

Onderzoeksnota :

Ik like dus ik ben

Het verkrijgen van inzichten in politieke voorkeur door middel van
Facebookdata

Stiene Praet^a, Peter Van Aelst^b, David Martens^c

^a Doctoraatsstudent aan de onderzoeksgroep Applied Data Mining, Universiteit Antwerpen
(stiene.praet@uantwerpen.be; +32 3 265 40 17)

^b Professor aan de onderzoeksgroep Media, Middenveld en Politiek, Universiteit Antwerpen
peter.vanaelst@uantwerpen.be; +32 494 17 24 57)

^c Professor aan de onderzoeksgroep Applied Data Mining, Universiteit Antwerpen
(david.martens@uantwerpen.be; +32 3 265 40 69)

Samenvatting voornaamste bevindingen:

1. Op basis van Facebooklikes zijn we vrij goed in staat om de politieke voorkeur van iemand te voorspellen.
 - In 60% van de gevallen kunnen we correct voorspellen of iemand zich links, in het midden of rechts op het politieke spectrum bevindt. Kiezers met een duidelijk links of rechts profiel kunnen we met een hogere betrouwbaarheid (66%) identificeren.
 - In 64% van de gevallen kunnen we correct voorspellen of iemand veel sympathie heeft voor een politieke partij. Onder veel sympathie verstaan we een score van 7 of meer op 10 op de vraag hoe groot de kans is om ooit voor die partij te stemmen.
 - In 36% van de gevallen kunnen we exact voorspellen voor welke partij iemand het meeste sympathie heeft op dit moment (dit lijkt weinig maar is substantieel beter dan willekeurig gokken wat leidt tot een accuraatheid van $1/7 = 14\%$). De partijvoorkeur voor nieuwere of meer extreme partijen laat zich ook heel wat beter voorspellen (40-50%) dan deze van de klassieke centrumpartijen.
2. Ook niet-politieke interesses zoals muziek, films, boeken, festivals, ... die je liket op Facebook zeggen iets over je politieke voorkeur. Zelfs zonder politieke Facebookpagina's in rekening te brengen, kan politieke voorkeur accuraat voorspeld worden.
3. Mensen met een verschillende politieke voorkeur hebben doorgaans ook andere interesses. De Facebookpagina's die het meest gecorreleerd zijn met een bepaalde politieke voorkeur kunnen gebruikt worden om de bredere interesses van kiezers te omschrijven.
4. Via een nieuwe index kunnen de Facebookpagina's gerangschikt worden volgens politieke homogeniteit. De Facebookpagina's van cafés, cultuur, boeken en evenementen hebben vaker een homogeen links of rechts publiek. Aan de andere kant brengen voornamelijk sport en TV-shows mensen met verschillende politieke voorkeur samen. In Vlaanderen hebben de meeste populaire pagina's een gemengd publiek, en zijn de organisaties of personen met een uitgesproken rechts of links profiel veelal kleinere nichespelers.

Inleiding

Stemmen staat centraal in het democratisch proces en ons politiek systeem. Stemgedrag en partijvoorkeur worden dan ook al jaren bestudeerd door politieke wetenschappers. Veelal doen politicologen dat via de studie van verkiezingsuitslagen en opiniesurveys bij kiezers. Vandaag de dag is er echter een nieuwe, en zeer rijke bron aan data beschikbaar om dit gedrag te bestuderen: sociale media. In deze studie willen we achterhalen wat Facebooklikes ons kunnen vertellen over iemands politieke voorkeur. Meer specifiek zijn we geïnteresseerd in de niet-politieke Facebooklikes, bijvoorbeeld muziek of films, en hoe deze gelinkt zijn aan politieke voorkeur. Alvorens onze bevindingen toe te lichten bespreken we beknopt hoe we de data hebben verzameld.

Dataverzameling

In juni 2018 werden via enkele populaire Vlaamse kranten burgers opgeroepen om deel te nemen aan de Facebookstudie van het NWS data project van de Universiteit Antwerpen (<https://www.uantwerpen.be/nl/projecten/nws-data/facebookstudie/>). Een totaal van 6.733 Facebookgebruikers nam deel aan onze studie. Deelnemers konden ons toegang tot hun Facebookdata (enkel de publieke pagina's die ze hebben geliket en de posts op hun tijdlijn) verlenen en tien vragen beantwoorden. Die vragen werden gebruikt om na te gaan wat we juist van mensen kunnen voorspellen op basis van hun paginalikes op Facebook. Na het invullen van de vragenlijst werd aan de deelnemer getoond wat voorspellingsmodellen over hem/haar zouden voorspellen op basis van zijn/haar Facebooklikes wat betreft geslacht, leeftijd en politieke voorkeur. We toonden hierbij ook welke specifieke likes (de Facebookpagina's die de deelnemer zelf geliket heeft) het meeste bijdragen tot deze voorspelling. Dit om gebruikers er bewust van te maken hoe veel persoonlijke informatie er afgeleid kan worden uit hun Facebooklikes. Deze persoonlijke data werden verzameld met complete transparantie van onze doelstellingen en met respect voor de privacy van de deelnemers. Daarin willen we ons uitdrukkelijk onderscheiden van bedrijven zoals Cambridge Analytica. Ons volledig privacybeleid kan teruggevonden worden op <https://www.uantwerpen.be/nl/projecten/nws-data/privacybeleid/>.

