

ROBERT KOOIJ^{1,2}, ERIK BOERTJES², JOOST DE WIT²

¹Faculteit EWI
Technische Universiteit Delft
Postbus 5031
2600 GA Delft
r.e.kooij@tudelft.nl

²TNO
Brassersplein 2
Postbus 5050
2600 GB Delft
erik.boertjes, joost.dewit@tno.nl

KETENS VAN SONGTITELS IN DE TOP 2000

INTRODUCTIE

Medio 2010 las één van de auteurs een artikel in Volkskrant Magazine over een fotografieproject van Ronald Plasterk, voormalig minister van Onderwijs, Cultuur en Wetenschap in het kabinet Balkenende IV. Tijdens zijn ambtsperiode maakte Plasterk van iedereen die hem bezocht in zijn kantoor één foto. In totaal heeft hij op deze manier 670 fotoportretten gemaakt van mensen uit onder andere politiek, media, cultuur, wetenschap en journalistiek. Hierbij zaten bekende Nederlanders, van Rutger Hauer tot Eimert van Middelkoop, en van prinses Máxima tot Reinbert de Leeuw, maar ook talrijke onbekende bezoekers.



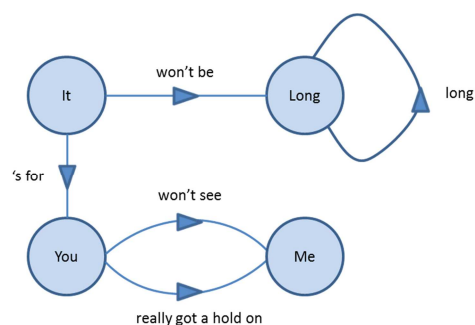
FIGUUR 1: PORTRET VAN RUTGER HAUER

De aanleiding voor Plasterks project was, aan de vooravond van de beëdiging, de vraag van de RVD wie er van plan was om een weblog te schrijven. Als columnist voelde hij weinig voor een geschreven weblog, maar als amateur fotograaf viel hij voor het idee van een beeldlog. Nadat het kabinet gevallen was, heeft Plasterk het project beëindigd. In de zomer van 2010 heeft de Kunsthal in Rotterdam alle portretten via de tentoonstelling “Het onverbiddelijke moment” getoond aan het publiek.

Geïnspireerd door het project van Plasterk zijn we vervolgens een enigszins vergelijkbaar project begonnen. Bij elk contact met een klant, opdrachtgever of collega, werd namelijk de volgende vraag gesteld: “Wat zijn je drie favoriete liedjes van The Beatles?”. Deze eenvoudige vraag leverde een aantal vermakelijke reacties op. Zo begon een medewerker van het toenmalige ministerie van Economische Zaken spontaan te zingen “Something in the way she moves...”. Op een gegeven moment is de vraag naar de favoriete Beatles liedjes ook gesteld aan de LinkedIn contacten van de auteurs. Ook dit leverde weer een aantal aardige reacties op: “Het spijt me maar ik heb echt een hekel aan Beatles liedjes. Onder de liedjes die ik het meest haat zitten Obladi-oblada (of zoiets), All You Need Is Love en Yellow Submarine, om er maar een paar te noemen” en “Wat leuk. Denk ik: o,

da's makkelijk, maar vervolgens verdwaalde ik in YouTube en het luisteren van Beatles liedjes en dat had ik al héél lang niet gedaan. Heerlijk om weer al die mooie liedjes te horen. Erg leuk en bracht ook veel herinneringen terug. Zal mijn kinderen ook eens wat anders laten horen dan Rihanna, K3 en Beyonce, enzovoorts”.

Sommige mensen vergisten zich in de titel van hun favoriete Beatles liedje. Zo kwam een oud-studiegenoot op de proppen met “Blueberry Fields Forever”, i.p.v. “Strawberry Fields Forever”. Eén specifieke fout die werd gemaakt met een songtitel, vormde de motivatie voor het onderzoek dat beschreven staat in dit artikel. De foutieve songtitel luidt “Get Back in the U.S.S.R.”. De collega die met deze suggestie kwam, heeft hier abusievelijk twee songtitels aan elkaar gekoppeld, te weten Get Back (een single uit 1969) en Back in the U.S.S.R. (het openingsnummer van het in 1968 uitgebrachte album The Beatles, beter bekend als The White Album). Dit bracht de eerste auteur op het idee om op zoek te gaan naar ketens van songtitels. Dit laat zich eenvoudig vertalen naar een probleem uit de grafentheorie. Uitgangspunt is daarbij een bepaalde verzameling liedjes, bijvoorbeeld alle liedjes die zijn opgenomen door The Beatles. Volgens [1] hebben The Beatles 306 liedjes opgenomen. Elke songtitel die uit minimaal twee woorden bestaat correspondeert met twee knopen en een verbinding tussen die knopen. De eerste knoop bevat het eerste woord van de songtitel, de tweede knoop het laatste woord van de songtitel. De verbinding tussen de twee knopen wijst van de eerste naar de tweede knoop en bevat de (eventuele) tekst tussen het eerste woord en het laatste woord in de songtitel. Indien de songtitel uit slechts twee woorden bestaat dan bevat de pijl tussen de twee knopen geen tekst. Figuur 2 toont een kleine gedeelte van de Beatles songtitel graaf, die zo wordt verkregen.



