



ILLUSTRATIE: JEROEN MURRÉ

#Data voor het oprapen

Honderden miljoen mensen delen op Twitter de meest uiteenlopende zaken met zoveel mogelijk mensen. De website lijkt een **onuitputtelijke bron** van bruikbare gegevens voor sociologisch onderzoek. Is Twitter werkelijk een goudmijn of juist een wetenschappelijk mijnenveld? **Door Erik Arends**

WAT IS BETER DAN ÉÉN deelnemer aan je onderzoek? Twee deelnemers natuurlijk. Of nog beter: 300 miljoen. Tot voor kort konden sociaal-wetenschappers alleen maar dromen van enorme aantallen deelnemers uit alle rangen en standen, die ook nog eens gratis hun gedachten delen. Met de groeiende populariteit van netwerksites als Twitter zien sociaal-wetenschappers kansen om op relatief eenvoudige manier representatieve meetgegevens te vergaren – en om af te rekenen met een paar hardnekkige problemen in hun vakgebied.

Vaak schakelen sociaal-wetenschappers studenten in als proefpersoon voor hun onderzoek. Een begrijpelijke keus: studenten zijn volop te vinden aan universiteiten, ze hebben genoeg tijd en ze zijn tevreden met een geringe onkostenvergoeding. Toch brengt die aanpak problemen met zich mee. Studenten vormen slechts een kleine groep in de samenleving. Resultaten uit dat

soort onderzoek zijn dan ook niet te generaliseren naar een brede groep mensen. Een bijkomend probleem is dat het ondervragen van een uitgebreide groep proefpersonen veel tijd kost, zodat het lastig is om een grote, betekenisvolle verzameling resultaten te krijgen. In dat opzicht is Twitter, dat werd opgericht in

'Twittergebruikers zijn welvarender en hoger opgeleid. Ze vormen een aparte groep.'

2006, een sociologisch geschenk uit de hemel. Zo'n 300 miljoen mensen gebruiken het netwerk, dat dagelijks groeit met ruim 300 miljoen berichten, ofwel tweets. Dat betekent een grote diversiteit aan 'respondenten' die met velen tegelijk laten weten wat ze doen, zien en vinden. Die 'schone' gegevens – niet beïnvloed door

observerende onderzoekers – liggen in compacte berichten opgeslagen, klaar om verslonden te worden door de wetenschap. 'Twitter biedt ons fantastische mogelijkheden,' zegt socioloog Scott Golder van Cornell University. 'We wilden altijd al live meekijken in het dagelijks leven van mensen, maar we hadden nooit de middelen om dat te doen.' Met Twitter meent Golder dat middel wel in handen te hebben. Op basis van twitterberichten onderzocht hij hoe het humeur van mensen gedurende de week verloopt (*Science*, september 2011). Hij analyseerde meer dan een half miljard berichten uit 84 landen en lette daarbij op bepaalde sleutelwoorden die iets verraden over het humeur van de twitteraars.

Apart

Golder concludeerde dat mensen 's ochtends goedgemutst opstaan, waarna hun stemming geleidelijk daalt en 's avonds weer omhoogschiet. Tijdens de eerste werkdagen van de week twitteren

► **Het onderzoeksteam van Michael Best analyseerde ongeregeldeheden tijdens de verkiezingen in Nigeria met behulp van Twitter.** MICHAEL BEST

mensen negatiever over het leven dan wanneer het weekend nadert. Op zaterdag en zondag hebben tweets de meest positieve toon (zie grafiek op pag. 43).

Werk lijkt dus een grote invloed te hebben op het humeur. Het is dan ook opmerkelijk dat de verschillen tussen de dagen erg groot zijn, vergeleken met de stemmingswisselingen gedurende de dag. Zo ligt de dinsdagse ochtendpiek op gelijke hoogte met het droevigste moment van donderdag. 'In het weekend heeft de grafiek dezelfde vorm als doordeweeks, dus werk heeft weinig te maken met de vorm,' verklaart Golder. 'Daar zijn biologische factoren voor verantwoordelijk. Sociale factoren verklaren de verschillen tussen de gemiddelden per dag. Daartoe hoort, naast werk, ook de stress die komt kijken bij de werkweek. Je hebt minder tijd voor bijvoorbeeld boodschappen.'