De deelnemers aan onze studie zijn divers, maar niet representatief voor de bevolking als geheel. Er is een ondervertegenwoordiging van kiezers van Vlaams Belang, N-VA en CD&V, en een oververtegenwoordiging van Groen-kiezers. Deze misrepresentatie heeft echter geen directe invloed op de resultaten van deze studie: er wordt getracht voorspellingen en inzichten te bekomen, en geenszins om een politieke peiling te bekomen.

Voorspellen van politieke voorkeur aan de hand van Facebooklikes

Voorgaande studies, zoals die van Kosinski¹, tonen aan dat persoonlijke kenmerken voorspeld kunnen worden aan de hand van Facebookdata. Wat je leuk vindt op Facebook zegt iets over bijvoorbeeld je geslacht en leeftijd. Zo kunnen we op basis van onze data met 86% zekerheid zeggen of iemand een man of vrouw is. Maar ons doel is om ook iemands politieke voorkeur in te schatten en daarom beantwoordden de deelnemers twee vragen over hun politieke voorkeur:

1. *Politiek spectrum.* Deelnemers positioneren zichzelf op een ideologische links-rechts schaal van 0 (meest links) tot 10 (meest rechts).
2. *Partijvoorkeur:* De deelnemers gaven aan voor welke van de 7 grote Vlaamse partijen ze meest waarschijnlijk zouden stemmen (0= nooit op stemmen; 10= zeker op stemmen).

Aan de hand van de Facebooklikes en de antwoorden op deze vragen van meer dan 6.700 mensen bouwden we voorspellingsmodellen om politieke voorkeur te voorspellen. In 60% van de gevallen kunnen we juist inschatten of iemand zich links, in het midden of rechts op het politieke spectrum bevindt. Dat is beduidend meer dan toevallige gok (kans van 1 op 3 of 33%), maar nog niet heel hoog. Dat is gedeeltelijk te wijten aan de moeilijkheid om de mensen 'in het midden' te voorspellen. Het is duidelijk dat zij een minder uitgesproken politiek profiel hebben, ook op Facebook. Kiezers met een duidelijk links of rechts profiel kunnen we met een hogere betrouwbaarheid (66%) identificeren.

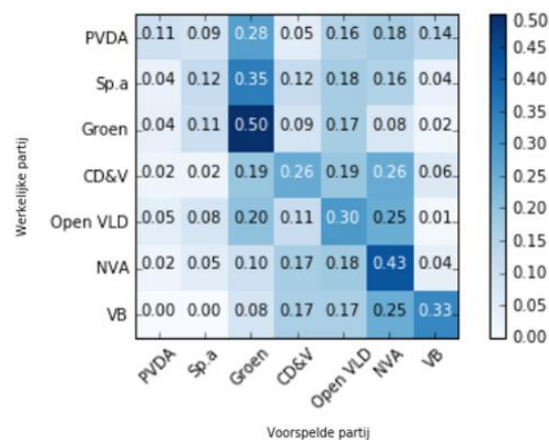
In 64% van de gevallen kunnen we correct voorspellen of iemand veel sympathie heeft voor een politieke partij. Onder veel sympathie verstaan we een score van 7 of meer op 10 op de vraag hoe groot de kans is om ooit voor die partij te stemmen. In 36%² van de gevallen kunnen we ook exact voorspellen voor welke partij iemand het meeste sympathie heeft op dit moment. We zien dat de voorkeur voor meer uitgesproken partijen, zoals PVDA, Vlaams Belang, maar ook Groen en N-VA gemakkelijker te voorspellen zijn dan voor de klassieke centumpartijen. Hun kiezers zijn dus gemakkelijker te onderscheiden van anderen op basis van de pagina's die ze leuk vinden op Facebook. Wanneer er een foute voorspelling plaatsvindt, zien we dat vaak een voorspelling voor een politiek nabije partij wordt gegeven (zie Figuur 1).

Daarnaast toont onze studie ook aan dat politieke voorkeur nog steeds te voorspellen is wanneer we de expliciet politieke Facebookpagina's (partijen, partijleden en politieke organisaties) weglaten. Eigenlijk daalt de totale voorspellingskracht van ons model nauwelijks. De muziek, films, boeken, festivals, ... die je liket op Facebook zeggen dus ook iets over je politieke voorkeur.

¹ <https://www.pnas.org/content/pnas/110/15/5802.full.pdf>

² Een random gok zou slechts in 1 op 7 gevallen juist zijn, dus 14%

Figuur 1: In de confusion matrix voor het voorspellen van politieke partij zien we welke ‘vergingingen’ het model maakt. Bijvoorbeeld, voor kiezers die in werkelijkheid aangeven op Open Vld te stemmen wordt ook vaak (25%) voorspeld dat ze op N-VA stemmen.



