Twijfel

HANS VAN MAANEN

Experiment met een muis. Foto Colourbox

De power van onderzoek

Neurowetenschap is een 'hot' onderzoeksgebied – zowel voor wetenschappers als voor journalisten. 'Overproductie van interleukine-6 verhoogt kans op alzheimer.' 'Migrainelijders hebben lange variant van het dopaminereceptorgen.' 'Prematuren vaker linkshandig' en 'Volwassenen met adhd mogelijk gebaat bij antidepressiva.' De kranten staan er vol mee, maar hoeveel waarde moeten wij hechten aan al die neurowetenschappelijke nieuwtjes?

Weinig, is het advies van een groep Britse en Amerikaanse onderzoekers vorige maand in, nota bene, *Nature Reviews Neuroscience*. De onderzoekers, onder aanvoering van Marcus Munafò, winden er geen doekjes om: 'Wij tonen aan dat het gemiddelde statistisch vermogen van studies in de neurowetenschap zeer laag is' – het statistisch vermogen zegt iets over de kans dat een onderzoek een werkelijk bestaand effect (of verschil, of wat dan ook) boven water krijgt. Wetenschappers spreken van power: een power van 80 procent betekent dat we acht van de tien keer het bestaande effect zullen vinden; twee keer komt de proef toevallig zo uit dat we het effect missen. Bijvoorbeeld: om met een power van 80 procent aan te tonen dat mannen gemiddeld langer zijn dan vrouwen – een groot, duidelijk effect – zijn zes mannen en zes vrouwen nodig. Twee op de tien keer zullen de mannen en vrouwen in de steekproef toch geen duidelijk verschil vertonen. Hoe kleiner het effect, des te groter de steekproef moet zijn: om 80 procent kans te hebben aan te tonen dat vrouwen meer schoenen hebben, zijn 16 mannen en 16 vrouwen nodig; om zo aan te tonen dat mannen zwaarder zijn, moeten we 47 mannen en 47 vrouwen vergelijken.

Het probleem van de neurowetenschap, zeggen Munafò en consorten, is dat ze op zoek zijn naar subtiele effecten maar daar evident te kleine steekproeven voor nemen. En ze kunnen het uiteraard niet laten ook andere wetenschapsgebieden – MRI-studies, experimenten met muizen en mensen – in hun beschouwing te betrekken.

De onderzoekers bedachten een mooie manier om te laten zien hoe 'underpowered' de meeste studies in de neurowetenschappen zijn. Naarmate meer onderzoek naar een verschijnsel wordt gedaan, komt langzamerhand het 'ware' effect er wel uit rollen – als er eindelijk een keer een echt grote studie wordt gedaan, of als iemand alle kleine studies bij elkaar optelt en analyseert. Terugkijkend kun je dan zien of de eerste studies groot genoeg waren om dit vastgestelde effect ooit te vinden. Dat blijkt dus maar zelden het geval te zijn, zegt Munafò. 'Onze resultaten wijzen erop dat de statistische power van studies in de neurowetenschappen waarschijnlijk niet veel hoger is dan 8 tot 31 procent.' En dat is vermoedelijk nog een overschatting.

'Het betekent,' vervolgt hij, 'dat de kans dat een significant resultaat op een echt effect berust, klein is.' Met name kleine studies met grote, opzienbarende effecten – pasmunt van de wetenschapspagina's – moeten met wantrouwen worden bekeken. Van de tien zullen er negen onzin wezen.

Het betekent ook dat er veel tijd, geld, moeite, proefdieren en proefpersonen worden verspild in al die te kleine studies. De onderzoekers geven het voorbeeld van ratjes die moeten leren door een doolhof te zwemmen. Leren vrouwtjes dat sneller dan mannetjes? Om dat aan te tonen, beperken onderzoekers zich veelal tot niet veel meer dan twintig ratten, maar om het bestaande effect met een power van 80 procent aan te tonen, zouden er in feite 134 nodig zijn geweest. Als twintig ratjes voldoende waren geweest, zou het verschil zo evident zijn dat het iedereen allang was opgevallen – en al die ratten dus evenmin hadden hoeven te zwemmen voor hun leven. 'De ethische implicaties zijn duidelijk,' aldus de onderzoekers.

