



THE LORD OF THE RINGS

De impact van sentiment op netwerkvorming

In dit artikel willen we laten zien dat er verband is tussen de dynamiek van een sociaal netwerk en de sentimentscore van de actoren in dat netwerk. Hiervoor gebruiken we de gegevens uit het filmscript van *The Lord of the Rings: The Fellowship of the Ring* die we vergelijken met wat er bekend is over het verhaal. De sentimentscore van het filmscript wordt verkregen met behulp van *machine learning*. Deze sentimentscores dienen vervolgens als input voor een sociale netwerkanalyse met behulp van het stochastische actor-gebaseerde model voor netwerkdynamiek, genaamd SIENA. Dit model schat de waarden van de parameters door een iteratief stochastisch simulatiealgoritme toe te passen. De uitkomst laat zien dat het effect van sentiment positief en significant is.

JADÉ DIETEREN & DEMI DE KORT

De wereld bestaat uit talloze complexe netwerksystemen die een belangrijk doel dienen in onze samenleving. Aan de hand van het sociale netwerk uit de film *The Lord of the Rings: The Fellowship of the Ring* tonen we aan dat er een verband is tussen de dynamiek van een sociaal netwerk en de sentimentscore van de actoren in dat netwerk.^{1,2} We maken gebruik van een stochastisch actor-gebaseerd model voor netwerkdynamiek, genaamd SIENA. Dit model maakt de statistische evaluatie van de netwerkeffecten mogelijk. Uit het filmscript kan ook het sentiment van de personages worden gehaald door gebruik te maken van *machine learning*-technieken. We onderzoeken vervolgens of de sentimentscore van een actor in een sociaal netwerk de dynamiek van de netwerkstructuur significant beïnvloedt.

Hoewel de dynamiek van sociale netwerken en sentimentanalyse zeer inzichtelijk kan zijn, zijn er nog steeds uitdagingen. Menselijke taal kan bijvoorbeeld moeilijk te begrijpen en zeer contextueel zijn. Het unieke aan het gebruik van een filmscript is dat elke beweging van elk personage bekend is. *The Lord of the Rings* is daarom een ideale testcasus voor netwerkanalyse, omdat we de mogelijkheid hebben om de modelresultaten op basis van data te vergelijken met de kennis die we hebben over het verhaal.

Sentimentscore

Sentimentanalyse is een toepassing van Natural Language Processing (NLP) en wordt gedefinieerd als de inter-

pretatie van teksten om de emotionele toon achter een reeks woorden te vinden. We vergeleken de prestaties van twee *supervised learning*-modellen en één *unsupervised learning*-model dat gebruikt maakt van een lexicon om de scènes te analyseren.

Het eerste model maakt gebruik van gegevens die op Twitter zijn verzameld, het tweede model gebruikt gegevens uit filmrecensies om zichzelf te trainen. Het laatste en best presterende model op basis van handmatige klassificering, is het AFINN Lexicon Based Model. De huidige versie van het lexicon bevat 2.477 woorden die elk een polariteitsscore tussen -5 en 5 hebben. Hoe hoger de score, hoe meer positiviteit aan dat woord wordt geassocieerd. De cumulatieve score van alle woorden is de score voor een zin of voor een hele scène. De output van dit model is een numerieke score. Hierdoor kan onderscheid worden gemaakt tussen zeer positieve scènes en licht positieve scènes.

Simulatie voor sociale netwerken

Een sociaal netwerk is een netwerk van sociale interacties en persoonlijke relaties. Stel dat er een sociaal netwerk is dat bestaat uit n actoren. Dit netwerk kan dan worden weergegeven door een n bij n matrix $X = (x_{ij})$ waarbij $x_{ij} = 1$ als er een verbinding tussen actor i en j aanwezig is en anders $x_{ij} = 0$.