Inzichten in de bredere interesses van kiezers

Door te kijken naar de correlaties tussen Facebookpagina’s en politieke voorkeur kunnen we de Facebookinteresses van kiezers omschrijven. Mensen die zichzelf links positioneren op het politieke spectrum volgen eerder alternatieve of linkse media (bijv. Apache, De Wereld Morgen) en progressieve organisaties (Klimaatzaak, Vluchtelingenwerk) op Facebook. Wie aan de andere kant Vlaams-nationalistische Facebookpagina’s liket op Facebook zal zichzelf ook eerder rechts op het politieke spectrum positioneren. Voor mensen die zichzelf in het midden plaatsen vinden we een brede diversiteit aan pagina’s terug met een minder duidelijk profiel.

Tabel 1: Deze Facebookpagina’s (met uitsluiting van pagina’s van politici of partijen) zijn het meest gerelateerd aan links, midden en rechts.

Links	Midden	Rechts
De Wereld Morgen	Roompot Vakanties	De Fiere Vlaamse Meme
MO*	MAC Cosmetics	Museum voor Vlaanderen
Apache	TEDxBussels	Onafhankelijk Verbond Der Vlaamse Meme
Klimaatzaak	Farm Heroes Saga	De Fiere Vlaamse Meme: Onderduikadres
Vluchtelingenwerk Vlaanderen	Mijn Leven in de Stad	Schild en meme
De Morgen	Vrouwen hebben altijd gelijk, ook als ze ongelijk hebben	VNZ - Vlaams & Neutraal Ziekenfonds
EVA	Hotel Hungaria	Pedagogisch Verantwoorde Vlaamse Memes
Ringland	TLC Vlaanderen	Ik werk extra hard op 15 december
Amnesty International België Vlaanderen		Make Vlaanderen Great Again
Climate Express		Vlaanderen Tegen Islamisering - VTI

Zo kunnen we ook per partij bekijken welke pagina's hun aanhangers leuk vinden op Facebook. Voor de PVDA en Sp.a zijn dit met name linkse media en vrijwilligersorganisaties, voor Groen is dat klimaat en milieu, voor CD&V onderwijs en gezondheid, voor Open Vld financieel nieuws en ondernemen, voor N-VA en voor Vlaams Belang tot slot zijn dat Vlaams-nationalistische media, organisaties en content.

Tabel 2: Deze Facebookpagina's (met uitsluiting van expliciete politieke pagina's) zijn het meest gerelateerd aan de zeven grote Vlaamse partijen.

PVDA	Sp.a	Groen	CD&V
Solidair	Curieus Oost-Vlaanderen	Klimaatzaak	KU Leuven Confessions
EPO Uitgeverij	Curieus	Dagen Zonder Vlees	KU Leuven confessions 3.0
De Wereld Morgen	Samenleving & Politiek	Greenpeace Belgium	Kerknet
Apache	Zij-kant	EVA	Make-A-Wish Belgium-Vlaanderen
ABVV	SamenSterker	Natuurpunt	KU Leuven
Intal Globalize Solidarity	Huiswaluw Home	De Morgen	Belgische Monarchie
Lava Tijdschrift	Kd Redactie	MO*	Ik heet Bert, niet Birt
G3W Geneeskunde voor de Derde Wereld	GO! onderwijs van de Vlaamse Gemeenschap	Ancienne Belgique - AB	Christelijke Mutualiteit
Jacobin Magazine	Kapitein Winokio	Balthazar	The Royal Family
Vluchtelingenwerk Vlaanderen	OLT Rivierenhof	Netwerk Bewust Verbruiken vzw	UZ Leuven

Open Vld	N-VA	Vlaams Belang
Liberale Vrouwen	SCEPTR	ReAct
De Tijd	Christiaan Janssens	SCEPTR
Financial Times	Doorbraak	Vlaanderen Tegen Islamisering - VTI
Liberaal Archief/Liberas	Museum voor Vlaanderen	Weekblad 't Pallieterke
Suitsupply	De Fiere Vlaamse Meme	Forum voor Democratie -FVD
Belgian Red Devils	VNZ - Vlaams & Neutraal Ziekenfonds	De Fiere Vlaamse Meme
European Liberal Forum	Weekblad 't Pallieterke	Make Vlaanderen Great Again
Abercrombie & Fitch	VL+ Vlaams-nationale duiding	Schild en meme
The Economist	Werkbond	Donald J. Trump
House of Cards	Kunnen wij Vlamingen 100 000 fans verzamelen op 1 pagina ?	Large Merchandising

Inzichten in ideologische homogeniteit

Het aantal mensen dat zichzelf links positioneert tegenover het aantal mensen dat zichzelf rechts positioneert dat een bepaalde pagina liket is een maatstaf voor homogeniteit. Als een pagina uitsluitend door linkse of uitsluitend door rechtse kiezers geliket wordt, dan kunnen we deze pagina als ‘ideologisch homogeen’ bestempelen. Als een pagina in min of meer gelijke mate door linkse en rechtse kiezers geliket wordt dan is ze ideologisch heterogeen. De gebruikte metriek is de “entropie”, dewelke maximaal is bij een 50/50 heterogene verdeling, en minimaal bij een volledig homogene (links of rechts) verdeling. Over het algemeen zijn politieke pagina’s en middenveldorganisaties zeer homogeen (lage entropie) maar ook de rubrieken nachtleven, cultuur en boeken zijn meer homogeen dan gemiddeld. Aan de andere kant zijn TV-shows, reizen en sport het meest heterogeen (hoge entropie) en worden dezelfde pagina’s dus zowel door linkse als rechtse kiezers evenveel geliket.

