



## UvA-DARE (Digital Academic Repository)

### De UvA webklas wiskunde

Brandts, J.

**Publication date**

2010

**Document Version**

Submitted manuscript

**Published in**

Euclides

[Link to publication](#)

**Citation for published version (APA):**

Brandts, J. (2010). De UvA webklas wiskunde. *Euclides*, 85(4), 145-148.  
<https://www.nvww.nl/19294/euclides/euclides-85-4.html>

**General rights**

It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

**Disclaimer/Complaints regulations**

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: <https://uba.uva.nl/en/contact>, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

# De UvA Webklas Wiskunde: Google's PageRank voor Dummies

Jan Brandts, Universiteit van Amsterdam

4 november 2009

## Samenvatting

Het Google PageRank probleem houdt zich bezig met de vraag hoe Google de resultaten van een zoekopdracht ordent: welke bladzijde staat er bovenaan in de lijst, en waarom. Het achterliggende model is wiskundig van aard en bevat graaftheorie, lineaire algebra, en numerieke wiskunde. Niet direct dagelijkse kost voor een middelbare scholier, en ook menig docent zal zich de fijnere details van die vakgebieden niet meer al te duidelijk herinneren. Maar dit hoeft ook niet. Als je de moeite neemt om precies na te gaan welke gereedschappen uit de diverse disciplines je nodig hebt, blijkt dat een groot deel van het succesverhaal rondom de PageRank zich laat vertalen in puzzels, eenvoudige diagrammen, en raadsels met kabouters en feetjes. De Webklas Wiskunde over Google's PageRank [1] laat leerlingen uit de bovenbouw van het VWO kennismaken met deze moeilijke wiskunde in een prettig ogende verpakking. Logisch redeneren staat centraal, rekenen is bijzaak.

## 1 Inleiding

Dit artikel heeft niet als doel om u volledig wegwijst te maken in de wiskunde die ten grondslag ligt aan de definitie en bepaling van de PageRank van Google. Mocht u daarin geïnteresseerd zijn, verwijs ik u naar de literatuur [2, 3, 4, 5], en naar de Webklas Wiskunde over Google's PageRank van de Universiteit van Amsterdam [1]. In dit stuk gaat het om de vertaalslag naar de wiskundige wereld van een leerling uit de bovenbouw van het VWO van de volgende, nogal specialistische, wiskundige tekst.

### Google's PageRank - definitie en eigenschappen

*De PageRank vector  $P$  is een eigenvector behorende bij de enkelvoudige, dominante eigenwaarde één van de positieve stochastische matrix  $G$ , die een convexe combinatie is van de connectiviteitsmatrix  $S$  van het World Wide Web, en een positieve rank-one update  $T = \frac{1}{n}ee^*$ , waarbij  $e$  de bij  $P$  horende linkereigenvector van  $G$  is, en  $n$  de lengte van  $e$ .*

*De Perron-Frobenius theorie voor positieve matrices toont de existentie aan van  $P$ , en laat bovendien zien dat  $P$  positief gekozen kan worden. De Machtsmethode voor de berekening van de dominante eigenvector van  $G$  convergeert naar  $P$  met een snelheid evenredig aan het relatieve gewicht van  $T$  in  $G$ .*

Als u zich wat ongemakkelijk voelt bij deze technische omschrijving, leest u dan vooral verder. Van een selectie van de voorkomende wiskundige begrippen zullen we namelijk laten zien dat

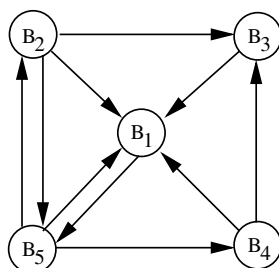
ze ook kunnen worden uitgelegd aan een goede bovenbouw VWO leerling met voldoende motivatie, middels raadsels en puzzels. Deze selectie is de volgende:

- eigenvectoren; verpakt als kabouter-raadsel,
- lineaire afbeeldingen en matrices; in termen van Sudoku-achtige Google puzzles,
- matrices met positieve enties (Perron Frobenius theorie); Goozzles met teleportatie,
- iteratieve methoden voor benadering van de eigendata van een matrix; herhaald Goozzelen.

Mocht u de vertaling van het gehele bovenstaande stukje willen lezen, dan kunt u kiezen uit referentie [3] die geschreven is voor docenten wiskunde, of de inhoud van de Webklas Wiskunde via de site [1] doornemen. Hierin worden echt alle bewijzen van bovenstaande beweringen stuk voor stuk vertaald tot een niveau dat voor een goede leerling te begrijpen is.