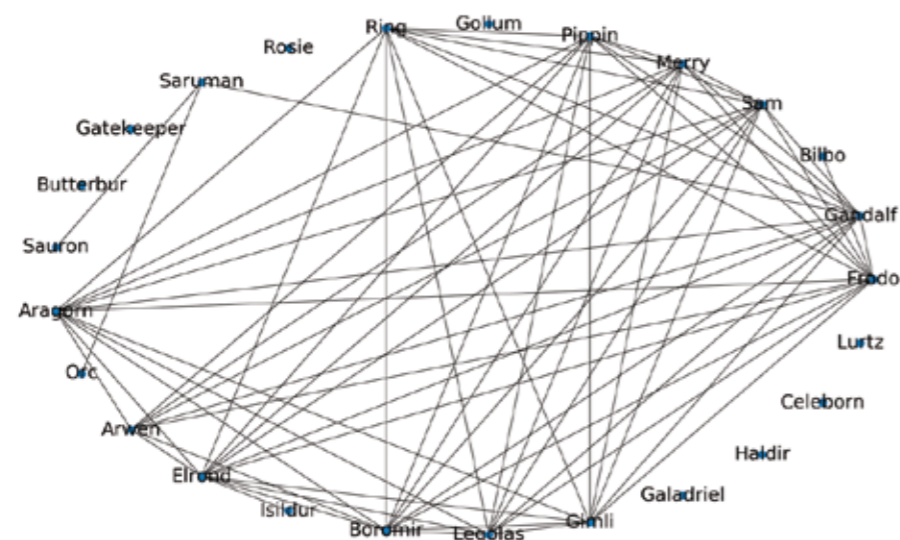
Network Dynamics is een onderzoeksgebied dat zich richt op netwerken die zich in de loop van de tijd ontwikkelen. Netwerkevolutie wordt gedreven door bepaal-

de netwerkeffecten. Ripley et al. (2011) geeft een gedetailleerde uitleg over hoe verschillende effecten worden gedefinieerd. Enkele prominente effecten zijn:

- Transitiviteit ontstaat als er een band bestaat tussen persoon A en persoon B, en weer een band tussen persoon B en persoon C, en dan ontstaat een band tussen persoon A en persoon C;
- Homofilie is de neiging om zich te verhouden tot mensen met vergelijkbare kenmerken;
- *Outdegree* is het aantal uitgaande banden dat een persoon heeft.

Voor onze analyse werken we met SIENA (Simulation Investigation for Empirical Network Analysis), een programma voor de statistische analyse van sociale netwerkgegevens. Het doel van het model is om de netwerkdynamiek weer te geven op basis van waargenomen longitudinale gegevens, en deze te evalueren volgens het paradigma van statistische inferentie (Snijders et al., 2010). Het model gaat ervan uit dat de onderliggende tijdparameter continu is en dat het veranderende netwerk het resultaat is van een Markov-proces. Daarnaast wordt aangenomen dat de actoren hun uitgaande verbindingen beheersen en dat op een gegeven moment één actor de kans krijgt om een uitgaande verbinding te veranderen.

Homofilie wordt gedreven door covariaten, die onafhankelijke variabelen in het model zijn. We zullen het sentiment van de actoren koppelen aan de sociale netwerkanalyse via een in de tijd variërende covariaat.



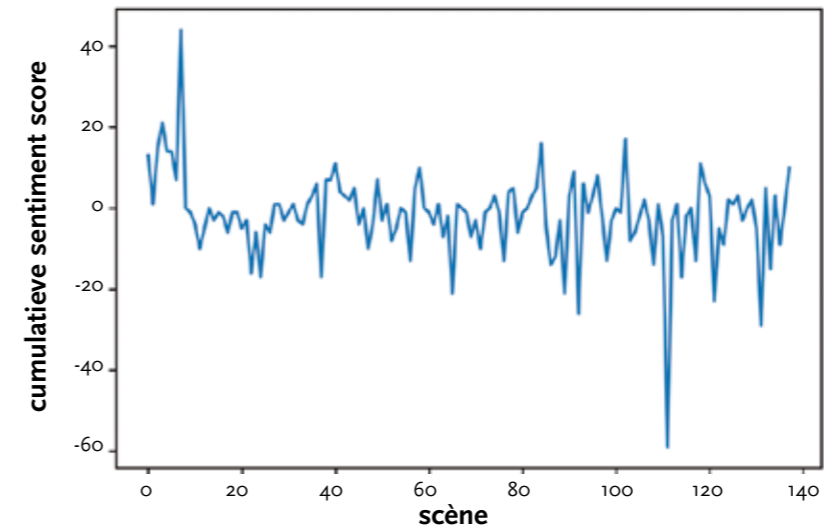
Figuur 1. Grafiek van een scènegroep; het Netwerk bevat de belangrijkste personages van het verhaal, die met elkaar verbonden zijn als ze samen in een scène voorkomen

De actor i die de kans krijgt om een verbinding te veranderen in netwerk x , heeft de mogelijkheid om niets te doen of één uitgaande connectie te veranderen. De kans voor alle mogelijkheden hangt af van de *objective function* (Snijders et al, 2010), gedefinieerd als:

$$\frac{\exp(f_i(\beta, x))}{\sum_{x' \in C} \exp(f_i(\beta, x'))}$$

waarbij $f_i(\beta, x) = \sum_k \beta_k S_{ik}(x)$. $S_{ik}(x)$ zijn de bovengenoemde effecten en de statistische parameters. Om de parameterwaarden β_k voor elk van de k effecten te vinden, wordt een iteratief stochastisch simulatie algoritme toegepast. Een groot aantal simulaties wordt herhaald om de netwerkevolutie na te bootsen tussen twee opeenvolgende observatiemomenten van het netwerk. De schattingsprocedure die in SIENA wordt gebruikt is de Method of Moments (MoM).