Tabel 3: 20 Rubrieken gerangschikt van lage (zeer homogeen) naar hoge (zeer heterogeen) entropie.

Rubriek	Omschrijving	Gemiddelde entropie
Politici en politieke organisaties	Politici, politieke partijen en overheidsinstellingen	75% (zeer homogeen)
Middenveld en goede doelen	Non-profit organisaties, vakbonden en religieuze organisaties	81%
Bars, cafés en clubs	Bars, cafés, pubs en clubs	83%
Kunst en cultuur	Kunst, cultuur en musea	89%
Boeken en auteurs	Boeken, uitgevers en auteurs	89%
Evenementen en festivals	Evenementen, concerten en festivals	89%
Muziek	Muziek, groepen, labels, albums, etc.	91%
Nieuws en media	Nieuws, media, radio en magazines	93%
Producten en diensten	Producten, merken, financiële diensten, marketing, etc.	94%
Groepen en interesses	Facebookgroepen (ludieke pagina’s bv. ‘I like coffee’) en interesses	94%
Humor en entertainment	Humor, komedie, amusement en games	94%
School, universiteit en onderwijs	Scholen, universiteiten, studentenverenigingen, onderwijs	94%
Apps, websites en blogs	Apps, websites en blogs	95%
Artiesten en beroemdheden	Artiesten en publieke figuren	95%
Films en acteurs	Films, acteurs en cinema	95%
Bedrijven en ondernemingen	Bedrijven, ondernemingen, shops, winkels, etc.	95%
Eten en drinken	Voeding, koken, brouwerijen, restaurants, etc.	96%
TV-programma’s en omroepen	TV-programma’s en openbare omroepen	97%
Toerisme en reizen	Reizen, reismaatschappijen en toerisme	97%
Sport en gezondheid	Sport, atleten en gezondheid	98% (zeer heterogeen)
Alle Facebookpagina’s		93%

Als we inzoomen op de verschillende categorieën dan zien we ook dat de populairste Facebookpagina's binnen een categorie doorgaans zeer heterogeen zijn en dat de kleinere nichepagina's vaker door een uitsluitend links of rechts publiek geliket worden (zie Figuur 2). Neem nu bijvoorbeeld de categorie 'Nieuws en media' waar VRT NWS zich als grootste speler pal in het centrum bevindt. Bij humor geldt nagenoeg hetzelfde voor de populaire pagina van Jeroom of Phillipe Goebels. Bij TV-programma's valt ook de centrale positie van De Ideale Wereld op, onder onze deelnemers het meest gelikte programma. Blijkbaar kunnen mensen aan beide zijden van het politieke spectrum dit programma dat vaak de draak steekt met politiek en politici appreciëren. Bij de middenveldorganisaties en goede doelen zien we niet verwonderlijk wat meer pagina's langs de linkerzijde. Al heeft ook hier de populairste pagina 'Music for Life' een meer centrumprofiel. De rubrieken kunst en cultuur en ook cafés en bars kennen een duidelijke oververtegenwoordiging bij mensen die zichzelf vaker als links beschouwen. De categorie 'Bedrijven en ondernemingen' heeft een bescheiden afwijking naar rechts. TV-programma's, Toerisme en Sport lijken de meest heterogene categorieën. Zo staan bij deze laatste de populaire Rode Duivels opnieuw pal in het midden.

Aangezien deze data vrij uniek zijn kunnen we ze moeilijk direct vergelijken met ander internationaal onderzoek. Voor nieuwsmedia zijn al wel gelijkaardige analyses gedaan (op basis van surveyonderzoek) in andere landen en dan blijkt dat er bijvoorbeeld in de VS veel grotere spelers met een homogeen links of rechts publiek zijn. In Vlaanderen blijkt dat voorsnog veel minder het geval te zijn, wat ook bevestigd werd in eerder onderzoek³.

Tot slot, moest ons land nog een nieuwe premier zoeken waarin zowel rechts als links zich kan vinden, dan raden we Barack Obama aan.