## 2 Eigenvectoren als kabouter-raadsels

Stel, er zijn vijf kabouters die we even  $B_1, B_2, B_3, B_4$  en  $B_5$  noemen. Ieder van deze kabouters heeft vrienden voor wie hij alles over heeft. In onderstaand plaatje kan je zien wie een vriend van wie is: de pijl van  $B_2$  naar  $B_1$  betekent dat  $B_1$  door  $B_2$  als vriend wordt beschouwd. De kabouters wijzen dus in Figuur 1 met pijlen hun vrienden aan.



**Figuur 1.** Niet-wederzijdse vriendschapsrelaties.

Behalve kabouters is er ook een goede fee, in het bezit van enorme hoeveelheden knikkers, die ze graag aan de kabouters wil uitdelen. Er is echter een probleem: zodra een kabouter knikkers krijgt, zal hij ze eerlijk verdelen over zijn vrienden. Zo zitten kabouters nu eenmaal in elkaar.



**Figuur 2.** Typische goede fee met knikkers.

De fee besluit daarom elke kabouter een zodanig aantal knikkers te geven (minstens één), dat nadat iedere kabouter zijn knikkers heeft verdeeld over zijn vrienden, ze allemaal weer net zoveel knikkers hadden als ervóór. Maar kan ze dat wel? En zo ja, hoe?

Omdat kabouter  $B_1$  iets krijgt van  $B_2$  tot en met  $B_5$ , en  $B_2$  iets krijgt van  $B_5$ , enzovoorts, zien we dat er 5 lineaire vergelijkingen zijn voor de 5 hoeveelheden knikkers  $P_1, \dots, P_5$ . Deze vergelijkingen kunnen natuurlijk in matrix-notatie in compacte vorm worden opgeschreven:

$$\begin{bmatrix} 0 & \frac{1}{3} & 1 & \frac{1}{2} & \frac{1}{3} \\ 0 & 0 & 0 & 0 & \frac{1}{3} \\ 0 & \frac{1}{3} & 0 & \frac{1}{2} & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{3} \\ 1 & \frac{1}{3} & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} P_1 \\ P_2 \\ P_3 \\ P_4 \\ P_5 \end{bmatrix} = \begin{bmatrix} P_1 \\ P_2 \\ P_3 \\ P_4 \\ P_5 \end{bmatrix}. \quad (1)$$

We zien dus dat de oplossing van het probleem van de fee een eigenvector is van de matrix  $H$  in bovenstaande vergelijking, immers, de oplossingsvector wordt door  $H$  op zichzelf afgebeeld. De bijbehorende eigenwaarde is dan gelijk aan één.

**Opmerking:** Uiteraard staan de kabouters voor webbladzijden en de pijlen voor hyperlinks tussen die bladzijden. Google definieert een webbladzijde als belangrijk als deze, in kaboutertermen, veel knikkers heeft. Duidelijk is dat een kabouter veel knikkers heeft als een andere kabouter met veel knikkers hem als vriend aanwijst, en deze verder weinig andere kabouters als vriend beschouwt. In webtermen: je bladzijde is belangrijk als belangrijke bladzijden jou als autoriteit aanwijzen, in plaats van je te noemen in een lange lijst van mogelijke bronnen.

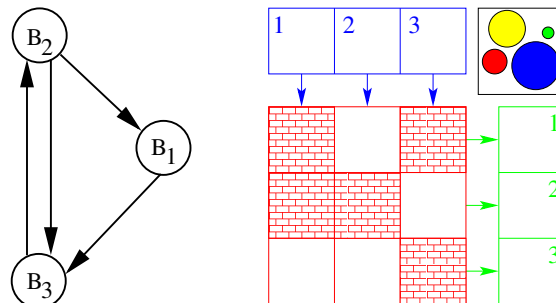
Het blijkt dat VWO leerlingen weinig moeite hebben om dergelijke puzzels te begrijpen en op te lossen, mits er eerst enkele eenvoudigere voorbeelden worden gegeven dan het  $5 \times 5$  voorbeeld hierboven. Ook het verband met webbladzijden en de heuristieken rondom het belang van een webbladzijde worden als schappelijk beschouwd. Op dit moment hebben ze echter nog geen matrix of vector gezien.

