1) = 0;

N_h(1) = S_h(1)+E_h(1)+I_h(1)+R_h(1);
N_v(1) = S_v(1)+E_v(1)+I_v(1);
H_v(1) = psi_v - (r_v/K_v)*N_v(1);
lambda_h(1) = (sigma_v*sigma_h*N_v(1))/(sigma_h*N_h(1)+sigma_v+N_v(1))*beta_hv*(I_v(1)/N_v(1));
lambda_v(1) = (sigma_h*sigma_v*N_h(1))/(sigma_v*N_v(1)+sigma_h+N_h(1))*beta_vh*(I_h(1)/N_h(1));

for i = 2:dt:t_size
    S_h(i) = S_h(i-1) + ( psi_h*H_0 - lambda_h(i-1)*S_h(i-1)-mu_h*S_h(i-1) ) *dt;
    E_h(i) = E_h(i-1) + ( lambda_h(i-1)*S_h(i-1) - v_h*E_h(i-1) -mu_h*E_h(i-1) ) *dt;
    I_h(i) = I_h(i-1) + ( v_h*E_h(i-1) - gamma_h*I_h(i-1) -mu_h*I_h(i-1) ) *dt;
    R_h(i) = R_h(i-1) + ( gamma_h*I_h(i-1) -mu_h*R_h(i-1) ) *dt;

    S_v(i) = S_v(i-1) + ( H_v(i-1)*N_v(i-1) - lambda_v(i-1)*S_v(i-1)-mu_v*S_v(i-1) ) *dt;

```

```

E_v(i) = E_v(i-1) + ( lambda_v(i-1)*S_v(i-1) - v_v*E_v(i-1) - mu_v*E_v(i-1) ) *dt;
I_v(i) = I_v(i-1) + ( v_v*E_v(i-1) - mu_v*I_v(i-1) ) *dt;

N_h(i) = S_h(i-1) +E_h(i-1) +I_h(i-1) +R_h(i-1);
N_v(i) = S_v(i-1) +E_v(i-1) +I_v(i-1);
H_v(i) = psi_v - (r_v/K_v)*N_v(i);
lambda_h(i) = (sigma_v*sigma_h*N_v(i)) / (sigma_h*N_h(i)+sigma_v+N_v(i))*beta_hv*(I_v(i)/N_v(i));
lambda_v(i) = (sigma_h*sigma_v*N_h(i)) / (sigma_v*N_v(i)+sigma_h+N_h(i))*beta_vh*(I_h(i)/N_h(i));
end

figure
subplot(1,2,1)
plot(1:t_size,S_h,1:t_size,E_h,1:t_size,I_h,1:t_size,R_h)
axis square
title('Mensen')
legend('Susceptible','Exposed','Infected','Recovered','Location','northeast')
xlabel('Tijd')
ylabel('Aantal mensen')
subplot(1,2,2)
plot(1:t_size,S_v,1:t_size,E_v,1:t_size,I_v)
axis square
title('Muggen')
legend('Susceptible','Exposed','Infected','Location','northeast')
xlabel('Tijd')
ylabel('Aantal muggen')

figure
subplot(1,2,1)
plot(1:t_size,R_h)
axis square
title('Recovered hosts')
xlabel('Tijd')
ylabel('Aantal mensen')
subplot(1,2,2)
plot(1:t_size,E_h,1:t_size,I_h)
axis square
title('Exposed/Infected')
legend('Exposed','Infected','Location','northeast')
xlabel('Tijd')
ylabel('Aantal mensen')

```

## B.5 Matlabcode bij vergelijking analytische benadering en numerieke benadering

```

clear all;
clc;
nu = 0.0277;
beta = 0.0592;
phi = sqrt(nu^2 + beta^2);
t_size = 15*365;
t = [1:t_size];
S = zeros(t_size,1);
I = zeros(t_size,1);
R = zeros(t_size,1);
S(1) = 91.8882;
I(1) = 0.1121;
R(1) = 0.5613;

```