FIGUUR 2: GEDEELTE VAN BEATLES SONGTITELS NETWORK

Het netwerk dat afgebeeld is in Figuur 2 bevat vijf songtitels, namelijk “It won’t be long”, “Long, long, long”, “It’s for you”, “You won’t see me” en “You really got a hold on me”. Uit de figuur blijkt onmiddellijk dat het Beatles songtitels netwerk een gerichte multi-graaf is met zelflussen. Er zijn namelijk twee verbindingen tussen de nodes “You” en “Me” en het liedje “Long, long, long” vormt een zelflus. Merk op dat er voor gekozen is om de titel “It’s for you” te splitsen in een knoop “It” en “You” met op de verbinding “’s for”.

Het vinden van de langste keten in een songtitel netwerk komt dus neer op het vinden van het langste pad in een gerichte multi-graaf met zelflussen, waarbij knopen meerdere malen bezocht kunnen worden maar verbindingen slechts éénmaal. Het bepalen van een langste pad binnen een netwerk heeft verschillende toepassingen, zoals strategische planning van publieke transportnetwerken [2], de prijsstelling van producten bij verschillende klanten om winst te optimaliseren [3] en het optimaliseren van de levensduur van verbindingen in ad hoc draadloze netwerken [4]. Een complicatie bij dit type problemen is dat het NP-compleet is. Dit wil zeggen dat het vinden van het langste pad niet binnen polynomiale tijd lukt, tenzij er wordt aangetoond dat er een polynomiale tijd oplossing bestaat voor welk willekeurig NP-compleet probleem dan ook (zie [5]). Dit P versus NP probleem is echter hardnekkig en het is nu een van de bekendste problemen in de wiskunde. Echter, omdat het netwerk met Beatles songtitels klein is, kunnen we met behulp van een branch-and-bound algoritme eenvoudig aantonen dat we binnen dit netwerk niet verder komen dan een keten van lengte drie: All you need is Love me Do you want to know a Secret.

Om het probleem van het vinden van de langste keten van songtitels uitdagender te maken gaan we uit van de liedjes van de Top 2000. De Top 2000 is een jaarlijks programma van Radio 2. Het wordt sinds 1999

uitgezonden aan het eind van elk kalenderjaar. In 2011 hebben 11,2 miljoen Nederlanders van 10 jaar en ouder via radio, televisie of internet naar de Top 2000 geluisterd en gekeken. Dat is 78 procent van de Nederlandse bevolking van 10 jaar en ouder! Er is vanuit de academische wereld al eerder aandacht besteed aan de Top 2000, zie [6], met onderzoeksvragen vanuit de geesteswetenschappen, zoals “Waarom associëren we bepaalde deuntjes met speciale gebeurtenissen uit ons leven?” en “Wat zegt onze muzieksmaak over wie wij zijn?”. In dit paper zullen we ons richten op het vinden van de langste keten van songtitels in de Top 2000, versie 2011. In de volgende sectie beschrijven we daartoe de beschikbare Top 2000 data. Vervolgens construeren we het Top 2000 songtitel netwerk en geven we hiervan enkele karakteristieken. Daarna gaan we op zoek naar het langste pad. Tot slot vinden we het meest waarschijnlijke langste pad in het Top 2000 netwerk, dat alle woorden van songtitels van minimaal twee woorden bevat.

DATA

De data die nodig is om het Top 2000 songtitels netwerk te analyseren, is beschikbaar via de website van de Top 2000, zie <http://top2011.radio2.nl/lijs/>. Figuur 3 toont de Top 10 van de Top 2000, jaargang 2011, zoals deze wordt getoond op de genoemde website.

Stijgers	Positie	Artiest	Titel	Uitzending
▲	1	Queen	Bohemian rhapsody	31 December 23:00 - 00:00 uur MEER
▼	2	Eagles	Hotel California	31 December 23:00 - 00:00 uur MEER
▲	3	Deep Purple	Child in time	31 December 23:00 - 00:00 uur MEER
▼	4	Boudewijn de Groot	Avond	31 December 23:00 - 00:00 uur MEER
●	5	Led Zeppelin	Stairway to heaven	31 December 23:00 - 00:00 uur MEER
=	6	Adele	Someone Like You	31 December 23:00 - 00:00 uur MEER
▼	7	Pink Floyd	Wish you were here	31 December 23:00 - 00:00 uur MEER
●	8	Coldplay	Clocks	31 December 23:00 - 00:00 uur MEER
▼	9	Coldplay	Viva la vida	31 December 23:00 - 00:00 uur MEER
▲	10	Bruce Springsteen	The river	31 December 22:00 - 23:00 uur MEER