### 3 Goozles: lineaire afbeeldingen verkleed als Google Puzzle

Een alternatieve grafische voorstelling van de kaboutertraadsels, Goozle genaamd, is de volgende. Het idee is:

- Plaats knikkers in bovenste drie vakjes,
- Verdeel ze eerlijk over de lege vakjes er verticaal onder,
- Verplaats ze horizontaal naar de rechter drie vakjes.

Als de hoeveelheden knikkers in de rechtervakjes nu hetzelfde zijn als waarmee bovenin werd begonnen, is de oplossing gevonden. Bekijk bijvoorbeeld het volgende eenvoudige  $3 \times 3$  puzzeltje:



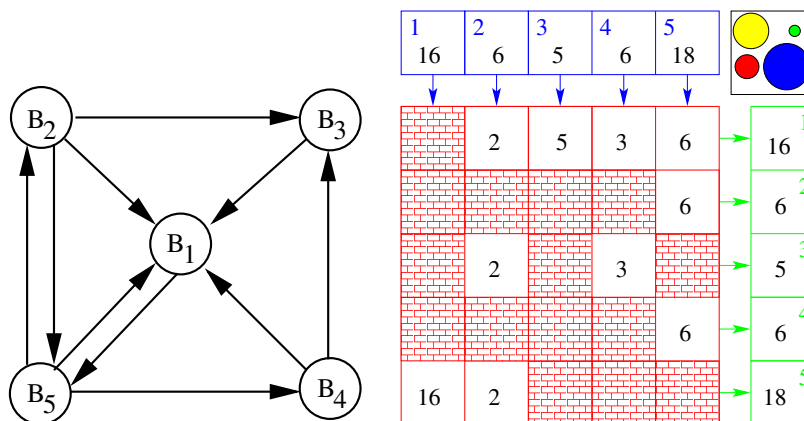
**Figuur 3.** Klein kaboutertraadsel met bijbehorende Goozle.

De representatie als Goozle lijkt iets overzichtelijker dan de representatie als graaf. Het weggeven van knikkers gaat verticaal, het ontvangen horizontaal.

Een belangrijk aspect van Goozles is, dat ze het resultaat zijn van een vertaalslag. Een puzzel met kabouters, waarvan inmiddels duidelijk is dat ze een internet(je) symboliseren, maar waarin de pijlen in principe alle kanten op kunnen wijzen, wordt herschreven in termen van een vierkant met volle en lege hokjes, een rijtje inputhokjes, en een rijtje outputhokjes. In feite staat hier een matrix, waarvan de gevulde vakjes een nul bevatten, en de open vakjes, impliciet, de breuk  $1/k$ , waarbij  $k$  het aantal open vakjes uit de betreffende kolom van het vierkant is. Het gaat richting wiskunde, maar vooralsnog lijkt het nog meer op een Sudoku. En als zodanig blijken Goozles een redelijke aantrekkingskracht te hebben op de gemiddelde leek, waaronder jongere kinderen en ook ouderen die niets met wiskunde hebben.

**Opmerking:** Ook belangrijk is om in eerste instantie in gehele getallen te blijven rekenen, zodat de gehele Goozle doorlopen kan worden zonder een knikker in stukken te hoeven breken. Behalve dat dit gemakkelijker puzzelen is, is ook het existentiebewijs van een oplossing van het PageRankprobleem in het uiteindelijke model relatief eenvoudig (zie [1]). In plaats van de dekpuntstelling van Brouwer op een compacte verzameling toe te passen, krijgen we te maken met een iteratie op een eindige verzameling. En dat is VWO leerlingen prima uit te leggen. Zie de website [1] voor verdere details over dit aspect.

Duidelijk is dat de PageRank  $P_j$  een  $k$ -voud moet zijn als er  $k$  lege vakjes in de  $j$ -de verticale kolom van de Goozle staan. Gebruik makend van dit soort inzichten (en ook van het feit dat  $P_2$  en  $P_4$  direct volgen uit  $P_5$ ) kan ook het oorspronkelijke kabouterprobleem uit Figuur 1 redelijk snel worden opgelost.