```

s(1) = log(S(1));
i(1) = log(I(1));
r(1) = log(R(1));
N = S(1) + I(1) + R(1);
S_anaytic = exp(0.1e1 / phi ^ 2 * (nu ^ 2 + beta ^ 2 * cos(t * phi / N)) - 0.1e1 - (phi ^ 2) ^
I_analytic = exp((phi ^ 2) ^ (-0.1e1 / 0.2e1) * beta * sin(t * phi / N) + cos(t * phi / N) - 0.
R_analytic = exp(0.1e1 / phi ^ 2 * beta * nu * (0.1e1 - cos(t * phi / N)) + (phi ^ 2) ^ (-0.1e1

for i=2:1:t_size
    S(i) = S(i-1) - beta*S(i-1)*I(i-1)/N;
    I(i) = I(i-1) + beta*S(i-1)*I(i-1)/N - nu*I(i-1)*R(i-1)/N;
    R(i) = R(i-1) + nu*I(i-1)*R(i-1)/N;

end
figure
subplot(1,2,1)
plot(t,S_anaytic,t,I_analytic,t,R_analytic)
axis square
title('Analytische benadering')
xlabel('Tijd in weken vanaf 2004')
ylabel('Genormeerde wekelijkse zoekdata')
subplot(1,2,2)
plot(1:2*365,S(1:2*365),1:2*365,I(1:2*365),1:2*365,R(1:2*365))
axis square
title('Numerieke benadering')
xlabel('Tijd in weken vanaf 2004')
ylabel('Genormeerde wekelijkse zoekdata')

```

## B.6 Matlabcode voor berekening SSE van SIR-model

```

function [ SSE ] = SSE1_SIR(x)

original = xlsread('FacebookWorldJuni.xlsx', 'E2:E667');
dt =1;
t_size = size(original,1);
beta = x(1);
gamma = x(2);
s0 = x(3);
i0 = x(4);
r0 = x(5);
% Initialiseren %
S = zeros(t_size,1);
I = zeros(t_size,1);
R = zeros(t_size,1);
% Beginwaarden %
S(1) = s0;
I(1) = i0;
R(1) = r0;
N = s0+i0+r0;
for i=2:dt:t_size
    S(i) = S(i-1) - (beta*I(i-1)*S(i-1))/N;
    I(i) = I(i-1) + (beta*I(i-1)*S(i-1))/N - gamma*I(i-1)/N;
    R(i) = R(i-1) + gamma*I(i-1)/N;
end
M = max(original);
normI = (I/M)*100;
SE = (normI-original).^2;
SSE = sum(SE);

```

```
end
```

## B.7 Matlabcode voor berekening SSE van irSIRsbd-model

```
function [ SSE ] = SSEIirSIRsdb( x )

original = xlsread('FacebookWorldJuni.xlsx', 'E2:E667');
t_size = size(original,1);
alpha = x(1);
beta = x(2);
nu = x(3);
s0 = x(4);
i0 = x(5);
r0 = x(6);
b = x(7);
d = x(8);
% Initialiseren %
S = zeros(t_size,1);
I = zeros(t_size,1);
R = zeros(t_size,1);
N = zeros(t_size,1);
% Beginwaarden %
N(1) = s0+i0+r0;
S(1) = s0;
I(1) = i0;
R(1) = r0;
for i=2:t_size
S(i) = S(i-1) + (-beta*S(i-1)*I(i-1) + alpha*R(i-1) - d*S(i-1))/N(i-1) +b;
I(i) = I(i-1) + (beta*S(i-1)*I(i-1) - nu*I(i-1)*R(i-1) - d*I(i-1))/N(i-1);
R(i) = R(i-1) + (nu*I(i-1)*R(i-1) - alpha*R(i-1) - d*R(i-1))/N(i-1);
N(i) = S(i) + I(i) + R(i);
end
m = max(original);
normI = I/m*100;
SE = (normI-original).^2;
SSE = sum(SE);
end
```

## B.8 Matlabcode voor berekening SSE van leiders-volgersmodel

```
function [ SSE ] = SSEleidervolg ( x )

original = xlsread('FacebookWorldJuni.xlsx', 'E2:E667');
t_size = size(original,1);
dt = 1;

gamma_v = x(1);
gamma_l = x(2);
beta_v = x(3);
beta_l = x(4);
sigma_I = x(5);
sigma_R = x(6);
n = x(7);
n_v = x(8);
s_l_0 = 0.16*(n-n_v);
i_l_0 = n_v*0.16;
r_l_0 = x(9);
```

```

s_v_0 = 0.84*n;
i_v_0 = 0;
r_v_0 = x(10);

%initialisatie S_h, I_h, R_h
S_l = zeros(t_size,1);
I_l = zeros(t_size,1);
R_l = zeros(t_size,1);