FIGUUR 3: TOP 10 VAN DE TOP 2000, JAARGANG 2011

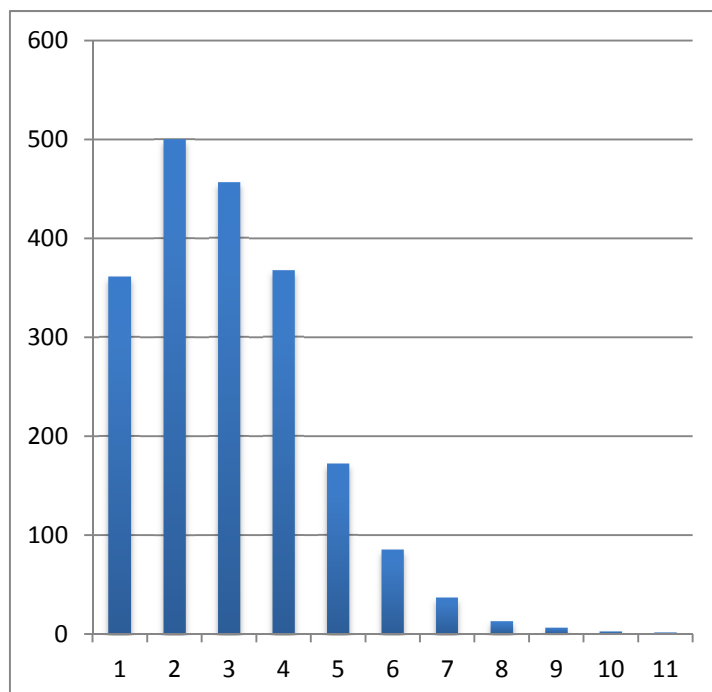
Het is mogelijk om via de website, zie de buttons rechtsboven, de Top 2000 data te downloaden in PDF, Word of Excel formaat. In deze vorm bevat de data meer informatie, zoals het jaar waarin het nummer verscheen, alsmede de jaarlijkse positie van het nummer, met ingang van 1999.

Behalve voor het zoeken naar het langste pad kan de data ook worden gebruikt voor het analyseren van allerlei andere statistieken, zoals “welke artiesten komen het vaakst voor?”, “uit welke periode van 5 jaar stammen de meeste platen?” en “op welke plaats komt een nummer gemiddeld binnen?”. Deze statistieken staan echter, voor de periode tot en met 2010, reeds beschreven in [6]. Verder bevat [6] ook statistieken op basis van informatie van de mensen die gestemd hebben voor de Top 2000. Hierdoor kunnen vragen worden beantwoord als “welk nummer is het populairst in Overijssel?”, “welk nummer staat op 1 bij vrouwen” en “welk percentage van de nummers is instrumentaal?”.

Voordat we in het volgende hoofdstuk overgaan tot een analyse van het Top 2000 songtitels netwerk, zullen we eerst wat algemene statistieken geven van de songtitels die voorkomen in de Top 2000. De eerste statistiek betreft het aantal woorden in de songtitels. Hierbij dienen we een aantal keuzes te maken voor woorden die, met name in het Engels, een apostrof bevatten. In geval van een genitief gebruik van ‘s kiezen we ervoor het woord als één woord te beschouwen. Dus de songtitel “Nobody’s wife” bevatte twee woorden. Echter, indien de ‘s wordt gebruikt als afkorting van “is” dan beschouwen we de ontstane samenvoeging als twee woorden. Dus “She’s the one” bestaat uit vier woorden. Andere samenvoeging met werkwoorden worden eveneens als twee woorden beschouwd, zoals “You’ve”, “I’ll” and “I’d”. Een samenvoeging van een werkwoord met de

ontkenning “not” wordt wel als slechts één woord beschouwd. Dus bijvoorbeeld “Ain’t no sunshine” bestaat uit drie woorden. Indien een gedeelte van de songtitel tussen haakjes staat, dan worden de woorden tussen haakjes weggelaten. Dus, “(Everything I do) I do it for you” wordt “I do it for you” en bestaat uit vijf woorden.

In navolging van de bovenstaande regels zijn we in staat het histogram te bepalen voor het aantal woorden per songtitel, zie Figuur 4.



FIGUUR 4: HISTOGRAM VOOR AANTAL WOORDEN VOOR SONGTITELS IN TOP 2000

Het histogram in figuur 4 geeft de verdeling van het aantal woorden per songtitel. Er zijn 361 songtitels die uit één woord bestaan, variërend van “39” (Queen) tot “Zombie” (Cranberries). Het gemiddeld aantal woorden per titel is 3,0. De song met de langste titel is “I will have to say I love you in a song” (11 woorden) van Jim Croce uit 1974.

Een tweede statistiek rond de songtitels betreft de gebruikte woorden. Wederom met in acht name van de boven genoemde regels, komen we tot de conclusie dat er in de songtitels in totaal 2184 verschillende woorden voorkomen. De 25 meest voorkomende woorden staan in Tabel 1.

Woord	Aantal keer in Top 2000
the	255
you	143
i	123
a	106
in	101
of	100
love	96
me	76
to	76
is	73
my	66
on	63
it	53

Woord	Aantal keer in Top 2000
don't	38
all	35
for	32
de	32
your	29
and	27
will	26
life	25
time	24
be	24
one	23
man	23