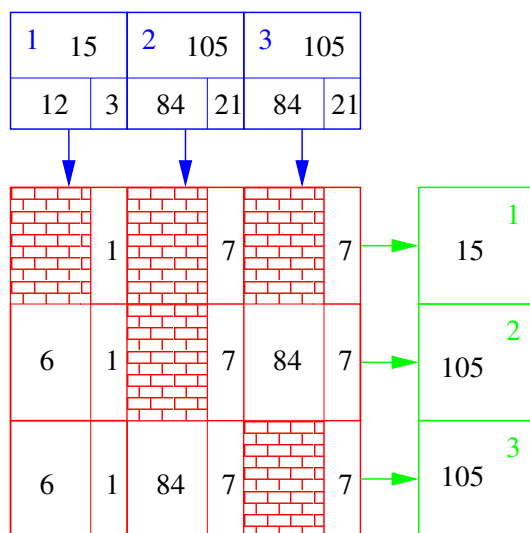
**Figuur 4.** Goozle behorende bij het kaboutertraadsel uit Figuur 1, met oplossing.

Puzzelend, of Goozzelend, zijn ook twee belangrijke eigenschappen van matrixrekening uit te leggen. Allereerst dat veelvoudigen van oplossingen (eigenvectoren) ook weer oplossingen zijn. En, door bijvoorbeeld met twee kleuren knikkers te werken, dat de output van een Goozle lineair afhangt van de input. De verschillende kleuren knikkers vinden namelijk onafhankelijk van elkaar de weg door de Goozle. Dit is van belang bij het bewijzen van de stelling dat er geen twee verschillende (niet elkaars veelvoud zijnde) oplossingen van een Goozle bestaan. Oftewel, dat de eigenruimte van de eigenwaarde één, ééndimensionaal is, en dus dat het PageRank probleem niet ineens met twee heel verschillende ranglijsten voor het belang van de webbladzijden op de proppen komt.

In de UvA Webklas Wiskunde [1] wordt, naast de manipulaties met Goozzles, ook wat lineaire algebra gegeven, maar dan in de hoedanigheid van een notatie voor Goozzles en hun oplossingen. Dit beperkt zich tot de matrix-vector vermenigvuldiging en de lineariteit ervan.

## 4 Goozzles met teleportatie: positive lineaire algebra

Het model zoals boven gesuggereerd, is nog niet het definitieve model. Er zijn eenvoudige voorbeelden te geven van Goozzles die alleen de nuloplossing hebben, en zelfs van Goozzles die inderdaad twee oplossingen hebben, die niet elkaars veelvoud zijn. Zonder in detail te gaan wat de oorzaak is van deze tekortkomingen (zie [1, 3] voor een uitgebreide beschrijving) geven we hier een voorbeeld van het uiteindelijke type Goozzle, nadat alle problemen de wereld uit zijn geholpen. Het wordt wel wat lastiger puzzelen.



**Figuur 5.** Goozzle met teleportatie.

Deze Goozzle kan worden gezien als een  $3 \times 3$  Goozzle van het eerdere type, waarbij we de doorstroom van knikkers garanderen door ieder van de drie kolommen een fractie  $1/5$  open te zetten. Dit is hier overigens omwille van de duidelijkheid van het plaatje niet in de juiste verhouding getekend! Hoe dan ook, knikkers uit een inputvakje moeten eerst in de verhouding  $4 : 1$  verdeeld worden over de twee kleinere inputvakjes (een soort voorselectie), en pas daarna worden ze eerlijk verdeeld over de open vakjes er recht onder. En weer wordt alles horizontaal naar de vakjes aan de rechterkant verplaatst. De vraag is hetzelfde gebleven, namelijk, om een stationaire oplossing te vinden.

**Opmerking:** Het openzetten van iedere kolom komt in het web overeen met het toevoegen van de mogelijkheid om vanuit een webbladzijde, met een bescheiden kans, naar een andere bladzijde te surfen, ook als daar geen hyperlink naar wijst. Dit kan bijvoorbeeld door in de browserbalk een nieuw adres in te tikken, en dit gebeurt natuurlijk ook geregeld. Het zich verplaatsen naar een andere webbladzijde zonder daarbij een hyperlink te hebben gevolgd wordt teleportatie genoemd.

We zijn nu beland in het vakgebied van de positieve matrices: Perron-Frobenius theorie. Door iedere kolom een klein beetje open te zetten, bevat de bijbehorende matrix geen nullen

meer. Een van de centrale stellingen uit het vakgebied, die in de Webklas Wiskunde [1] ook daadwerkelijk bewezen wordt in termen van Goozzles, is dat de dominante eigenvector positief gekozen kan worden, wat voor het concept PageRank natuurlijk wel van belang is. Immers, hoe vergelijk je een PageRank van 1 en  $-2$ ? Deze vraag hoeft gelukkig niet te worden beantwoord.