Zij bepleiten, onder veel meer, een veel nauwkeuriger schatting van de nodige proefdieren en proefpersonen vooraf. Dat zijn er meestal beduidend meer dan onderzoekers zouden wensen, want grote experimenten zijn moeilijker gefinancierd te krijgen dan kleine.

Het streven naar snel leuke, nieuwe en onverwachte resultaten – van onderzoekers, geldschieters en tijdschriften – verspilt niet alleen de tijd van krantenlezers, het kost levens.

Oordeel:
Onzin (negen van de tien keer)

Digitale recherche

REPORTAGE HET PROFIEL VAN EEN MAILER

Veel reacties op het Amerikaanse spionageprogramma Prism komen neer op: ik heb niks te verbergen. Echt niet? Ook niet als digitale experts eens goed snuffelen?

Door **Bard van de Weijer** Illustratie **Thijs Balder**

Wie ben ik? TNO ging op zoek in mijn 32.272 mails

Het voelt alsof ik weer een jongetje van 6 ben dat bij Sinterklaas op schoot zit, mijn hersenen pijnigend met de vraag of ik het afgelopen jaar wel goed geluisterd heb naar mijn ouders, mijn best gedaan heb op school en het hok van de cavia vaak genoeg heb verschoond. Het is de vraag die elk kind zich stelt als de Goedheiligman nabij is: ben ik wel braaf geweest?

Maar ik zit niet bij Sinterklaas, ik zit naast Wessel Kraaij, gespecialiseerd in het verzamelen en analyseren van grote hoeveelheden ongestructureerde data. Tegenover hem zit Mark van Staalduinen, expert digital profiling. Hij adviseert namens TNO onder meer politie over het maken en analyseren van digitale profielen. Hoewel ik de twee mannen nog nooit heb ontmoet, weten ze alles van me. Ze weten met wie ik de afgelopen vijf jaar heb gecommuniceerd en hoe vaak en wat ik met al die mensen heb besproken. Het zou zomaar kunnen dat ze de krochten van mijn liefdesleven kennen en weten met wie ik zoal in onmin leef.

Ik was dus een beetje nerveus toen ik mijn paspoort toonde aan de portier van het anonieme TNO-gebouw in Delft en ik door het poortje liep waarachter Van Staalduinen me stond op te wachten. 'Hallo, ik ben Bard. Ik hoef me denk ik niet verder te introduceren.'

Ik heb dit zelf veroorzaakt. Ik heb Kraaij, behalve onderzoeker bij TNO ook bijzonder hoogleraar aan de Radboud Universiteit, enkele dagen eerder het wachtwoord gegeven van mijn privémailbox. Daarin staan alle inkomende berichten van de afgelopen vijf jaar: 32.272 mails, waarvan trouwens 28.076 ongezien – kennelijk ben ik niet zo'n zorgvuldige lezer.

De aanleiding dat Kraaij en Van Staalduinen mijn mail bekeken is Prism, het omstreden Amerikaanse afluisterprogramma. En de reactie daarop van veel Nederlanders die zeggen dat 'ze' alles van me mogen weten omdat ik toch niks te verbergen heb. Ik denk dat ik ook niks te verbergen heb (al weet ik dat van de afgelopen vijf jaar niet zo zeker), maar heb geen idee wat 'ze' dan allemaal te weten komen als mijn mailbox wordt geanalyseerd. Als Facebook op basis van een paar likes al weet wat mijn seksuele geaardheid is, wat zal er dan wel niet uit vijf jaar mailverkeer geconcludeerd kunnen worden? Welk beeld zal er van mij opdoemen bij iemand die me niet kent? Wat voor persoonlijkheid komt daar uit? Wellicht zaken die minder rooskleurig zijn, of zelfs iets waarop een overheid zou 'aanslaan'?