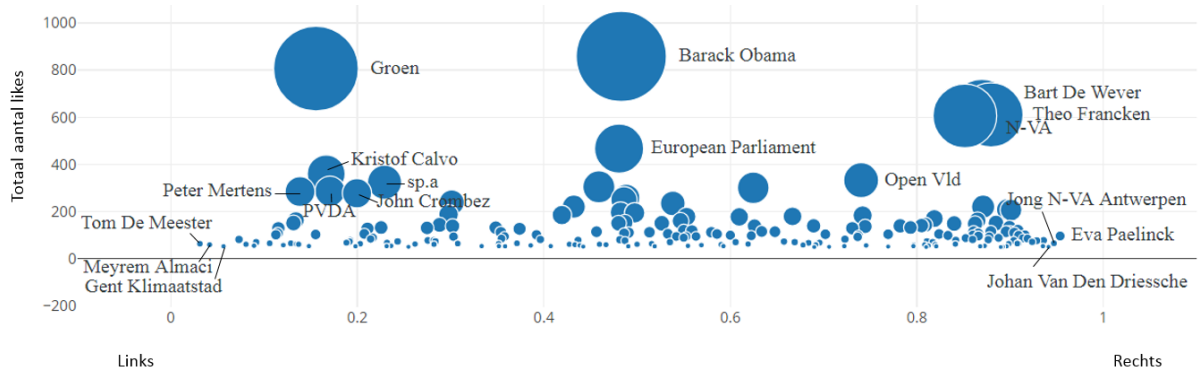
Tot slot

Het gebruik van dit soort fijnmazige data heeft tot voordeel dat zeer uitgebreide interesseprofielen bekomen kunnen worden, wat zeer moeilijk te bevragen is in een survey. Helaas heeft Facebook in reactie op het Cambridge Analytica schandaal de toegang voor onderzoekers tot dit soort data bemoeilijkt. Hoewel bepaalde restricties wenselijk zijn om de privacy van Facebookgebruikers te beschermen hebben deze strenge maatregelen een nefast effect op transparantie en toekomstig onafhankelijk onderzoek.

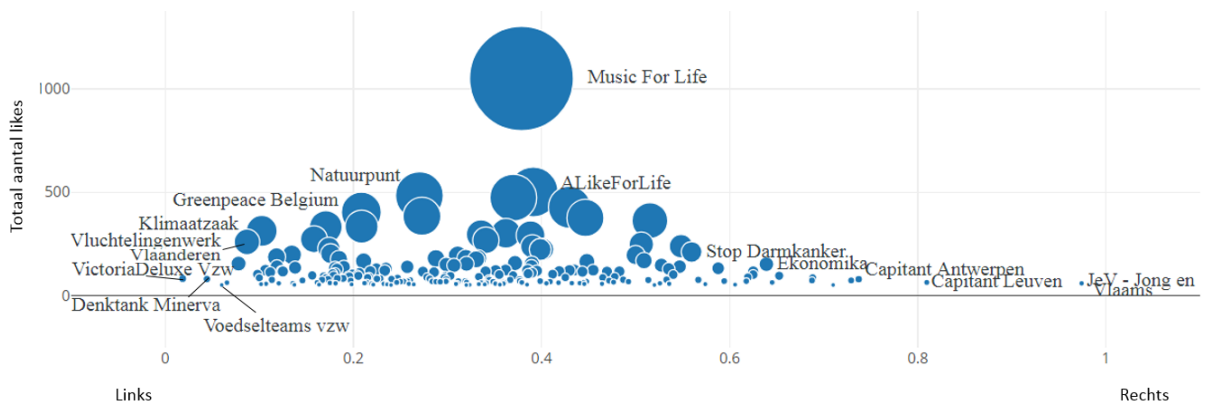
³ Picone, I. & Vandenplas, R. (2018). Ook Vlaming maakt zich zorgen over fake news. *Studies Media Innovation Technology (SMIT)*. 33 #16. <http://smit.vub.ac.be/wp-content/uploads/2018/06/POLICY-BRIEF16-1.pdf>

Figuur 2: Facebookpagina's gerangschikt van links naar rechts⁴ (gebaseerd op het aantal linkse en rechtse kiezers die de pagina geliket hebben). De grootte van de bol geeft het totaal aantal likes van de pagina weer.