Wat direct aan de puzzel te zien is, is dat de oplossing niet kan bestaan uit én nullen, én positieve getallen. Immers, een positieve hoeveelheid in één van de inputvakjes verdeelt zich in het rechterdeel van de bijbehorende kolom over alle outputvakjes. Dit eenvoudige resultaat is al een stap in de goede richting voor de iets lastigere stellingen in deze context.

## 5 Stijgende ondergrenzen en de Machtsmethode

Als laatste voorbeeld van hoe een Goozzle kan helpen bij de uitleg van een wiskundig begrip kijken we nogmaals naar de Goozzle uit Figuur 5. Uitgaande van het feit dat er een oplossing bestaat in positieve gehele getallen die ook gehele getallen oplevert in alle vakjes van de Goozzle, kan deze gevonden worden middels een iteratie. Het idee is als volgt.

Als er een oplossing bestaat die overal in de Goozzle positieve gehele getallen heeft, kunnen we een eerste ondergrens afleiden voor de oplossing. Immers, teneinde gebroken knikkers te vermijden moet de input van de Goozzle uit Figuur 5 minimaal 15, 15, 15 zijn. Deze getallen zijn dus een (eerste) ondergrens voor de gezochte oplossing. Een tweede volgt door deze 15, 15, 15 knikkers volgens de Goozzle regels naar de outputvakjes te brengen. Het resultaat is 3, 21, 21. Ook deze drie getallen zijn ondergrenzen voor de gezochte oplossing. Voor twee van de drie vakjes levert dit zelfs een verbeterde ondergrens op: combinatie van de gegevens geeft als beste ondergrens 15, 21, 21. Zoveel knikkers zijn er dus minimaal nodig. Maar omdat 21 niet deelbaar is door 15, en deelbaarheid door 15 noodzakelijk is om gebroken knikkers te vermijden, kunnen we de ondergrenzen weer ophogen tot 15, 30, 30. Deze hoeveelheden brengen we weer volgens de Goozzle regels naar de outputvakjes, enzovoorts.

Even nadenken over dit algoritme laat zien dat het convergeert naar de kleinste oplossing in positieve gehele getallen, die ook gehele getallen oplevert in alle vakjes van de Goozzle. Een monotone begrensde rij gehele getallen convergeert, en dat zelfs in eindig veel stappen. En, niet onbelangrijk, de limiet is inderdaad een oplossing.

Vervolgens kan met wat meer werk ook worden aangetoond dat dit proces convergeert zonder de aanname van de existentie, en ook in dit geval kan worden laten zien dat de limiet een oplossing is. En dus levert dat de existentie van die oplossing op. Een wonderlijke constructie, maar wel zonder moeilijke wiskundige begrippen, allemaal vanwege het discrete karakter van gemakkelijk te begrijpen puzzels.

In de praktijk wordt de dominante eigenvector (de oplossing van de Goozzle) van de stochastische Google matrix  $G$  (van de Goozzle) bepaald door diezelfde iteratie, uiteraard in termen van lineaire algebra, genaamde de Machtsmethode. De Machtsmethode is niets anders dan het herhaald toepassen van de Goozzle op een gegeven beginsituatie, waarbij breuken wel worden toegestaan en de afrondstap dus ontbreekt. Gegeven de Google matrix  $G$  benadert de Machtsmethode de dominante eigenvector (dus de PageRank vector) van  $G$  middels de eenvoudige iteratie  $x_{k+1} = Gx_k$ , waarbij  $x_0 \in \mathbb{R}^n$  een willekeurig gekozen startvector is. Convergentie van de machtsmethode, voor een niet-symmetrische matrix zoals  $G$ , kan worden

bewezen [6] met behulp van een analyse van de Jordan normaalvorm van  $G$ . Maar daar kan je natuurlijk in 5 VWO niet mee aankomen. Het mooie is dat bovenstaande variant met gehele getallen en afronding naar bepaalde veelvouden een veel eenvoudiger convergentiebewijs heeft, zonder dat dit ten koste gaat van de convergentiesnelheid.

## 6 Conclusies

In een toepassing zoals de PageRank van Google wordt vaak gebruik gemaakt van pittige wiskunde en stellingen met moeilijke bewijzen. De reden dat die bewijzen vaak moeilijk zijn, is dat stellingen vaak zo breed en algemeen toepasbaar worden geformuleerd. Dat is de kracht van de wiskunde. Aan de andere kant, de toepassing zelf heeft vaak bij lange na niet die algemeenheid nodig, en is in die context eigenlijk slechts een speciaal geval. Om die reden kan het goed mogelijk zijn om de wiskundige algemeenheid weer teniet te doen ten gunste van de begrijpelijkheid ervan. Dit is precies wat de UvA Webklas Wiskunde over Google's PageRank heeft geprobeerd te doen. Gevolg is dat nu ook bovenbouw VWO leerlingen kunnen meegeenieten van de slimme truuks die Larry Page en Sergey Brin bedachten om hun zoekmachine zo optimaal mogelijk te laten functioneren.

