**KAARTKLEUREN: Een minimaal aantal kleuren**

Casimir Boot

Universiteit van Amsterdam

[casboot@gmail.com](mailto:emailadres@email.uva.nl)

Lonneke Lammers

Universiteit van Amsterdam

[lonnekelammers@gmail.com](mailto:lonnekelammers@gmail.com)

Alwin Lijdsman

Universiteit van Amsterdam

[alwinlijdsman@outlook.com](mailto:alwinlijdsman@outlook.com)

**1. Inleiding**

Het minimaliseren van het aantal kleuren wanneer een kaart wordt ingekleurd is een bekend probleem. In 1852 vermoedde Guthrie dat, wanneer elk land wordt ingekleurd waarbij aangrenzende landen niet dezelfde kleur mogen hebben, er maximaal vier kleuren nodig zijn [2]. Dit wordt het Guthrie’s Problem genoemd [2].

Het eerste wetenschappelijke artikel over dit probleem werd geschreven in 1878. In 1976 werd Gurthrie’s probleem opgelost door Kenneth Appel en Wolfgang Haken [1], hiermee werd de Four Color Theorem ontwikkeld [1]. Het was de eerste grote stelling die bewezen werd met behulp van een computer [1].

In dit verslag wordt de Four Color Theorem getoetst met drie verschillende kaarten, een kaart van Pennsylvania, een kaart van een Indiase provincie en een kaart van een sociaal netwerk. Het doel is om met zo min mogelijk kleuren de kaarten in te kleuren. De hoofdvraag van dit verslag luidt daarom als volgt: hoeveel kleuren zijn er minimaal nodig om een kaart in te kleuren? De uitkomt hiervan zou volgens de Four Color Theorem vier moeten zijn. Dit zal daarom getoetst worden.



*Figuur 1: Kaart van Penssylvania en kaart van een Indiase provincie*

Om de hoofdvraag te beantwoorden zijn in dit verslag twee heuristieken gebruikt die beide tot een correct antwoord komen. Dit om een beter begrip te krijgen van het probleem van het kaartkleuren.

De bruto toestandsruimtegrootte in dit verslag wordt geschat op Kn, waar K gelijk is aan het aantal kleuren en n gelijk is aan het aantal landen op de kaart. Als ieder land op de kaart een andere kleur toegewezen krijgt, is er altijd een oplossing. De restrictie op de transitie is, zoals eerder gesteld, dat aangrenzende landen nooit dezelfde kleur mogen kennen en dat het minimaal aantal kleuren moet worden gebruikt.

Een sociaal netwerk inkleuren zonder dat personen die een connectie hebben een zelfde kleur krijgen is in principe vergelijkbaar met het inkleuren van een kaart. Echter wordt dit probleem complexer naarmate je meerdere netwerken aan elkaar wil koppelen waarbij niet elke persoon aangesloten is in ieder netwerk. In dit verslag is een sociaal netwerk van drie platformen ingekleurd (Facebook, Twitter en Myspace) met 100 anonieme gebruikers. Hierbij zijn sommige personen slechts lid een netwerk, sommigen van twee en een aantal van alledrie.

**2. Methodes**

In deze paragraaf wordt besproken welke methodes zijn gebruikt om te komen tot een antwoord op de hoofdvraag. Om het minimaal aantal kleuren voor het kleuren van een kaart te vinden wordt er in dit verslag gewerkt met een random sampling algoritme en een zelf ontwikkeld gericht algoritme. Als eerste is het random sampling algoritme ontwikkeld besproken. Daarna wordt het gerichte algoritme besproken.

*2.1 Random sampling algoritme*

Het random sampling algoritme is als eerste ontwikkeld omdat dit algoritme relatief gemakkelijk te ontwikkelen was. Het algoritme werkt als volgt: als eerste worden twee CSV-bestanden ingelezen met data over de kaart die geanalyseerd wordt. De CSV-bestanden worden met de hand gemaakt, omdat deze verschillend zijn voor iedere kaart. De kaarten die hiervoor zijn gebruikt zijn de kaart van Pennsylvania, een Amerikaanse staat, een regio van India en aantal verbindingen op een sociaal netwerk. Bij de kaart van het sociale netwerk, zijn alle drie de netwerken in elkaar gevoegd. Hierdoor ontstaat er een driedimensionale kaart, want naast “buurlanden” die naast een “land” kunnen liggen, liggen landen nu ook boven en onder elkaar.