%initialisatie S_v, E_v, I_v
S_v = zeros(t_size,1);
I_v = zeros(t_size,1);
R_v = zeros(t_size,1);

%initialisatie hulpwaarden
N_l = zeros(t_size,1);
N_v = zeros(t_size,1);
lambda_I= zeros(t_size,1);
lambda_R= zeros(t_size,1);

%start waarden populaties
S_l(1) = s_l_0;
I_l(1)= i_l_0;
R_l(1)= r_l_0;

S_v(1) = s_v_0;
I_v(1)= i_v_0;
R_v(1)= r_v_0;

%start waarden totale populaties
N_l(1) = s_l_0+i_l_0+r_l_0;
N_v(1) = s_v_0+i_v_0+r_v_0;

%start waarden lambdas
lambda_I(1) = sigma_I*(I_l(1)/N_l(1));
lambda_R(1) = sigma_R*(R_l(1)/N_l(1));

for i = 2:dt:t_size
    S_v(i) = S_v(i-1) + ( -beta_v*I_v(i-1)*S_v(i-1)/N_v(i-1) - lambda_I(i-1)*S_v(i-1)
) *dt;
    I_v(i) = I_v(i-1) + ( beta_v*I_v(i-1)*S_v(i-1)/N_v(i-1) + lambda_I(i-1)*S_v(i-1) - gamma_v*I_v(i-1)
    R_v(i) = R_v(i-1) + ( gamma_v*I_v(i-1)/N_v(i-1) + lambda_R(i-1)*I_v(i-1)
) *dt;

    S_l(i) = S_l(i-1) + ( -beta_l * I_l(i-1)*S_l(i-1)/N_l(i-1)
) *dt;
    I_l(i) = I_l(i-1) + ( beta_l * I_l(i-1)*S_l(i-1)/N_l(i-1) - gamma_l*I_l(i-1)/N_l(i-1)
) *dt;
    R_l(i) = R_l(i-1) + ( gamma_l*I_l(i-1)/N_l(i-1)
) *dt;

    N_l(i) = S_l(i) +I_l(i-1) +R_l(i-1);
    N_v(i) = S_v(i) +I_v(i-1) +R_v(i-1);
    lambda_I(i) = (sigma_I*(I_l(i)/N_l(i)));
    lambda_R(i) = (sigma_R*(R_l(i)/N_l(i)));
end

S = S_v + S_l;
I = I_v + I_l;

```

```
R = R_v + R_l;
```

```
m = max(original);
normI = I/m*100;
SE = (normI-original).^2;
SSE = sum(SE);
end
```

```
function [ SSE ] = SSE_leidervolg_metplot ( x )
% opt = [0.044329576642123 0.872149548845434 0.090546963019920 0.094315283723067 4.244972568047686 0
original = xlsread('FacebookWorldJuni.xlsx', 'E2:E667');
t_size = size(original,1);
dt = 1;

gamma_v = x(1);
gamma_l = x(2);
beta_v = x(3);
beta_l = x(4);
sigma_I = x(5);
sigma_R = x(6);
n = x(7);
n_v = x(8);
s_l_0 = 0.16*(n-n_v);
i_l_0 = n_v*0.16;
r_l_0 = x(9);
s_v_0 = 0.84*n;
i_v_0 = 0;
r_v_0 = x(10);

%initialisatie S_h, I_h, R_h
S_l = zeros(t_size,1);
I_l = zeros(t_size,1);
R_l = zeros(t_size,1);

%initialisatie S_v, E_v, I_v
S_v = zeros(t_size,1);
I_v = zeros(t_size,1);
R_v = zeros(t_size,1);

%initialisatie hulpwaarden
N_l = zeros(t_size,1);
N_v = zeros(t_size,1);
lambda_I= zeros(t_size,1);
lambda_R= zeros(t_size,1);

%start waarden populaties
S_l(1) = s_l_0;
I_l(1) = i_l_0;
R_l(1) = r_l_0;

S_v(1) = s_v_0;
I_v(1) = i_v_0;
R_v(1) = r_v_0;

%start waarden totale populaties
N_l(1) = s_l_0+i_l_0+r_l_0;
N_v(1) = s_v_0+i_v_0+r_v_0;
```