Er ontstaan clusters van afzenders die iets met elkaar hebben

Met deze vragen als uitgangspunt had ik contact gelegd met Kraaij en Van Staalduinen, die het een aardige casus vonden. Suzan Verberne, postdoctoraal onderzoeker aan de Radboud Universiteit, en promovendus Maya Sappelli trokken mijn mailbox leeg en analyseerden de ruim dertigduizend berichten: 86 megabyte tekst, exclusief attachments, die niet zijn bekeken. 'Als iemand deze hoeveelheid tekst 'handmatig' moet lezen', zegt Van Staalduinen, 'is hij ongeveer 500 uur kwijt.' Handmatig 'afluisteren' is alleen al om die reden onbegonnen werk. Daarom wordt de hulp ingezet van algoritmen die teksten kunnen analyseren. De computer van Verberne en Sappelli deed er een half uur over en stelde op basis van mijn inbox een netwerk samen van mensen met wie ik contact heb gehad.

Hoe gaat zoiets in zijn werk? Een belangrijke aanwijzing vormen de ontvangers die in de mails worden ge-cc'd, zegt Sappelli. Kennelijk zijn al de ontvangers binnen deze groep aan elkaar gelinkt. Als hetzelfde e-mailadres in meerdere cc's van verschillende afzenders voorkomt, is de bijbehorende persoon belangrijk en wordt zijn cirkel in het netwerkoverzicht groter. Zo iemand kan de spil zijn in een subnetwerk.

Ook wordt gekeken hoe vaak iemand mij mailt. Wie vaak contact legt, staat vermoedelijk dichter bij me en verschijnt daarom ook letterlijk dichterbij in het netwerk. Mensen die niet met elkaar communiceren zullen elkaar waarschijnlijk niet kennen. Zij worden daarom ondergebracht in een andere subgroep, waarvan opnieuw gekeken wordt wie met wie contact heeft. Op basis van deze vergelijkingen ontstaan clusters van afzenders die iets met elkaar hebben. Deze clusters hebben hun eigen kleur.

Met deze analyse is mijn sociale netwerk in kaart gebracht. Alleen zegt een overzicht van al deze mensen nog niet zo veel, zegt Van Staalduinen. 'Je ziet dat je kennelijk vrij veel communiceert met de Vereniging van Eigenaren. Als je weet dat dat niet belangrijk is, kun je stukken uitsluiten. Dan hoef je verder niet meer naar die VvE te kijken en kan worden ingezoomd op andere clusters.' Dat inzoomen betekent letterlijk kijken waarover wordt gemaild.

Van mensen uit een cluster die dicht bij me staan of die als belangrijk worden beschouwd, kan per afzender een termenwolk worden gemaakt. Zo'n wolk, bekend van sommige zoekmachines op internet, toont allerlei veelgebruikte woorden. Hoe vaker een woord wordt gebruikt, hoe groter het wordt afgebeeld. 'Met een termen-

JE EIGEN NSA

Wie zelf NSA'tje wil spelen: op internet staan allerlei hulpmiddelen die je berichtenverkeer op sociale netwerken als Facebook en Twitter analyseren.

Tweetgenie.nl
Tweetgenie, een onderzoeksproject van Dong Nguyen van de Universiteit Twente, probeert op basis van Nederlandstalige tweets het geslacht en de leeftijd te raden van de twitteraar. Blijkt heel aardig te werken.

youarewhatyoulike.com
Probeert persoonlijkheidskenmerken vast te stellen op basis van 'likes' op Facebook.

www.mapmyfollowers.com
Geografische weergave van volgers.

twitonomy.com
Geeft een overzicht van meestgenoemde twitteraars, vaakst gebruikte hashtags, vaakst geretweete gebruikers en hoe vaak je geretweet bent. En hoe laat je meestal twittert. En op welke dagen van de week.