Hieronder de mening van Dorien Lugt die in 2008 één van de deelnemers aan de webklas was en inmiddels wiskunde studeert, terwijl ze dat vóór de webklas zelfs nog nooit overwogen had.

*In het voorjaar van 2008 heb ik de wiskundewebklas gevolgd aan de UvA. Toen ik besloot een webklas te gaan volgen dacht ik niet meteen aan wiskunde. Ik bekeek het aanbod van webklassen en wiskunde leek me wel interessant. Het onderwerp van de webklas was: Google's PageRank. Ik had daar nog nooit van gehoord, en ik werd nieuwsgierig naar de manier waarop Google bepaalt welke zoekresultaten het belangrijkste zijn.*

*Aan het begin wist ik niet goed wat ik moest verwachten. Wiskunde, dat klinkt zo moeilijk. Maar toen ik aan de opdrachten begon, bleek het heel erg leuk! De eerste week was redelijk makkelijk, later moesten we ingewikkeldere dingen doen, zoals zelf algoritmes (her)schrijven.*

*Dit soort wiskunde kende ik helemaal niet van school. Je moest vooral zelf manieren bedenken om problemen op te lossen. Gelukkig kon je met je medewebklassers discussiëren over de opdrachten op het forum. Daar kon je ook vragen stellen aan de webklasdocenten.*

*Het leuke van de webklas vond ik vooral dat alles wat je leerde ook echt wordt toegepast in de realiteit. De vraag die ik op school nog wel eens heb bij wiskunde: 'Waar gebruik je dit in hemelsnaam voor?', was hier dus helemaal niet nodig. Ik vond het zelfs zo interessant, dat ik besloten heb dit als onderwerp voor mijn profielwerkstuk te gebruiken.*

*Aan het einde van de webklas werd er een middag georganiseerd waar alle deelnemers aan de wiskundewebklas naar toe mochten als ze wilden. Best leuk om de personen, die je alleen via het forum hebt gesproken, in het echt te zien.*

*Ik vond het ontzettend leuk om via deze webklas met een 'ander soort' wiskunde in contact te komen. Ik denk dat het voor mensen die erover denken om wiskunde te gaan studeren een hele goede kennismaking is. Maar ook als je daar helemaal niet aan denkt is aan te raden een wiskunde webklas te volgen. Het is gewoon leuk!*



Jan Brandts (1968) promoveerde begin 1995 aan de Universiteit Utrecht en werkte vervolgens in Bristol, Jyväskylä, Kiel, Praag, en Sydney alvorens in 1999 als KNAW Fellow terug te keren aan de Universiteit Utrecht. Sinds 2002 heeft hij een vaste aanstelling aan het Korteweg-de Vries Instituut voor Wiskunde van de Universiteit van Amsterdam, alwaar hij onderzoek doet in de (Numerieke) Lineaire Algebra en de Computatieve Meetkunde, en onderwijs geeft in zowel de Bachelor als de (Nationale) Master. Daarnaast is hij betrokken bij outreach-activiteiten als Proefstudereren, Leve de Wiskunde, en de Webklas Wiskunde [1] van de UvA.

## Referenties

- [1] J.H. Brandts (2008). UvA Webklas Wiskunde: de PageRank van Google in 12 uur. Via <http://staff.science.uva.nl/~brandts>
- [2] J.H. Brandts (2007). Google's PageRank: de grootste matrixberekening ooit. *De Nieuwe Wiskrant*, 27(2):8–12.
- [3] J.H. Brandts (2008). De wiskunde die Google groot maakte. *De Uitwisseling*, 24(4):8–18.
- [4] J.H. Brandts (2008). Google danst. *Pythagoras*, 48(1):8–12.
- [5] A.N. Langville and C.D. Meyer (2006) Google's PageRank and Beyond: the science of search engine rankings. *Princeton University Press*, Princeton and Oxford.
- [6] Wikipedia. Power Iteration. [http://en.wikipedia.org/wiki/Power\\_iteration](http://en.wikipedia.org/wiki/Power_iteration)