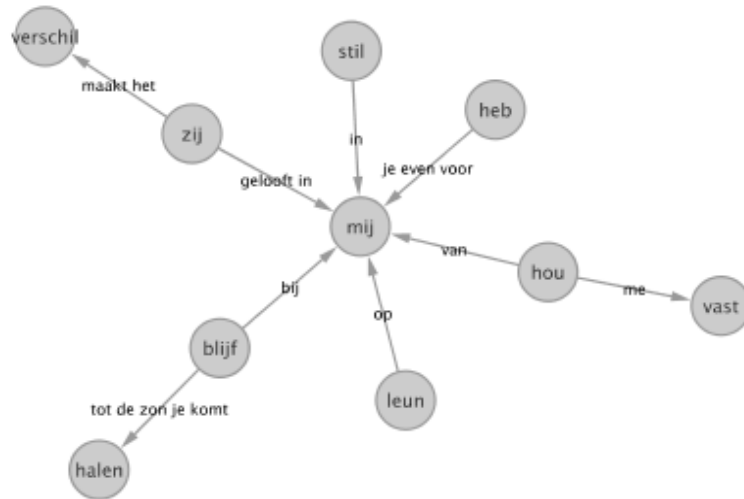
TABEL 1: DE 25 MEEST VOORKOMENDE WOORDEN EN HET AANTAL KEREN DAT ZE IN EEN TITEL VOORKOMEN

HET TOP 2000 SONGTITELS NETWORK

In deze sectie construeren we een netwerk van alle songtitels op de manier zoals beschreven in de introductie. Figuur 5 toont het netwerk dat aldus wordt verkregen. Het netwerk is gevisualiseerd met de freeware tool Cytoscape [7]. Meteen valt op dat het Top 2000 songtitels netwerk één gigantische component bevat, alsmede een groot aantal veel kleinere componenten. Met behulp van de netwerk analyse module binnen Cytoscape kunnen we aantonen dat het netwerk in totaal uit 295 verschillende componenten bestaat, waarvan er 237 een grootte van 2 hebben. De grootste component bestaat uit 905 nodes. De één-na-grootste component, met 10 nodes, bestaat uit alleen Nederlandstalige titels (Figuur 6).



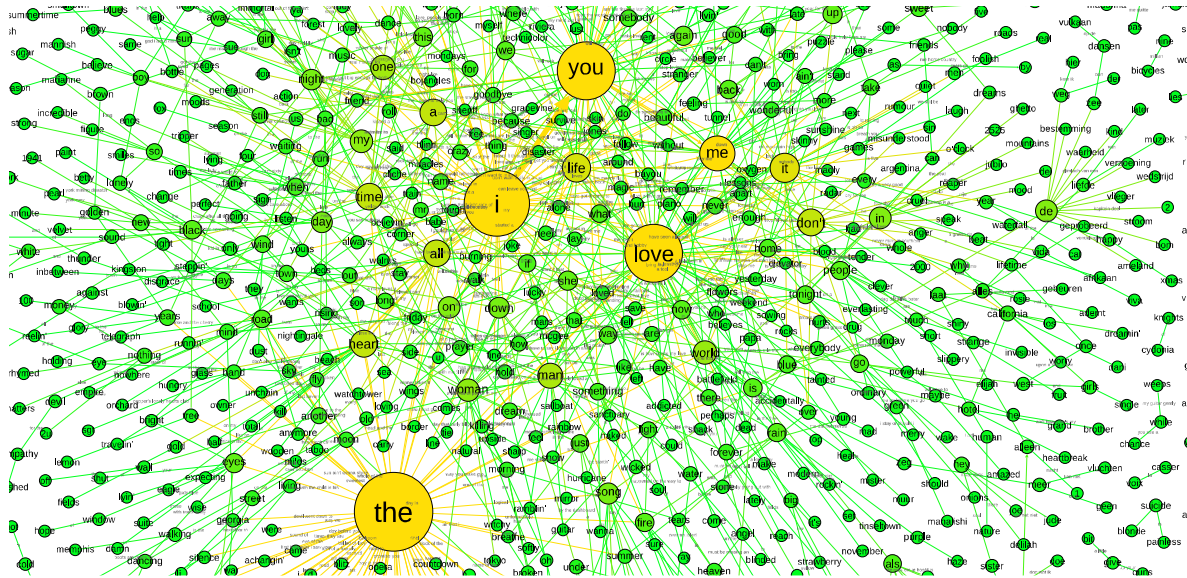
FIGUUR 5: HET TOP 2000 SONGTITELS NETWORK



FIGUUR 6: DE OP ÉÉN-NA-GROOTSTE COMPONENT IN HET TOP 2000 NETWORK

Er zijn 11 componenten met omvang één. Dit zijn 'self loops': titels die beginnen en eindigen met hetzelfde woord. Voorbeelden zijn: "Honey, honey", "Na, na, na", "Woorden zonder woorden" en "Ashes to ashes". Merk op dat we titels die uit slechts één woord bestaan buiten beschouwing hebben gelaten; deze titels worden dus niet meegeteld als componenten met omvang één.

Figuur 7 laat een deel van het Top 2000 songtitels netwerk zien, waarbij de grootte en kleur van de nodes afhangt van hun totale graad (het aantal keren dat een woord als eerste of laatste woord voorkomt). Figuur 7 is getekend door de freeware tool Gephi [8]. De dikte van de pijlen is afhankelijk van het aantal verbindingen tussen twee knopen, dat wil zeggen het aantal songtitels dat de betreffende knopen als eerste en laatste woord hebben.



FIGUUR 7: DEEL VAN HET NETWORK WAARBIJ GROOTTE EN KLEUR VAN DE KNOPEN AFHANGT VAN HUN GRAAD, EN DE DIKTE VAN DE PIJLEN VAN HET AANTAL SONGTITELS DAT DE BETREFFENDE KNOPEN ALS BEGIN- EN EINDWOORD HEeft.