```

%start waarden lambdas
lambda_I(1) = sigma_I*(I_l(1)/N_l(1));
lambda_R(1) = sigma_R*(R_l(1)/N_l(1));

for i = 2:dt:t_size
    S_v(i) = S_v(i-1) + ( -beta_v*I_v(i-1)*S_v(i-1)/N_v(i-1) - lambda_I(i-1)*S_v(i-1)
) *dt;
    I_v(i) = I_v(i-1) + ( beta_v*I_v(i-1)*S_v(i-1)/N_v(i-1) + lambda_I(i-1)*S_v(i-1) - gamma_v*
    R_v(i) = R_v(i-1) + ( gamma_v*I_v(i-1)/N_v(i-1) + lambda_R(i-1)*I_v(i-1)
) *dt;

    S_l(i) = S_l(i-1) + ( -beta_l * I_l(i-1)*S_l(i-1)/N_l(i-1)
) *dt;
    I_l(i) = I_l(i-1) + ( beta_l * I_l(i-1)*S_l(i-1)/N_l(i-1) - gamma_l*I_l(i-1)/N_l(i-1)
) *dt;
    R_l(i) = R_l(i-1) + ( gamma_l*I_l(i-1)/N_l(i-1)
) *dt;

    N_l(i) = S_l(i) + I_l(i-1) + R_l(i-1);
    N_v(i) = S_v(i) + I_v(i-1) + R_v(i-1);
    lambda_I(i) = (sigma_I*(I_l(i)/N_l(i)));
    lambda_R(i) = (sigma_R*(R_l(i)/N_l(i)));
end

S = S_v + S_l;
I = I_v + I_l;
R = R_v + R_l;

%parameters
beta_SIR = 0.0292195733641;
gamma_SIR = 1.1378327523449;
dt = 1;

%initialisatie S, I, R
S_SIR = zeros(t_size,1);
I_SIR = zeros(t_size,1);
R_SIR = zeros(t_size,1);

S_SIR(1)=199.4796356526135;
I_SIR(1)= 0.0841490763223;
R_SIR(1)=0;
N_SIR = S_SIR(1) + I_SIR(1) + R_SIR(1);

for i = 2:dt:t_size
    S_SIR(i) = S_SIR(i-1) - (beta_SIR*I_SIR(i-1)*S_SIR(i-1))*dt/N_SIR;
    I_SIR(i) = I_SIR(i-1) + (beta_SIR*I_SIR(i-1)*S_SIR(i-1))*dt/N_SIR - gamma_SIR*I_SIR(i-1)*dt
    R_SIR(i) = R_SIR(i-1) + gamma_SIR*I_SIR(i-1)*dt/N_SIR;
end

week = transpose(1:size(original));
jaar = 2004+week/52;

figure
plot(jaar,I,jaar,original,jaar,I_SIR)
legend('Leider-volgersmodel','Originele data','SIR-model','Location','northwest')
xlabel('Tijd in jaren')
ylabel('Genormeerde zoekdata')

figure
plot(jaar,I,jaar,original)

```

```

legend('Leider-volgersmodel','originele data','Location','northwest')
xlabel('Tijd in jaren')
ylabel('Genormeerde zoekdata')

figure
plot(jaar,I_v,jaar,I_l,jaar,S_v,jaar,S_l,jaar,R_v,jaar,R_l)
legend('Geïnfecteerde volgers','Geïnfecteerde leiders','Vatbare volgers','Vatbare leiders','Resisten
xlabel('Tijd in jaren')
ylabel('Genormeerde zoekdata')

figure
plot(jaar,I_v,jaar,I_l)
legend('Volgers','Leiders','Location','northwest')
xlabel('Tijd in jaren')
ylabel('Genormeerde zoekdata')

figure
plot(jaar,I_l,jaar,S_l,jaar,R_l)
legend('Geïnfecteerde leiders','Vatbare leiders','Resistente leiders','Location','northwest')
xlabel('Tijd in jaren')
ylabel('Genormeerde zoekdata')

m = max(original);
normI = I/m*100;
SE = (normI-original).^2;
SSE = sum(SE);

figure
I_v_norm = I_v/max(I_v);
I_l_norm = I_l/max(I_l);
plot(jaar,I_l_norm,jaar,I_v_norm)
legend('Geïnfecteerde leiders','Geïnfecteerde volgers','Location','northwest')
xlabel('Tijd in jaren')
ylabel('Relatieve aantallen geïnfecteerden')