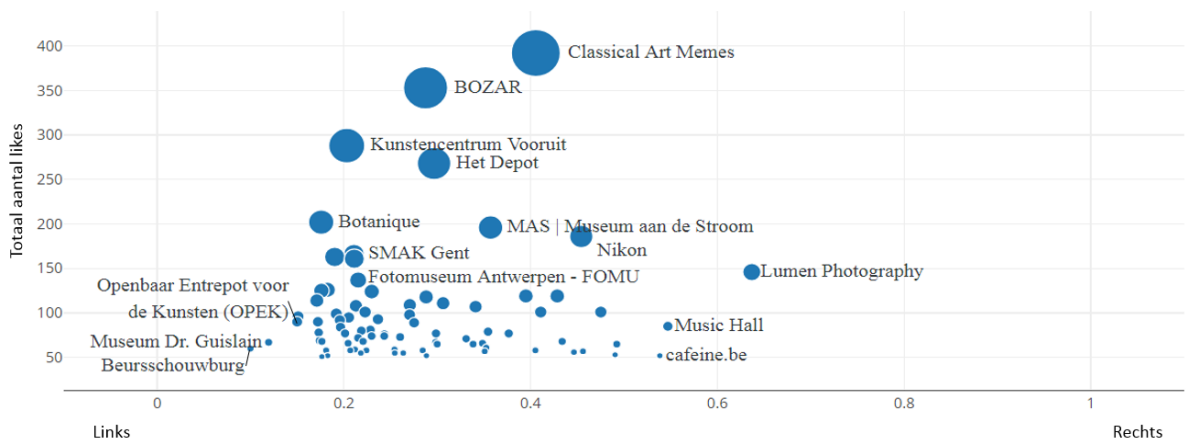
(a) Politici en politieke organisaties



(b) Middenveld en goede doelen

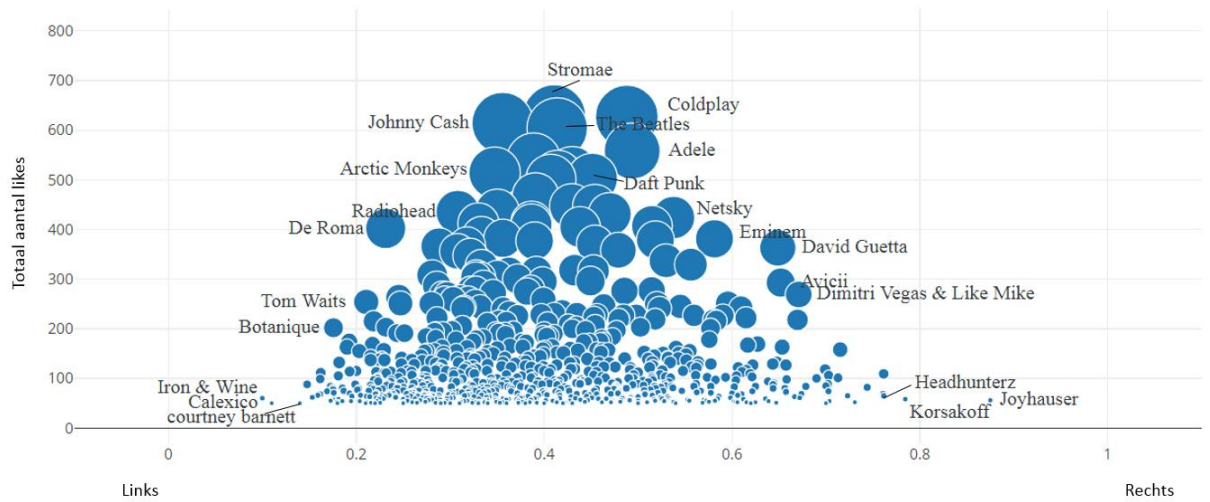


(c) Kunst en cultuur

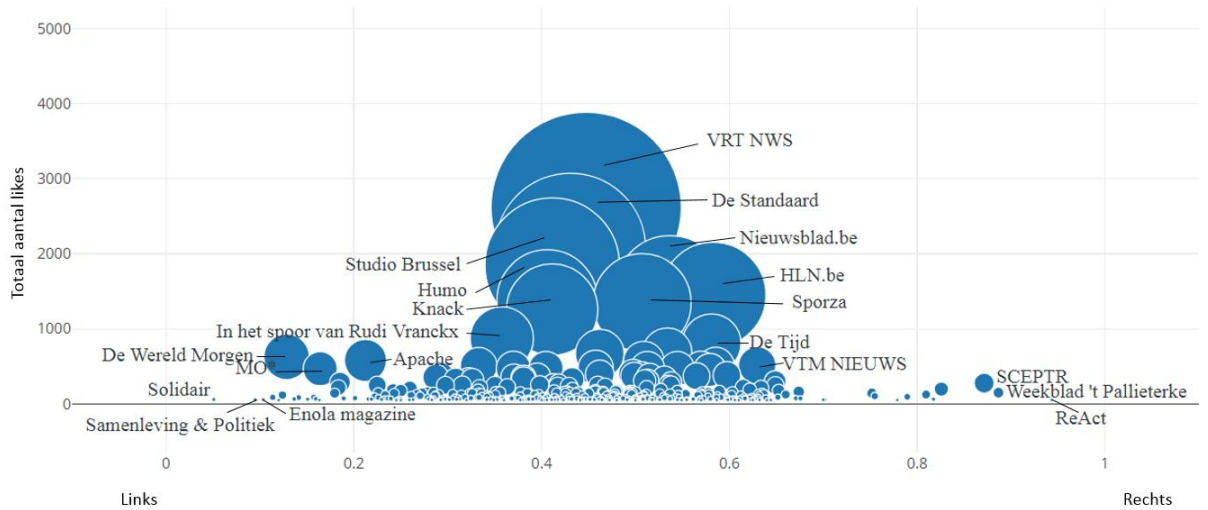


⁴ Enkele numerieke waarden van deze grafieken zijn terug te vinden op onze website: <https://www.uantwerpen.be/nl/projecten/nws-data/facebookstudie/>