Representatie

De verleiding is groot om Twitter te zien als gegevensmachine waaruit de wetenschap naar hartenlust kan putten en, net als Golder, verrassende conclusies kan trekken. Toch kleven aan onderzoek op basis van tweets de nodige nadelen. Een meerderheid van de mensen heeft bijvoorbeeld geen account bij Twitter, waardoor het twitteronderzoek niet representatief is voor de wereldbevolking. Golder: 'Twittergebruikers vormen inderdaad een aparte groep. Ze zijn welvarender en hoger opgeleid. Toch zijn onze gegevens rijker dan die van traditionele onderzoeken, die vooral kleine groepjes studenten bestuderen.'

Nog een nadeel van onderzoek met Twitter is dat het weinig diepgang heeft. Klassieke proefpersonen werken uitgebreide vragenlijsten af, terwijl een tweet maximaal 140 leestekens bevat. Daarbij is het maar de vraag of mensen op Twitter wel het achterste van hun tong laten zien, en bijvoorbeeld niet veel vaker over positieve zaken berichten om goed voor de dag te komen.



Golder: 'De culturele verschillen tussen de 84 landen zijn groot. In sommige culturen is het *not done* om op te scheppen, zodat die mensen juist minder positieve geluiden laten horen.'

Het is maar de vraag of mensen op Twitter wel het achterste van hun tong laten zien

Andere onderzoekers werken aan manieren om twittergegevens meer diepte te geven. De Nijmeegse taalwetenschappers Eric Sanders en Maaske Treurniet verzamelen bijvoorbeeld tweets in combinatie met persoonsgegevens, zoals geslacht en

leeftijd. Sanders: 'Uit onze gegevens blijkt bijvoorbeeld dat vrouwen het woord 'ik' vaker gebruiken dan mannen. Dat kan betekenen dat vrouwen het vaker over zichzelf hebben. Maar het is ook mogelijk dat ze meer waarde hechten aan volledige zinsconstructies, terwijl mannen 'ik' vaker weglaten. Daar doen we nog verder onderzoek naar.'

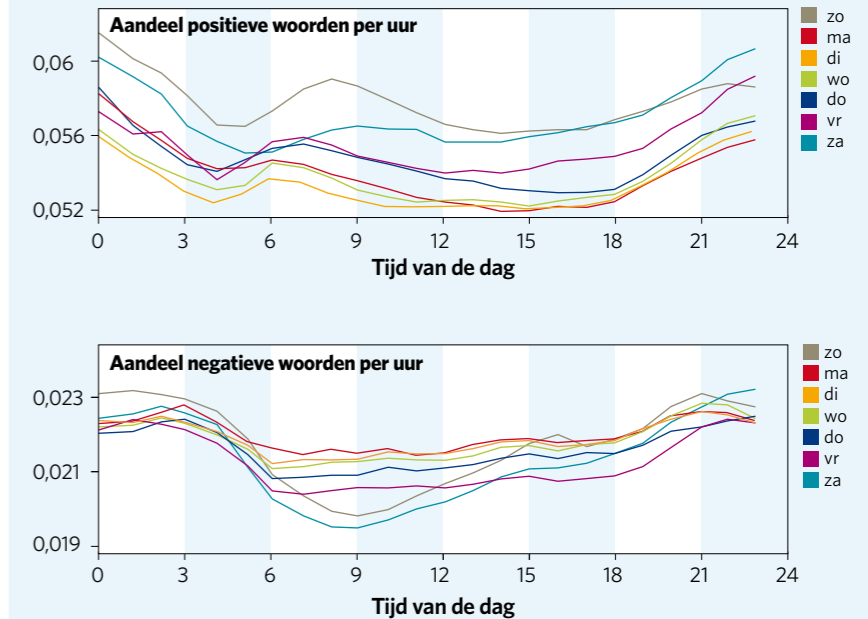
Aanvullende informatie wordt niet zomaar met tweets meegestuurd. De twee taalkundigen riepen twittergebruikers op hun gegevens op te sturen. Door de extra moeite die het kost om de gegevens te verzamelen, halen deze nooit de enorme kwantiteit van 'ruwe' twittergegevens waarop bijvoorbeeld Golder zich baseert.