end

```

## B.9 Minimalisatiecode voor SSE van SIR-model

```

clc; clear all;
format long
dt= 1;

beta = 0.0292195733641;
gamma = 1.1378327523449;
s0 = 199.4796356526135;
i0 = 0.0841490763223;
r0 = 0;
SSE_SIR([beta, gamma, s0, i0, r0])
options = optimset('Display','iter','MaxIter', 2000, 'TolFun', 10^-6, 'TolX', 10^-6);
opt = fminsearch(@SSE_SIR, [beta, gamma, s0, i0, r0], options)
SSE_SIR(opt)

original = xlsread('FacebookWorldJuni.xlsx', 'E2:E667');
t_size = size(original,1);
%initialisatie S, I, R
S = zeros(t_size,1);
I = zeros(t_size,1);

```

```

R = zeros(t_size,1);
N=s0+i0+r0;
beta = opt(1);
gamma = opt(2);
S(1)=opt(3);
I(1)=opt(4);
R(1)=opt(5);

for i = 2:dt:t_size
    S(i) = S(i-1) - (beta*I(i-1)*S(i-1))*dt/N;
    I(i) = I(i-1) + (beta*I(i-1)*S(i-1))*dt/N - gamma*I(i-1)*dt/N;
    R(i) = R(i-1) + gamma*I(i-1)*dt/N;
end

```

```

week = transpose(1:size(original));
jaar = 2004+week/52;
SSE = (I-original).^2;
sum(SSE)
plot(jaar, I, 'r')
hold on
plot(jaar, original, 'b')
legend('Benadering', 'Originele data', 'Location', 'northwest')
title('Facebook met het SIR-model');
xlabel('Tijd in jaren');
ylabel('Genormeerde wekelijkse zoekdata');
hold off

```

## B.10 Minimalisatiecode voor SSE van irSIRSD-model

```

clc; clear all;
format long
original = xlsread('FacebookWorldJuni.xlsx', 'E2:E667');
t_size = size(original,1);
alpha = 1.239244494510886;
beta = 0.0407002600712227;
nu = 0.045411746122411;
s0 = 10.256377170091664;
i0 = 0.089634046479910;
r0 = 57.557789765325481;
b = 0.546388510009792;
d = 0.435866288773971;
%% Optimaliseren van parameters
SSE_irSIRSDb([alpha, beta, nu, s0, i0, r0, b, d])
options = optimset('Display','iter', 'MaxIter', 600, 'TolFun',10^-6, 'TolX', 10^-6, 'MaxFunEval', 10000);
opt = fminsearch(@SSE_irSIRSDb, [alpha, beta, nu, s0, i0, r0, b, d],options)
SSE5_irSIRSDb(opt)

alpha = opt(1);
beta = opt(2);
nu = opt(3);
s0 = opt(4);
i0 = opt(5);
r0 = opt(6);
b = opt(7);
d = opt(8);

```

```

% Initialiseren %
S = zeros(t_size,1);
I = zeros(t_size,1);
R = zeros(t_size,1);
N = zeros(t_size,1);
% Beginwaarden %
N(1) = s0+i0+r0;
S(1) = s0;
I(1) = i0;
R(1) = r0;

for i=2:t_size
S(i) = S(i-1) + (-beta*S(i-1)*I(i-1) + alpha*R(i-1) - d*S(i-1))/N(i-1) +b;
I(i) = I(i-1) + (beta*S(i-1)*I(i-1) - nu*I(i-1)*R(i-1) - d*I(i-1))/N(i-1);
R(i) = R(i-1) + (nu*I(i-1)*R(i-1) - alpha*R(i-1) - d*R(i-1))/N(i-1);
N(i) = S(i) + I(i) + R(i);
end

week = transpose(1:size(original));
jaar = 2004+week/52;
SSE = (I-original).^2;
sum(SSE)
plot(jaar, I, 'r')
hold on
plot(jaar, original, 'b')
legend('Benadering', 'Originele data', 'Location', 'northwest')
title('Facebook met het irSIRS-model met geboorte- en sterftecijfer');
xlabel('Tijd in jaren');
ylabel('Genormeerde wekelijkse zoekdata');
hold off

```

## B.11 Minimalisatiecode voor SSE van leiders-volgersmodel