Nederlandse gamemakers hebben een – enigszins mellige - online-game ontwikkeld over de zaak-Snowden. Daarbij kruipt de gamer in de huid van klokkenluider Edward Snowden, die usb-sticks laadt met gevoelige informatie, om deze uit het raam van het NSA-gebouw te werpen zonder te worden betrapt door beveiligers en camera's. bit.ly/1aPUsxn

ANALYSE E-MAIL

Voorbeelden van termenwolken van onderwerpen waarover veel is gemaild

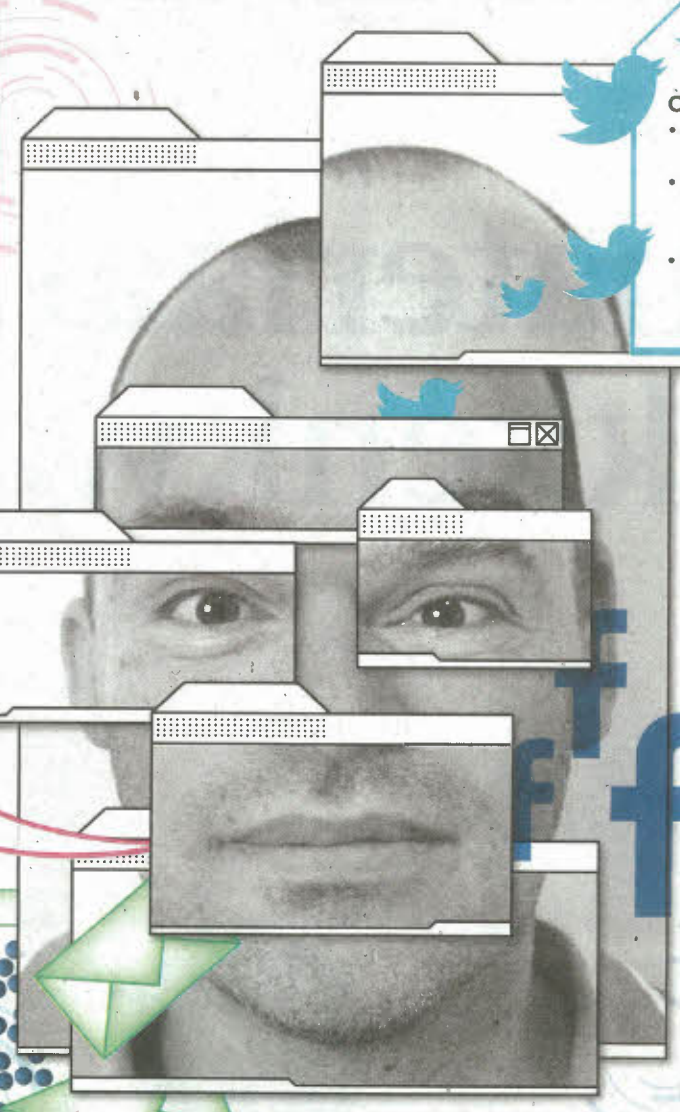
TWITTER

Volkskrant-datajournalist Sybren Kooistra heeft een analyse gemaakt van Bards uitingen op Twitter en Facebook. Daaruit kwam het volgende naar voren:

Communiceert vooral met
- Wiskundemeisje
  - Ionica Smeets
- wetenschapcollega's
  - Martijn van Calmthout
  - Maarten Keulemans
- Andere journalisten
  - Toine Heijmans
  - Peter Vandermeersch

Onderwerpen
- (Elektrische) auto's
- Ruimtevaart
- Technologie

volgt David Pogue, Elon Musk



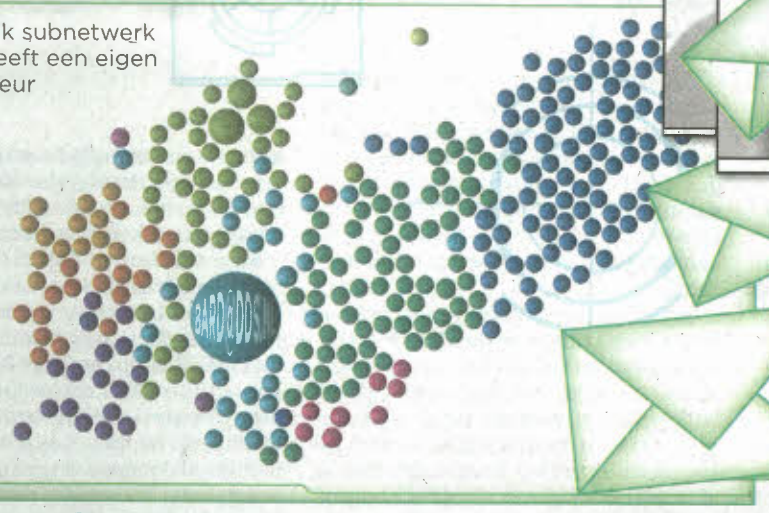
EN WAT ZEGT FACEBOOK?