Maar ook zonder die extra gegevens kunnen tweets nuttig zijn voor taalonderzoek, meent Treurniet. 'Je kunt bijvoor-

beeld een taalmodel bouwen voor dreig tweets. Die bevatten meestal specifieke woorden en slecht taalgebruik. Als je op basis van dat model software ontwikkelt die dreigtweets detecteert, kan de politie Twitter continu monitoren en bij een bedreiging meteen actie ondernemen.'

Ook uitingen van geweld in het echte leven zijn via Twitter op te merken. Dat bewijst onderzoeker Michael Best van het Georgia Institute of Technology, die ordehandhavers hielp met het opsporen van ongeregeldeheden tijdens de Nigeriaanse verkiezingen van afgelopen april. Hij ontwikkelde software die twitterberichten analyseert in *real-time*; het programma leest alle tweets uit Nigeria onmiddellijk in en beoordeelt met combinaties van sleutelwoorden en locaties of het gaat om een uitbraak van geweld.

Twitter als stemmingsmeter



▲ **De stemmingsontwikkelingen die socioloog Scott Golder waarnam in twitterberichten.** SCOTT GOLDER

'We zagen dat Twitter altijd sneller is dan traditionele media en informatiekanaal van de politie,' zegt Best. 'Ons systeem detecteerde bijvoorbeeld tweets van enkele vrouwen die vastzaten in een schoolgebouw. Ze waren omsingeld door bendes. De politie werd meteen gewaarschuwd. Ze hebben daar geen degelijk alarmnummer dat ze altijd kunnen bellen.'

Aandelen

Met Twitter kun je niet alleen live in de gaten houden wat er gebeurt, maar mogelijk ook voorspellingen doen. Onderzoekers van de universiteit van Tokio bestudeerden bijvoorbeeld de verspreiding van een aardbeving aan de hand van Twitter (*Proceedings of the World Wide Web*, april 2010). Een programma scant tweets op sleutelwoorden als 'aardbeving' en 'schudden', en let tegelijkertijd op de context waarin ze zijn geschreven. Wanneer een explosie van dit soort twitterberichten plaatsvindt, kijkt de software naar de plaatsnamen die mensen erbij vermelden en op welk tijdstip ze hun tweets verstuurden. Op die manier reconstrueert het

systeem de schok en wijst het razendsnel het epicentrum aan. Geregistreerde gebruikers in de buurt krijgen vervolgens automatisch een e-mail toegestuurd die waarschuwt voor de naderende aardbeving. In theorie kan dit systeem snel genoeg zijn om levens te redden. Een aardbeving beweegt zich voort met ongeveer honderd kilometer in twintig seconden. Mensen op meer dan driehonderd kilometer van het epicentrum worden dus op tijd gewaarschuwd door het bericht, dat ze volgens de auteurs meestal binnen een minuut ontvangen.

De Japanse onderzoekers detecteerden via Twitter 96 procent van de aardbevingen die schade veroorzaakten. Het is de vraag of bewegingssensoren geen degelijker en effectiever netwerk vormen, maar de Japanners tonen met hun onderzoek aan dat Twitter potentie heeft als opsporingsmiddel voor omvangrijke gebeurtenissen. Denk bijvoorbeeld aan een waarschuwingssysteem voor files of het in kaart brengen van een epidemie.