In het eerste CSV-bestand worden de namen en te gebruiken afkortingen van de landen opgeschreven. Bijvoorbeeld: ”cra,crawford”. In het tweede CSV-bestand worden vervolgens de buurlanden aangegeven. Bijvoorbeeld: “cra,eri,law,lyo”. Dit betekent dat naast de regio Crawford, de landen Erie, Lawrence en Lycoming liggen. Door deze gemaakte CSV-bestanden weet het algoritme welke landen aan welke andere landen grenzen. De lijst met landen wordt door elkaar geschud, waardoor er willekeurig wordt begonnen met het kleuren van een land. Dit land wordt als eerste gekleurd en krijgt kleur “1”.

Vervolgens wordt er gekeken naar het tweede land in de lijst. Als dit land grenst aan het eerste land, krijgt dit land de kleur “2”. Zo wordt de gehele lijst afgegaan. Uiteindelijk wordt er gekeken naar wat de hoogste “kleur” was in de kaart. Als er vier kleuren zijn gebruikt is de hoogste waarde “4”.

Het inkleuren van de kaart wordt 1000 maal herhaald. Hieruit komen verschillende maximale waarden. De minimale waarde van deze lijst van maxima, vormt het antwoord op de hoofdvraag. Daarnaast worden alle maxima in een CSV-bestand gestopt verdere statistische berekeningen mogelijk te maken.

*2.2 Gericht algoritme: land/persoon met meeste connecties als eerste kleuren*

Ondanks het feit dat er goede antwoorden kwamen uit het random sampling algoritme, is er gekozen om een gericht algoritme te ontwikkelen. Dit om te onderzoeken of er een betere methode is om tot het juiste antwoord te komen.

Het gerichte algoritme heeft dezelfde basis als het random sampling algoritme. Het wijkt af van het random sampling algoritme bij het bepalen van de volgorde van de te kleuren landen. Waar het random algoritme willekeurig alle ingelezen landen door elkaar schudt, zoekt het gerichte algoritme eerst naar de landen met de meeste aangrenzende buurlanden en zet deze vervolgens vooraan in de lijst om gekleurd te worden. Landen die een gelijk aantal buren hebben worden wel in een willekeurige volgorde naast elkaar gezet.

Ook dit algoritme wordt 1000 maal herhaald. Die maximale waarden worden geteld en in een CSV-bestand gestopt voor verdere statistische analyse.

Een sociaal netwerk inkleuren zonder dat personen die een connectie hebben een zelfde kleur krijgen is in principe vergelijkbaar met het inkleuren van een kaart. Echter wordt dit probleem complexer naarmate je meerdere netwerken aan elkaar wil koppelen waarbij niet elke persoon aangesloten is in ieder netwerk. In dit verslag is een sociaal netwerk van drie platformen ingekleurd (Facebook, Twitter en Myspace) met 100 anonieme gebuikers. Hierbij zijn sommige personen slechts lid een netwerk, sommigen van twee en een aantal van alledrie.

**3. Resultaten**

In deze paragraaf wordt besproken welke resultaten er uit het onderzoek naar de meeste kaartkleuren is gekomen. De uitkomsten worden besproken per kaart, waarbij er een vergelijking wordt gemaakt tussen het random sampling algoritme en het gerichte algoritme. Onder statistieken wordt in dit verslag gezien: het antwoord op de hoofdvraag, oftewel het minimaal aantal kleuren, de gevonden maximale waarden en de gemiddelde waarden.

*3.1 Kaart van Pennsylvania*

Als eerste is de kaart van Pennsylvania geanalyseerd. Bij het random sampling algoritme kwam hier een minimum aantal kleuren van vier uit. Dit aantal is overeenkomstig met de Four Color Theorem. Opvallend is wel dat er bij random sampling opvallend weinig een optimale oplossing van vier kleuren uitkomt in tegenstelling tot het tweede, het gerichte kleur algoritme. Bij de random sampling komt er bij een herhaling van 1000 maal 1 keer een uitkomst van vier kleuren uit, 661 keer een uitkomst van een vijf-kleurige kaart en 238 keer een uitkomst van een zes kleuren.