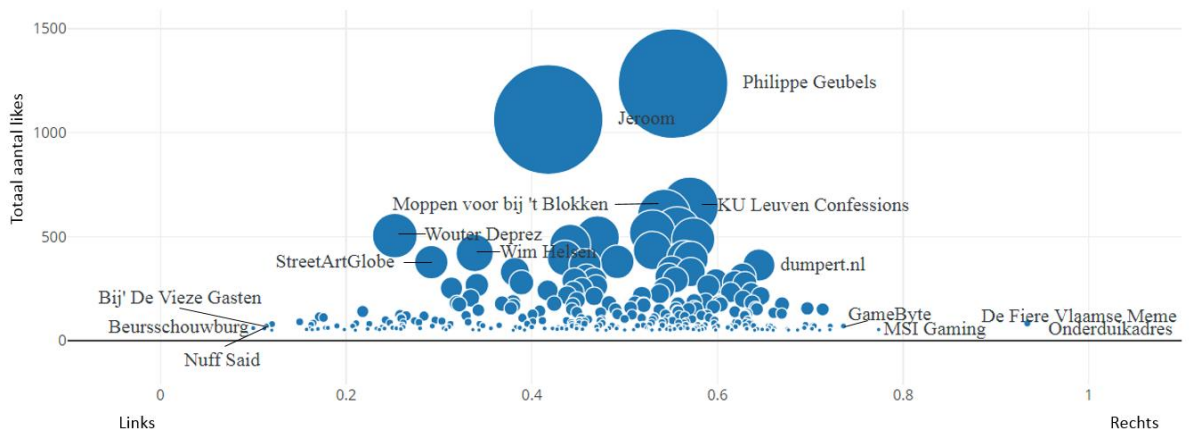
(d) Muziek



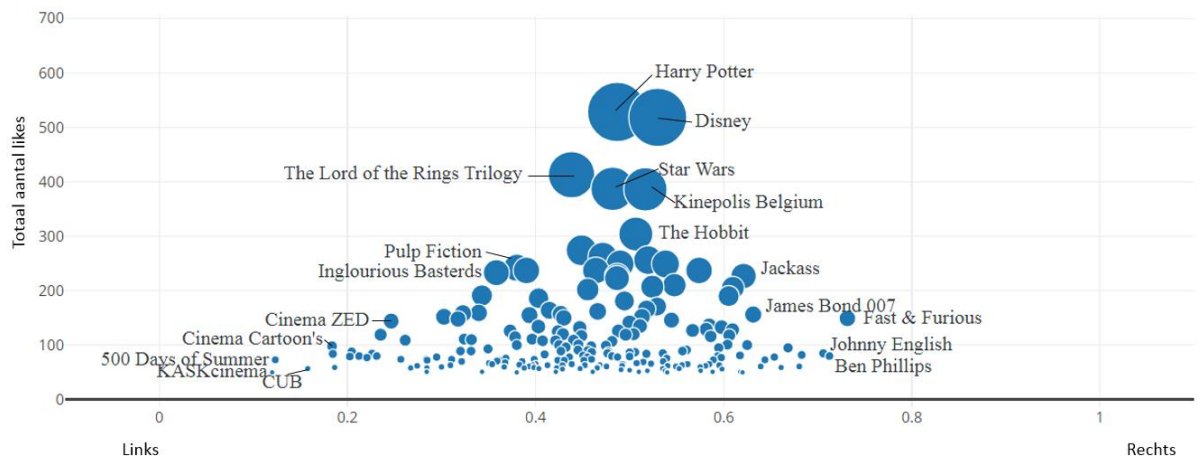
(e) Nieuws en media



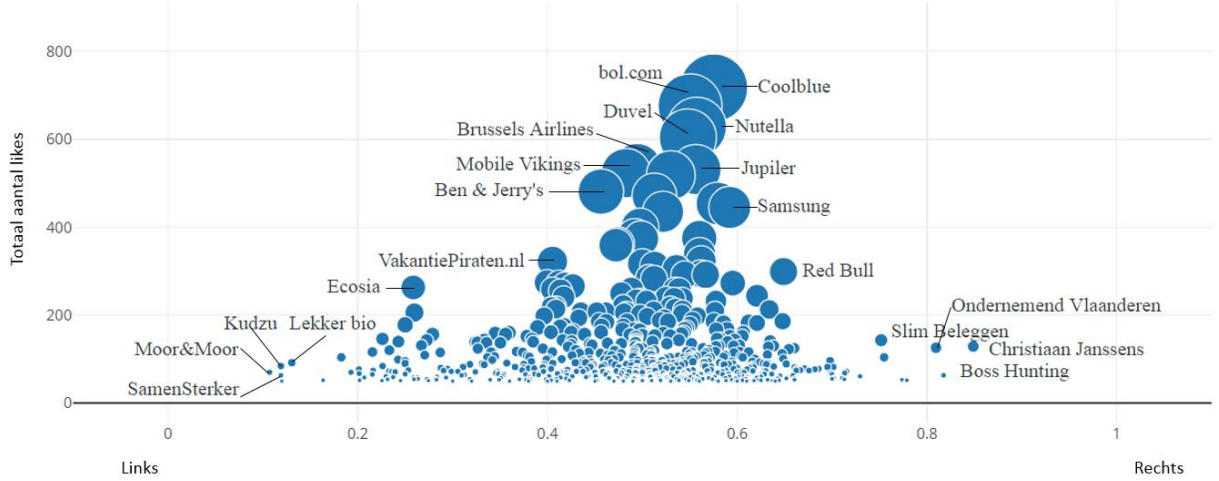
(f) Humor en entertainment



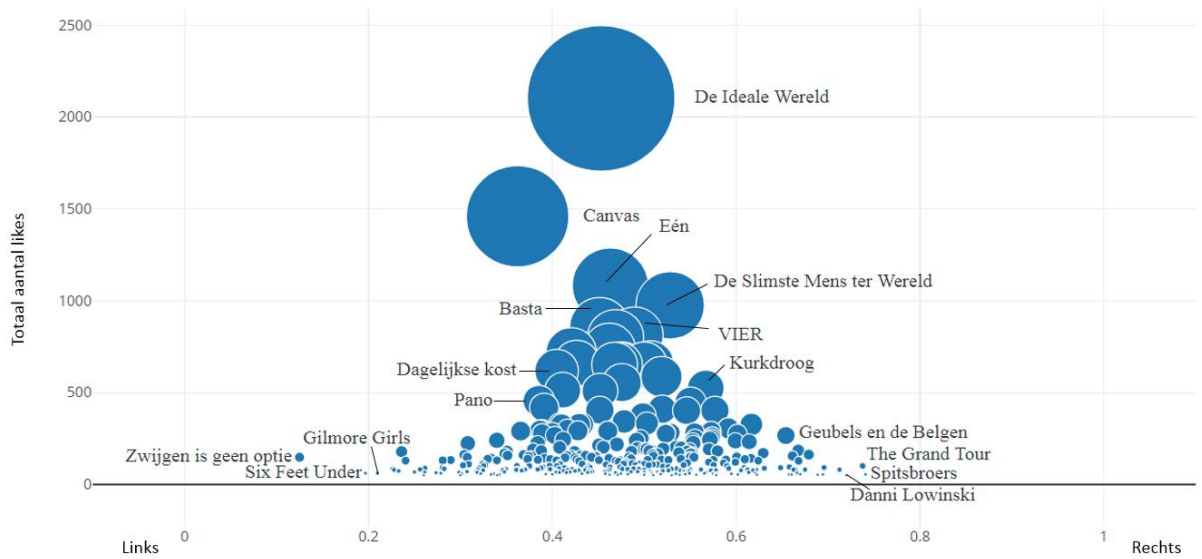
(g) Films en acteur



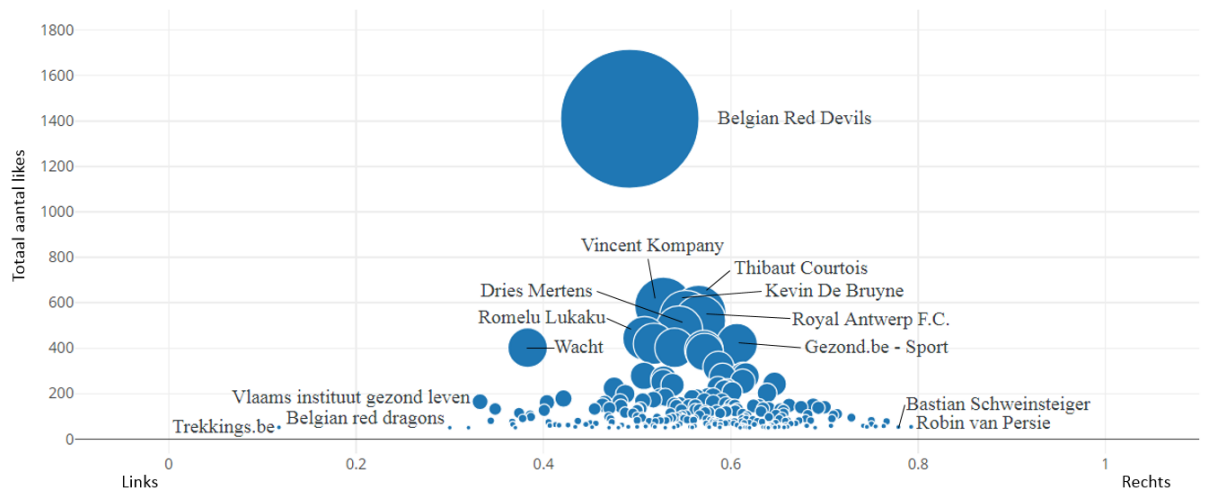
(h) Bedrijven en ondernemingen



(i) TV-programma's en omroepen



(j) Sport en gezondheid



De volledige paper (geschreven in het Engels) kan u hier terugvinden:

Praet, S., Van Aelst, P., & Martens, D. (2018). *I like, therefore I am: predictive modeling to gain insights in political preference in a multi-party system* (FBE Research Paper: RPS-2018-014).

Full text (open access): <https://repository.uantwerpen.be/docman/irua/e7ffcf/155787.pdf>