- Houdt van jazz en motown
- Valt op mannen* en vrouwen
- Interesse in Eerste Wereldoorlog/ ruimtevaart/koken
- Is op het werk actiever op FB dan thuis, vooral op donderdag
- Vriendin is niet de moeder van de kinderen*
- Actief op Facebook sinds-6 dec 2009
- Woont in Oude Diemen*
- Jarig op 29 september

** Klopt niet*

HET NETWERK VAN BART

Elk subnetwerk heeft een eigen kleur



wolk kun je heel snel belangrijke conversatie-onderwerpen ontdekken', zegt onderzoekster Verberne. Woorden die daarin veel voorkomen kun je gebruiken als 'ingang' voor verder onderzoek.

Om te voorkomen dat veelgebruikte woorden als 'het' en 'groetjes' telkens het beeld domineren, wordt alle mail eerst vergeleken met een referentie-e-mailbestand. Woorden die in beide datasets veel voorkomen, worden uitgesloten. Wat overblijft zijn als het goed is woordwolken met zeggingskracht. 'Ik zag bijvoorbeeld dat een afzender Lieve Bart in zijn aanhef zette', zegt Verberne. 'Dat is merkwaardig, omdat 'lieve' een intieme band suggereert terwijl de afzender je voornaam toch verkeerd spelt.'

Termenwolken zijn de vuurtorens van de datazee. Ze wijzen op de onderwerpen waarover gesproken wordt, maar kunnen bijvoorbeeld ook helpen bij het ontdekken van gecodeerd taalgebruik. In de drugsscene was het in de jaren negentig gebruikelijk om bij de dealer 'versnellingsbakken' te bestellen als speed bedoeld werd. Dat zou de eventueel meeluisterende politie om de tuin moeten leiden. Bij de analyse van het e-mailverkeer zou deze omschrijving er in de termenwolk zonder meer uitknallen en zou al snel duidelijk worden wat ermee bedoeld wordt.

De algoritmen kunnen ook zien of de toon van de berichten positief of negatief is, door te kijken of woorden die als negatief worden beschouwd vaak voor komen. Zo ontdekte Sappelli dat ik kennelijk 'een issue heb gehad' met de buren over poezenoverlast. Inderdaad: we hadden tot voor kort een kat, net als enkele andere buren, en een van deze katten poepte soms op het gazon van een van de buren, waar kleine kinderen spelen. Dat leidde enkele jaren terug tot wat geagiteerd mailverkeer. Ik was het alweer vergeten, mijn mailbox niet.

De grote vraag is natuurlijk wat de onderzoekers met al hun graafwerk van me te weten zijn gekomen. Hebben ze een beeld van wie ik ben? Hoogleraar Kraaij gaat er even goed voor zitten. Ze hebben achterhaald dat ik op IJburg woon (inclusief adres), dat mijn vriendin psycholoog is en in de verslavingszorg werkt (ook bij welke instelling). Ze weten zelfs op welke dagen ze werkt. Ze weten dat ik een zoon Teun heb, op welke school hij zit en waar hij zwemles heeft. Ze weten dat ik een kat heb en bovengemiddeld geïnteresseerd ben in Mercedes-Benz (tot ongeveer een jaar geleden reden we in een klassieke Mercedes en ik was lange tijd actief in een club).