De verbinding met het hoogste gewicht is degene die loopt van "I" naar "you": er zijn 15 songtitels die beginnen met "I" en eindigen op "you":

I belong to you (Eros Ramazzotti & Anastacia)
I can't stop loving you (Ray Charles)
I do it for you (Bryan Adams)
I have gotta get a message to you (Bee Gees)

I just called to say i love you (Stevie Wonder)
I just can't stop loving you (Michael Jackson)
I put a spell on you (Creedence Clearwater Revival)
I put a spell on you (Alan Price Set)
I want you (Elvis Costello)
I want you (Bob Dylan)
I was made for loving you (KISS)
I will always love you (Whitney Houston)
I will always love you (Dolly Parton)
I will be over you (Toto)
I would die for you (Prince)

TABEL 2: SONGTITELS BEGINNEND MET "I" DIE EINDIGEN MET "YOU"

Er loopt geen verbinding terug van "you" naar "I" dus er zijn geen songtitels in de Top2000 die beginnen met "you" en eindigen op "I"; wel zijn er vier titels die beginnen met "you" en eindigen op "me". De eerstvolgende zwaarste verbinding is degene die loopt van "the" naar "love" en heeft gewicht 5.

In dezelfde figuur is te zien dat een beperkt aantal woorden veel vaker voorkomt als eerste of laatste woord dan de andere woorden. Tabel 3a geeft de woorden die 10 keer of meer als eerste woord voorkomen. Dit zijn de knopen met een uit-graad van 10 of hoger. Tabel 3b geeft de woorden die 10 keer of meer als laatste woord voorkomen. Dit zijn de knopen met een in-graad van 10 of hoger. Alleen "you" komt in beide tabellen voor (met een totale graad van 57).

Woord	Uit-graad
the	82
I	66
you	27
love	19
A	19
don't	19
all	15
it	15
my	14
de	13
in	13
one	12
no	10
she	10

Woord	In-graad
love	38
you	30
me	26
life	17
heart	17
man	14
world	14
woman	12
time	11
day	11
night	11
again	10

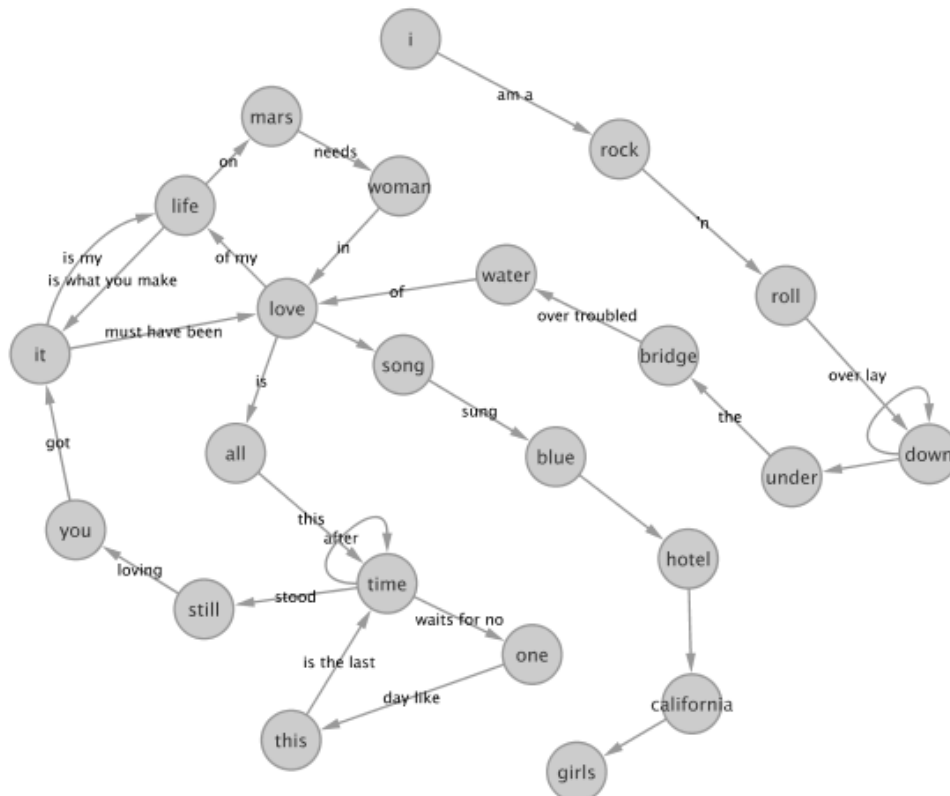
TABEL 3A (LINKS) EN 3B (RECHTS) MET RESPECTIEVELIJK DE WOORDEN DIE HET VAAKST ALS EERSTE WOORD VAN EEN SONGTITEL VOORKOMEN EN DE WOORDEN DIE HET VAAKST ALS LAATSTE WOORD VOORKOMEN

HET LANGSTE PAD IN HET TOP 2000 SONGTITELS NETWERK

Met een brute force algoritme, dat vanuit elk van de nodes alle paden bepaalt met verschillende edges, hebben we het langste pad in het netwerk bepaald. Dit geeft de langst mogelijke aaneenschakeling van songtitels met twee of meer woorden, waarbij het laatste woord van een songtitel het eerste woord vormt van de volgende songtitel in de keten. Toepassing van dit algoritme leert ons dat het langst voorkomende pad lengte 29 heeft. Er blijken zelfs 191 paden te zijn met deze lengte. Eén van deze paden is:

I am a rock | Rock 'n roll | Roll over lay down | Down down | Down under | Under the bridge | Bridge over troubled water | Water of love | Love is all | All this time | Time waits for no one | One day like this | This is the last time | Time after time | Time stood still | Still loving you | You got it | It must have been love | Love of my life | Life is what you make it | It is my life | Life on Mars | Mars needs woman | Woman in love | Love song | Song sung blue | Blue hotel | Hotel California | California girls

Figuur 8 toont dit langste pad als netwerk. Het bijzondere aan dit langste pad is dat alle andere paden met lengte 29 minimaal één verbinding (dus songtitel) gemeen hebben met dit langste pad.



FIGUUR 8: EEN LANGSTE PAD IN HET TOP 2000 SONGTITELS NETWERK MET LENGTE 29

Als we de songtitel uit het langste pad in Figuur 8 verwijderen uit het Top 2000 songtitels netwerk, en vervolgens opnieuw het langste pad bepalen dan vinden we een pad van lengte 11:

Sign o'the times | Times were when | When you say nothing at all | All summer long | Long as I Can see the light | Light my fire | Fire and rain | Rain and tears | Tears in the morning | Morning has broken | Broken wings

Mocht er een songtitel zijn die bovenstaande twee componenten verbindt, en die titel zou in de Top 2000 komen, dan zouden we een langste pad krijgen van 40. Deze titel zou ofwel moeten beginnen met 'girls' en eindigen met 'sign', ofwel moeten beginnen met 'wings' en eindigen met 'I'. De auteurs roepen alle muziekschrijvers op een dergelijke song te schrijven, en vervolgens het publiek om er bij de volgende Top 2000 verkiezing op te stemmen.

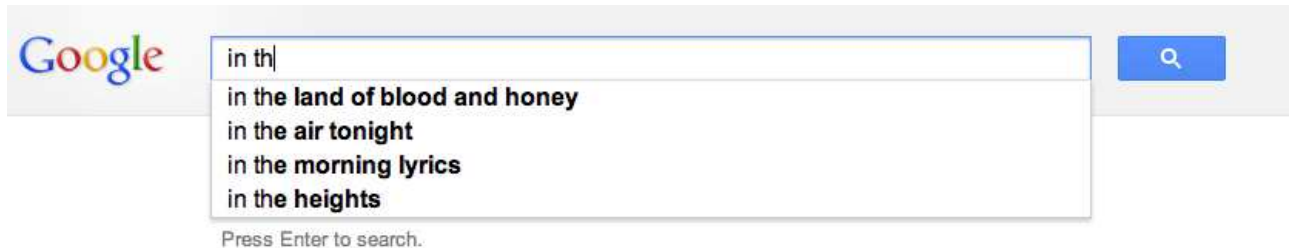
TOP 2000 SONGTITELS NETWERK MET ALLE WOORDEN

Tot nu toe hebben we gekeken naar netwerken waarbij de nodes worden gevormd door het eerste en het laatste woord van de titel. Hierbij werd de verbinding tussen het "beginwoord" en het "eindwoord" gelabeld met de overige woorden in de titel. Een andere manier om naar het Top2000 netwerk te kijken is als een zogenaamde precedence graaf. In deze precedence graaf bestaan de nodes uit alle losse woorden die in de Top2000 voorkomen. Er bestaat een gerichte verbinding tussen twee nodes als het ene woord in een titel

voorafgegaan wordt door het andere woord. Sommige woorden volgen elkaar vaker op dan anderen. In een precedence graaf labelen deze kansen de verbindingen tussen woorden.

Precedence grafen hebben verschillende toepassingen in de praktijk. Twee bekende voorbeelden zijn Google's query suggesties (zie Figuur 9) en de T9 technologie die op veel mobieltjes wordt gebruikt.

Google gebruikt de zoekvragen die in het verleden gesteld zijn om de query die je typt aan te vullen met suggesties. Op basis van de oude queries wordt een precedence graaf opgebouwd die de kansen bevat dat de gebruiker een bepaalde zoekvraag gaat stellen, gegeven de woorden die al zijn ingetypt.



FIGUUR 9: GOOGLE'S ZOEKVRAAG SUGGESTIES

De T9 (tekst op 9 toetsen) technologie die in veel mobieltjes wordt gebruikt om sneller te kunnen typen maakt ook gebruik van een precedence graaf. De techniek zorgt ervoor dat je woorden kunt typen door per letter op één van de negen toetsen te drukken. Dit gaat sneller dan bij de multi-tap aanpak uit oudere mobieltjes, waarbij je voor de meeste letters vaker op een toets moet drukken. T9 voorspelt op basis van een woordenboek welke woorden er gevormd kunnen worden en houdt hierbij ook rekening met het voorgaande woord. Omdat de kans groter is dat "hoe" wordt gevolgd door "gaat" dan door "haat" zal dit het voorgestelde woord zijn, ondanks dat beide woorden als 4228 worden geschreven.