Dit zijn echter slechts voorspellingen van iets dat even daarvoor al is gebeurd op



een andere plek. Econoom Timm Sprenger van de technische universiteit van München vroeg zich af of Twitter een echt voorspellende gave heeft, en wel op de aandelenmarkt. Zijn voorlopige antwoord: waarschijnlijk wel. Sprenger schreef een programma dat tweets over de effectenbeurs analyseert. Het schat voor elke tweet in of het om een koop- of verkoopadvies voor een bepaald aandeel gaat. Een simpel voorbeeld is een bericht met als enige relevante woorden *buy* en \$GLD, wat een koopadvies is voor goud. Sprenger: 'Als we terugkijken op de voorspellingen die onze software maakte, zien we dat het mogelijk is om winstgevend te handelen met behulp van Twitter.' Over een periode van zes maanden realiseerde het programma vijftien procent *abnormal return* – de winst bovenop de verwachte stijging, veroorzaakt door speciale gebeurtenissen als fusies en schandalen.

Het aantal tweets over een aandeel staat volgens Sprenger in verband met het volume waarmee gaat worden gehandeld, iets wat beleggers ook graag van tevoren willen weten. 'Ons onderzoek geeft aan dat daar ongeveer een dag tussen zit.' Hoewel dat lang genoeg is om ernaar te handelen, geeft het ook meteen de handicap aan van Sprengers systeem. Twitter voorspelt slechts de kleine hobbeltjes in een aandelenkoers. Een langetermijnstrategie is er

'Het is mogelijk om met behulp van Twitter winstgevend te handelen'

niet op te bouwen. 'Twitter is zeker niet de heilige graal, maar het biedt degelijk, actueel financieel advies op basis van de collectieve wijsheid van honderden investeerders,' zegt Sprenger.

Het kenmerkt precies de waarde van twittergegevens. Elke individuele tweet bevat zeer weinig informatie en is onbruik-

baar voor onderzoek. Maar de dagelijkse stortvloed aan tweets heeft als collectief inmiddels gezorgd voor een kolossale verzameling van gegevens met een grote statistische waarde. Golder: 'Ik denk dat Twitter in de toekomst nog belangrijker wordt voor onderzoekers. Daarvoor moeten we sociale wetenschappers wel anders opleiden, want ze hebben betere computervaardigheden nodig.'

Best is het met hem eens. 'Zonder Twitter zouden we nooit miljoenen stukjes dialoog kunnen opvangen van enkel een verkiezing in Nigeria. Voor wetenschappers is het een geweldig instrument.'

De kwantiteit van twitterberichten compenseert het gebrek aan diepgang echter niet, waardoor onderzoekers niet zomaar conclusies mogen trekken over het echte leven. Sanders: 'In de taalwetenschap zeggen we: *'There is no data like more data.'* Maar tweets vormen wel heel specifieke data, die niet te vergelijken is met alledaagse spraak. Dat is ook interessant, maar het vervangt niet de traditionele onderzoeksmethoden.'

Bovendien zit er nog een schaduwkant aan gemakkelijk verkregen gegevens. Die zijn namelijk niet door de wetenschap zelf gegenereerd. Hoe meer de wetenschap onderzoek doet naar sociale media, des te afhankelijker ze is van de bedrijven die websites als Twitter en Facebook beheren. Best: 'Met name Facebook is zuinig met het vrijgeven van gegevens. Als sociale media niet alle informatie delen, beïnvloeden ze daarmee wetenschappelijke resultaten.'

Golder blijft echter onverminderd enthousiast: 'Er gaat een deur voor ons open.' ■



NWTmagazine
@NWTonline

Volg de twittervoorspellingen die de software van econoom Timm Sprenger genereert op www.tweettrader.net [#twitterwetenschap](https://twitter.com/tweettrader)

1:16 PM · 2 Feb via web · Embed this Tweet



Retweet



Delete



Favorite