Dat is een behoorlijk compleet gezinsplaatje, al is een belangrijk aspect gemist: er is nog een tweede zoon. Die is wellicht over het hoofd gezien door zijn bijzondere naam: Dieks. Ooit een populaire voornaam, inmiddels hevig in onbruik geraakt; er zijn nog maar 19 Dieksen in Nederland. 'We hebben hem voorbij zien komen, maar hem niet als zoon opgemerkt', zegt Verberne. Waarmee blijkt dat de analyse van de data mensenwerk blijft, waarbij vergissingen mogelijk zijn.

'Verder niks spannends?', vraag ik. Geen minnaressen? Geen extreme politieke voorkeuren? Geen zaken waar een overheid op zou aanslaan? 'Nee, niets van dat alles', zegt Verberne. 'Als we je uitgaande mail hadden kunnen bekijken, hadden we misschien meer gevonden. Wat je zelf verstuurt, zegt meer over je dan wat je ontvangt.' Maar die data is niet beschikbaar. De conclusie is helder. 'Wat er niet in zit, komt er ook niet uit', zegt Kraaij. 'Dit is kennelijk wie je bent.'

Ik voel een lichte teleurstelling dat in mijn leven de VvE kennelijk zo'n grote rol speelt. Maar er is ook opluchting, omdat het nu officieel vaststaat: ik ben braaf geweest.

☢ MAG DE OVERHEID (OF JE BAAS) ALLES GEBRUIKEN WAT OPENBAAR IS?

De overheid mag niet zomaar doen wat de onderzoekers van de Radboud Universiteit en TNO hebben gedaan: iemands mailbox lezen. 'Ook de politie heeft natuurlijk iemands wachtwoord nodig om in de mail te kunnen komen, en toestemming van de rechter-commissaris voor een e-mailtap', zegt Mark van Staalduinen, die voor TNO de politie adviseert over digital profiling, het maken van profielen op basis van sporen die iemand digitaal nalaat.

Door de onthullingen over Prism, het snuffelnetwerk dat de Amerikaanse veiligheidsdienst NSA gebruikt om de gangen van burgers na te gaan op onder meer sociale netwerken, is ook in Nederland de discussie opgelaaid over wat de recherche en veiligheidsdiensten mogen inzien. Een van de vragen is of gebruik mag worden gemaakt van openbare bronnen als Twitter en de publiek toegankelijke data op Facebook. 'Je kunt argumenteren dat mensen die informatie daar nooit hebben achtergelaten om de politie eventueel van dienst te zijn', zegt Van Staalduinen. Aan de andere kant verlangt de samenleving dat de overheid ingrijpt als iemand op een openbaar prikbord verkondigt dat hij een slachting wil aanrichten op een middelbare school, stelt hij. 'In dat spanningsveld moeten we opereren. En waar de grens precies ligt, zal gaandeweg duidelijker worden.'

De techniek die hier gebruikt is om de mailbox van de verslaggever te analyseren, is niet alleen nuttig bij opsporingsdoeleinden, ze kan ook helpen het welzijn van bijvoorbeeld werknemers te verbeteren. Hoogleraar Wessel Kraaij werkt aan de Radboud Universiteit aan het project Swell, waarbij hij vergelijkbare methoden toepast om ervoor te zorgen dat werknemers bijvoorbeeld niet overbelast raken. 'Je kunt bijvoorbeeld iemands mail scannen en dan ontdekken dat hij steeds meer woorden gebruikt die duiden op stress', zegt Kraaij. In zo'n geval zou de software, met wat de hoogleraar 'zachte signalen' noemt, een werknemer daarop kunnen wijzen en hem adviseren de e-mail per project te lezen in plaats van op volgorde van ontvangst.

De software zou ook binnenkomende mail kunnen filteren op relevantie. 'Als je op je werk zit is een bericht van de Vereniging van Eigenaren minder relevant, dat kan zo'n systeem dan even elders parkeren. Terwijl een mailtje van iemand die je net hebt geïnterviewd wel doorkomt', aldus Kraaij. Doel van het project is om op langere termijn gezondere en dus productievere werknemers te krijgen. Ook hier vormt privacy een heet hangijzer. 'Mensen willen niet dat de manager kan meekijken om te zien hoe gestrest iemand is. Men wil de baas blijven over de eigen data.'