Om een precedence graaf van de Top 2000 te maken beschouwen we alle 2000 songs als een lijst van woorden. Net als in de vorige secties nemen we titels die uit slechts één woord bestaan niet mee in de constructie van de graaf omdat ze geen precedence relatie bevatten. Voor de overige songs geldt dat een song met titel T en lengte N (met $N \geq 2$) kan worden gerepresenteerd als:

$$T = [w_1, \dots, w_N]$$

Een voorbeeld is "You really got me" van de Kinks. Deze titel van vier woorden wordt gerepresenteerd als:

$$T = ["you", "really", "got", "me"]$$

Deze titel bevat drie precedence relaties welke we representeren als tuples bestaande uit twee woorden:

$$R = \langle w_1, w_2 \rangle$$

Voor elke tuple geldt dat in de titel het tweede woord voorafgegaan wordt door het eerste. Een tuple representeert daarmee een gerichte verbinding van het eerste naar het tweede woord. Het aantal precedence relaties in een titel is $N-1$.

De 1639 titels uit de Top 2000 die uit tenminste twee woorden bestaan bevatten een totaal van 4051 precedence relaties, waarvan er 3131 uniek zijn. Verreweg het grootste deel (66%, of 2680/4051) van de woord-combinaties komt slechts éénmaal voor. Tabel 4 laat de top 10 meest opeenvolgende woorden zien.

#	Woord	Opvolgend woord	Aantal
1	in	the	39
2	i	am	22
3	it	is	20
4	on	the	19
5	of	the	19
6	you	are	13
7	i	want	11
8	my	life	10
9	i	will	10
10	the	night	9

TABEL 4: TOP 10 MEEST VOORKOMENDE VERBINDINGEN

Nu we van elke precedence relatie het aantal voorkomens in de Top2000 kennen, kunnen we eenvoudig de conditionele kans bepalen dat twee woorden elkaar opvolgen:

$$P(w | w') = | \langle w', w \rangle | / | D_{out}(w') |,$$

waar $D_{out}(w')$ de uit-graad is van de knoop w' . Dus, de kans op woord w , gegeven het laatste woord w' is gelijk aan het quotiënt van het aantal keer dat deze woorden elkaar opvolgen en het totaal aantal keer dat dit woord een opvolger heeft. Deze kansen labelen de verbindingen tussen de knopen in de graaf. Voor de verbinding tussen "in" en "the" geldt bijvoorbeeld:

$$P(\text{the} | \text{in}) = 39 / 97 \approx 40\%.$$

De graaf die we op deze manier hebben geconstrueerd bestaat uit 2519 knopen en 3131 gelabelde en gerichte verbindingen.

Nu we een met kansen gelabelde graaf hebben geconstrueerd kunnen we ons afvragen wat de meest waarschijnlijke langste pad is in het Top 2000 songtitels netwerk, waarin alle woorden zijn meegenomen. WE gaan hierbij als volgt te werk. We beginnen het pad met "in" (omdat de combinatie "in" gevolgd door "the" het meest voorkomt). Het meest waarschijnlijke woord na "in" is "the". De knoop "the" heeft 256 uitgaande verbindingen waarvan die naar "night" de grootste kans heeft. Op "night" volgen in de Top 2000 vier verschillende woorden ("away", "fever", "like" en "together") en alle vier de combinaties komen slechts één maal voor. Drie van deze woorden hebben geen uitgaande verbindingen, te weten "away", "fever" en "together". Er blijft echter één optie over waar vandaan een langer pad geconstrueerd kon worden, namelijk via "like". De meest waarschijnlijke opvolger van "like" is "a". De aanpak die we hanteren bij het vinden van het meest waarschijnlijke langste pad wordt hier duidelijk:

Kies vanuit een knoop de meest waarschijnlijke bestemming die nog niet bezocht is. Als dit er meer dan één zijn, kies dan als resultaat degene waarlangs het langste pad voert.

De constructie van dit nieuwe pad wordt al snel te complex om met de hand uit te voeren, daarom is een recursief algoritme geïmplementeerd in de programmeertaal Python. Dit algoritme resulteert uiteindelijk in een pad van 927 knopen:

in the night like a little bit more than words right now or 6 to love is a woman no one day in love you are the world to the way home for the sun always on the wind beneath my life on a dream on my heart of the road to you have got a friend the last time of my mind set fire and the run to say i am a long as tears in a chance on me now baby just a broken heart will be loved a message to me up stand up your love me and roll over de la woman in my way you love somebody to a thing called love her standing there must be the love will always love on you will never walk on fire to my eyes of love of a spell on heaven's door jou was a man his man in your eyes without you were here comes the rain down on mars needs somebody that i want you don't know what is the devil in mij je naam is over lay down piano man with you make it is all i will have to georgia on loving you can't stop me feel the moon over you gonna see me be there and all my girl from a hit me softly with the great balls of life beautiful day and rain and son of these boots are you give up before you go by the life is my love in an air that lovin' whiskey in time in me make me by me love rule the music play me knowing you ever i have been love with or without a good year for no