```

clc;
clear all;
original = xlsread('FacebookWorldJuni.xlsx', 'E2:E667');
t_size = 666;

gamma_v =0.044329576642123;
gamma_l =0.872149548845434 ;

beta_v = 0.090546963019920;
beta_l = 0.094315283723067;

sigma_I = 4.244972568047686;
sigma_R = 0.015326573347482;

n = 80.994104287546040;
n_v =1.104302509381040e-05;
r_l_0 = -2.130130486141346;
r_v_0 = 1.524628910365468e+02;

SSE_leidervolg([gamma_v, gamma_l, beta_v, beta_l, sigma_I, sigma_R, n, n_v, r_l_0, r_v_0])
options = optimset('Display','iter', 'MaxIter', 200, 'TolFun', 10^-6, 'TolX', 10^-6, 'MaxFunEvals', 50)
opt = fminsearch(@SSE_leidervolg, [gamma_v, gamma_l, beta_v, beta_l, sigma_I, sigma_R, n, n_v, r_l_0, r_v_0],

```

SSE\_leidervolg (opt)

SSE\_leidervolg\_metplot (opt)



## Bijlage C

# Parameters

### C.1 Parameters bij voor facebook een jaar later

$\alpha$	1.239244494510886
$\beta$	0.040700260071227
$\nu$	0.045411746122411
$s_0$	10.256377170091664
$i_0$	0.089634046479910
$r_0$	57.557789765325481
$b$	0.546388510009792
$d$	0.435866288773971

### C.2 Geoptimaliseerde parameterwaarden SIR-model

$\beta$	0.0291800405161
$\gamma$	0.014225843490041
$s_0$	219.5671688775670
$i_0$	0.1034987587665
$r_0$	-0.0018768416290

Gevonden sum of squared errors:  $8.466840518500198 \cdot 10^3$ .

### C.3 Geoptimaliseerde parameterwaarden irSIRsdb-model

$\alpha$	1.175221171419798
$\beta$	0.040686811468789
$\nu$	0.043737072036003
$s_0$	3.756814693185675
$i_0$	0.131143580059496
$r_0$	63.553471995040468
$b$	0.560177957054893
$d$	0.445595190361699

Gevonden sum of squared error: 2888.126444





# Bibliografie

- [1]
- [2] Kimitaka Asatani. Rise and decline of process of online communities: Modeling social balance of participants. [http://ddd.uab.cat/pub/poncom/2014/128040/ssc14\\_a2014a35iENG.pdf](http://ddd.uab.cat/pub/poncom/2014/128040/ssc14_a2014a35iENG.pdf).
- [3] Dorwin en Frank Haraway. Cartwright. Structural balance: a generalization of heidler's theory. *Psychological Review*, page 277, 1956.
- [4] Simrat Singh Chhabra, Ajit Brundavanam, and Saswata Shannigrahi. An alternative explanation for the rise and fall of myspace. *CoRR*, abs/1403.5617, 2014.
- [5] Fabian Hadiji Christian Bauckhage, Kristian Kersting. How viral are viral videos? <http://www.aaai.org/ocs/index.php/ICWSM/ICWSM15/paper/view/10505/10485>.
- [6] Carrie A. Manore et al. Comparing dengue and chikungunya emergence and endemic transmission in a. aegypti and a. albopictus. <http://www.ncbi.nlm.nih.gov/pubmed/24801860>.
- [7] Carrie A. Manore et al. Mathematical models for fighting zika virus. *Siam News*.
- [8] J. Ginsberg et al. Detecting influenza epidemics using search engine query data. *Nature*.
- [9] Kazuyuki Fujii. Comment on 'epidemiological modeling of online social network dynamics'. <https://arxiv.org/pdf/1402.1225v2.pdf>.
- [10] Sandra Maring. Epidemologische modellen voor de groei en afname van sociale netwerken. [http://repository.tudelft.nl/assets/uuid:fa20871c-ea0e-4a56-acdd-e5de40e246b7/BEP\\_SMaring.pdf](http://repository.tudelft.nl/assets/uuid:fa20871c-ea0e-4a56-acdd-e5de40e246b7/BEP_SMaring.pdf).
- [11] J-P Onnela, Jari Saramäki, Jorkki Hyvönen, György Szabó, David Lazer, Kimmo Kaski, János Kertész, and A-L Barabási. Structure and tie strengths in mobile communication networks. *Proceedings of the National Academy of Sciences*, 104(18):7332, 2007.
- [12] E.M. Rogers. *Diffusion of innovations*. Free Press of Glencoe, 1962.