Het algoritme kent drie fasen waarin verschillende stappen worden doorlopen om de hyperparameters van het model te bepalen. In de eerste fase wordt een eerste ruwe schatting van de matrix van derivaten gemaakt (Ripley et al., 2011). Het doel van de tweede fase is om de parameterwaarden te vinden waarvoor de afwijkingen tussen gegenereerde en waargenomen waarden gemiddeld nul zijn. Dit resulteert in een uiteindelijke waarde voor de parametervector β_k . De derde fase produceert schattingen voor de covariantiematrix en de matrix van derivaten. Deze kunnen worden gebruikt om de standaardfouten van de parameterwaarden te berekenen.



Figuur 2. Sentimentscore van het AFINN lexicon model

Case study: Lord of the Rings

Om de netwerkanalyse uit te voeren, zijn de scènes in het filmscript van *The Lord of the Rings: The Fellowship of the Ring* verdeeld in vijf groepen (zie figuur 1 voor scènegroep 3). De belangrijkste locaties van de vijf sociale netwerken zijn de Shire, Bree, Rivendell, Moria en Lothlórien. In deze netwerken werd aangenomen dat personages als ze samen in een scène voorkomen een connectie hebben.

De sentimentscore van het AFINN lexicon model voor elke scène is te zien in figuur 2. Scène 8 krijgt een score van 43 en is de enige scène met een sentimentscore boven de 20. In deze scène vindt Bilbo's verjaardagsfeest plaats. Het is een vrolijke gebeurtenis, waar vuurwerk wordt afgestoken en mensen dansen. De laagste senti-

mentscore wordt behaald in scène 112. In deze scène, die zich afspeelt in de Mines of Moria, stoot Pippin een scheidel om die naar beneden valt en veel lawaai maakt. Door dit te doen, weten de Orks waar de groep zich in de mijn bevindt en komen ze hen aanvalen. De scène is gevuld met veel angst, gevolgd door een felle en gewelddadige strijd. Gemiddeld is het sentiment van de film negatief, namelijk -1,73. Dit past bij de sombere sfeer van de film.

Het SIENA-model moest in totaal twee keer worden uitgevoerd voordat de gewenste convergentie werd verkregen. De parameterschattingen voor de effecten zijn te vinden in tabel 1.

De parameters in tabel 1 kunnen worden geïnterpreteerd met de evaluatiefunctie:

$$f_i = \sum_{j=1}^6 \beta_k s_{ik}(x) ,$$

		β	S.E.	t	p
1	Outdegree	-8,59	1,86	0,10	0,0000
2	Evenwicht	0,44	0,15	-0,01	0,0037
3	Transitiviteit	5,96	1,70	0,11	0,0005
4	4-cycles	-0,26	0,11	0,11	0,0147
5	Populariteit	0,59	0,22	0,11	0,0061
6	Sentimentovereenkomst	1,88	0,86	0,08	0,0292

Tabel 1. Selectie van geschatte waarden

waarbij k overeenkomt met de effecten. Dus via f_i hebben we een proxy van de verdeling over de mogelijke acties die actor i kan ondernemen. Een directe vergelijking van de waarden van parameterschattingen is mogelijk, maar houd er rekening mee dat dit niet-gestandaardiseerde coëfficiënten zijn (Ripley et al., 2011).

De parameterschatting heeft de grootste absolute waarde voor het *outdegree*-effect. De negatieve schatting kan worden verklaard door het feit dat het een inspanning kost om vriendschappen en connecties te onderhouden. Het effect is significant, zoals blijkt uit de p -waarde. Uit het resultaat blijkt ook dat er sprake is van een hoge mate van transitiviteit. Dit geeft aan dat het principe 'vrienden van mijn vrienden zijn mijn vrienden' in dit netwerk op zijn plaats is.

De parameter voor sentimentovereenkomst is positief en significant. Dit betekent dat actoren met sentiment scores die vergelijkbaar zijn, meer kans hebben om een verbinding met elkaar aan te gaan. Aangezien dit sociale netwerk is geconstrueerd op basis van personages die samen in een filmscène zitten, zouden ze bovendien de positieve of negatieve ervaring delen en daarom waarschijnlijk een vergelijkbare sentiment score hebben.