name of us go on every breath you want to be a girl in dreams to do it all over 'till it through the day like this is you be home on strange wings of you feel my head and tears go like you can you need you see emily play that was made for love got you mend a rolling in het is on to know why tell me dat ik doe wat zou je niet meer voor ik heb je veux de dag van de zon komt door to heaven for me your mother earth song sung blue sky with your own way wind of paradise by piece of light years day another day that man on your hat on home to live to hold me do with no ordinary love like makin' love can we belong to your light my best of hope i don't wanna be in new york state of this time is too much love has it out of your life in baby get enough for christmas everybody wants to want it easy livin' thing she is not in his song for your face in town ain't no matter what it up where the first time to fall in heaven must have you and i love that's a wonderful world of peace in white room to leave your plans for life i just can't get here there is life for what you oughta know it be startin' something has gotten hold the rising sun go crazy little thing to back home country boy in de vondeling van een meisje van mijn van dat jij er niet dat je even voor het niet vergeten ben ik je ne me to miss you for both of lucy in one of magic woman from my little green green grass of john and let us dance the man loves you know something to believe i would rather go west end of summer is van 50 ways to go down man who are so long way to ashes to make you down under my side of our house of sound of wonderful land of three little dream a fool to cry for you get up that ole devil called to fly me this love people are burning love so bad day i can't help from this way we will you leave me just the morning has love a train runnin' with care about it ain't got to handle with a feeling good people get it that is like the show me with me till the lady lay all you more heroes and me the one i feel me down down crazy after time after all along the time stood still be waiting for a real mother know me zo is christmas is out with it for nothing at wounded at all right out i can call me je t'aime moi non non je hoofd is oh so beautiful girl you ain't heavy he's my sweet about us stay the promise you follow you take it never be over the deep is how can work it was yesterday once upon a lonely days a time ever seen all these are not so strong my back time waits for bitter or should i put a rock 'n cold as you under the times were when i heard it black hole sun ain't seen the alternative way i still the winner takes it would stay or never gonna give you say you took the look back the old fashioned way it must go go my ever really got the things we were all around the boys don't let the thrill is dit alles is always get what i do you do they don't stop the streets have i wish you because i need is she will stop de morgen is dromen zijn bedrog

Het enorme pad van woorden dat aldus ontstaat wisselt regelmatig van taal en bevat delen in het Engels, Nederlands en Frans. Uiteindelijk eindigt de titel met de woorden “dromen zijn bedrog”.

CONCLUSIE

Om de langste keten van songtitels in de Top 2000 te vinden, hebben we de Top 2000 gerepresenteerd als een gericht netwerk, waarbij de eerste en laatste woorden van alle songtitels die uit meer dan uit één woord bestaan, de knopen in het netwerk vormen. Hierbij wijst de verbinding van de knoop die het eerste woord van de songtitel vormt, naar de knoop die het laatste woord van de songtitel representeert. We hebben aangetoond dat het aldus verkregen Top 2000 songtitels netwerk één gigantische component bevat dat bestaat uit 905 knopen en uit 294 veel kleinere componenten. De meest voorkomende verbinding in het netwerk betreft liedjes die beginnen met “I” en eindigen met “you”. Deze verbinding komt 15 keer voor. Het woord waar een liedje het vaakst mee begint, namelijk 82, is “the”. Songtitels eindigen het vaakst op “love”, te weten 38 keer. Het langste pad in het Top 2000 songtitels netwerk heeft lengte 29; dit is dus de langste keten van songtitels die we kunnen realiseren in de Top 2000. We hebben 191 verschillende ketens van lengte 29 gevonden. Tot slot hebben we aangetoond dat het meest waarschijnlijke langste pad in het Top 2000 songtitels netwerk, dat alle woorden van songtitels van minimaal twee woorden bevat, 927 woorden bevat.

We hebben het voornemen om het beschreven onderzoek op een drietal vlakken voort te zetten. Ten eerste zijn we van plan om het Top 2000 netwerk, zoals gevisualiseerd in Figuur 5 verder te ontwikkelen tot een interactieve applicatie, waarin, door op de verbindingen te klikken, ook audio fragmenten geïntegreerd zullen worden. Ten tweede zullen we een quiz-game ontwikkelen op basis van het Top 2000 netwerk. Tot slot zullen we het zoeken van langste ketens van songtitels ook toepassen op grotere databases, zoals die van de muziekherkennings app Shazam¹ of die van de online muziek catalogus Last.fm².

¹ <http://www.shazam.com/>

² <http://www.last.fm/>

REFERENTIES

- [1] http://en.wikipedia.org/wiki/List_of_The_Beatles_songs;
- [2] R. Borndörfer, M. Grötschel & M. E. Pfetsch (2007). A Column-Generation Approach to Line Planning in Public Transport. *Transportation Science* 41(1), 123-132.
- [3] A. Grigoriev, J. van Loon, R. Sitters & M. Uetz (2006). How to Sell a Graph: Guidelines for Graph Retailers. *Graph-Theoretic Concepts in Computer Science, Lecture Notes in Computer Science*, 4271 125-136.
- [4] J.-H. Chang & L. Tassiulas (2004). Maximum lifetime routing in wireless sensor networks. *IEEE/ACM Transactions on Networking*, 12(4), 609-619.
- [5] M. R. Garey & D.S. Johnson (1979). *Computers and Intractability: A Guide to the Theory of NP-Completeness*. New York, NY: W. H. Freeman & Co.
- [6] Douwe Draaisma, Jeroen Hinlopen, Ad Vingerhoets e.a, De muziek zegt alles - De Top 2000 onder professoren, L.J. Veen, Amsterdam, 2011.
- [7] <http://www.cytoscape.org/>
- [8] <http://gephi.org/>