De parameterschattingen voor evenwicht en populariteit zijn laag, maar significant. Voor evenwicht betekent dit dat er een lichte voorkeur is voor de personages om een connectie te hebben met andere personages die veel van dezelfde connecties hebben als ze al hebben. De positieve schatting van de populariteit wijst op een lichte voorkeur om een connectie te maken met populaire personages. De reden hiervoor kan zijn dat de hoofdpersonen in de film het meest voorkomen in scènes en daarom als populair kunnen worden beschouwd. Omdat het verhaal om hen draait, creëren ze ook meer nieuwe banden. De schatting voor 4-cycli is negatief, wat wijst op een neiging om niet uit clusters van grootte vier te komen.

Conclusie

In dit artikel zijn gegevens uit de film *The Lord of the Rings: The Fellowship of the Ring* gebruikt om het sociale netwerk en sentiment scores van de personages te extraheren. Voor de extractie van het karaktersentiment werd

gebruik gemaakt van het AFINN Lexicon Based Model. Vervolgens werd een sociale netwerkanalyse uitgevoerd met behulp van SIENA. We vonden een positieve parameterwaarde voor sentiment die significant verschilt van nul.

Meestal kwamen onze bevindingen met behulp van de modellen overeen met wat we weten over het verhaal. Er zijn echter nog steeds uitdagingen bij het toepassen van sentimentanalyse en sociale-netwerkanalyse. Bij sociologisch empirisch onderzoek is het daarom noodzakelijk om de resultaten van de analyse te interpreteren en na te denken over wat de data-analyses je wel en niet kunnen laten zien. We weten bijvoorbeeld dat de film om de ring draait. Het is echter niet eenvoudig om hiermee rekening te houden bij het uitvoeren van de analyse.

Voor toekomstig onderzoek kan dezelfde analyse worden uitgevoerd met andere filmscripts uit verschillende genres. Bovendien zou dit onderzoek kunnen worden uitgevoerd met behulp van echt bestaande sociale netwerken. In het echte leven zal het echter moeilijk zijn om gegevens te verzamelen. Hoewel de filmgegevens erg waardevol zijn omdat er geen ontbrekende gegevens zijn en het mogelijk is om resultaten te valideren, weerspiegelt het niet helemaal een situatie in het echte leven.

NOOT

1. Dit artikel is gebaseerd op de masterscriptie van Demi de Kort voor haar studie Econometrie & Operations.
2. Het volledige filmscript is te vinden op <https://www.imsdb.com/scripts/Lord-of-the-Rings-Fellowship-of-the-Ring-The.html>.

LITERATUUR

- Snijders, T. A. B., van de Bunt, G. G., & Steglich, C. E. G. (2010). Introduction to stochastic actor-based models for network dynamics. *Social Networks*, 32, 44–60.
- Ripley, R. M., Snijders, T. A. B., Boda, Z., Vörös, A., & Preciado, P. (2020). *Manual for SIENA version 4.0*. University of Oxford, Department of Statistics; Nuffield College.

JADÉ DIETEREN is forecasting business analyst bij Amazon en PhD student aan de Vrije Universiteit.
E-mail: jjdd@amazon.cz

DEMI DE KORT is data scientist bij izi. Met hun analyses willen de data scientists van izi bijdragen aan een betaalbare en doelmatige zorg in Nederland.
E-mail: demidekort@hotmail.com



COMBINATORISCHE OPTIMALISATIE EN QUANTUM COMPUTING

De Quantum Computer komt met rasse schreden dichterbij. Eén van de veelbelovende toepassingsgebieden is optimalisatie. Voor Operations Research practitioners is het van belang deze ontwikkeling te volgen, aangezien ze binnen een aantal jaren quantum algoritmes aan hun gereedschapskist zullen kunnen toevoegen. In dit artikel bespreken we twee manieren waarop optimalisatieproblemen kunnen worden aangepakt met de Quantum Computer.

FRANK PHILLIPSON

Een nieuwe computer die in slechts enkele seconden het antwoord op vragen kan geven waar de huidige generatie computers vele jaren mee bezig zouden zijn. Dat is wat er beloofd wordt door de wetenschappers die, onder an-

deren in Delft, werken aan de Quantum Computer (QC). Deze QC maakt gebruik van kwantummechanische verschijnselen, zoals superpositie en verstrengeling, om bewerkingen uit te voeren.