Het gerichte kleur algoritme geeft bij een herhaling van 1000 maal 150 keer een uitkomst  om de kaart in te kleuren met vier kleuren en 850 keer een kaart met vijf kleuren.

*3.2 Kaart van provincie in India*

Als tweede is de kaart van een provincie in India geanalyseerd. Opnieuw zijn beide algoritmes 1000 maal gerund en op basis van de resultaten zijn onderstaande histogrammen tot stand gekomen.

Uit de histogram van random sampling valt op te maken dat de kaart slechts enkele keren met vier of zeven kleuren kan worden ingekleurd. Zes kleuren komt vaker voor, maar het meeste aantal keren werd het probleem met vijf kleuren opgelost.   
 Als er gericht gekleurd werd, zijn de resultaten anders. Alle 1000 maal werd de kaart met vier kleuren ingekleurd. Het is echter niet zo dat er 1000 maal dezelfde oplossing gevonden werd. Uit de code valt op te maken dat landen steeds verschillende kleuren toegewezen krijgen, wat dus betekent dat het algoritme elke keer een andere oplossing vindt.

*3.2 Sociaal Netwerk*

Een sociaal netwerk visueel maken is een een iets complexer probleem. Het is complexer omdat, zoals eerder vermeldt, personen in een sociaal netwerk drie dimensionaal connecties kunnen hebben. Als het algoritme 1000 keer gedraaid wordt kunnen de volgende histogrammen worden gemaakt.

Bij het random selecteren wordt er 103 keer met 5 kleuren gekleurd, 775 keer met 6 kleuren en 122 keer met 7 kleuren. Als er gericht gekleurd wordt verschilt het resultaat aanzienlijk. 947 keer heeft het algoritme genoeg aan 5 kleuren, 53 keer aan 6 kleuren en het hoefde nooit 7 kleuren te gebruiken.

**4. Conclusies**

Als beide algoritmes vergeleken worden kunnen enkele interessante conclusies getrokken worden. Het random sampling algoritme is ondanks dat het een sneller algoritme is een minder geschikte manier om een kaart te kleuren. Dit is het geval omdat het random sampling algoritme minder vaak tot een oplossing komt met het minimaal aantal kleuren dan het gericht kleuren algoritme. Dit blijkt zowel het geval te zijn met de kaart van Pennsylvania als met de kaart van India. In beide gevallen werd met het gerichte algoritme vaker een oplossing gevonden waar 4 kleuren werden gebruikt. Volgens het Four Color Theorem moet dit bij elke twee dimensionale kaart kunnen. Het lijkt dus effect te hebben op de oplossing door te beginnen met kleuren bij de landen met de meeste aangrenzende landen.

Ook met het inkleuren van de sociale netwerken blijkt het beter om met het gerichte kleuren algoritme te werk te gaan.

Hetzelfde geldt voor het inkleuren van een sociaal netwerk. Echter gaat hier de Four Color Theorem niet op, omdat het een driedimensionale kaart betreft.

Je kunt dit probleem nader onder de loep nemen. Zo zou verder onderzoek wellicht tot een ontdekking kunnen leiden dat andere aspecten ook effect zouden kunnen hebben op de effectiviteit van het kleuren van een kaart. Zoals bijvoorbeeld aantal “landenpunten”. Dit zou je mogelijk met de heuristiek hill climber kunnen proberen.

Een nadeel aan de algoritmes in dit verslag is is dat er een kans is op gelijke oplossingen, vooral in het gericht zoeken. Er is geen formele check op gelijke oplossingen, maar bij het analyseren van de resultaten lijken er geen dubbele oplossingen voor te komen.

Er kan geconcludeerd worden dat het gericht kleuren voor de kaarten van Pennsylvania, India en het sociale netwerk veel efficienter is in vergelijking met random sampling.

**5. Referenties**

[1] [http://mathworld.wolfram.com/Four-ColorTheorem.html](http://mathworld.wolfram.com/Four-ColorTheorem.html" \t "_blank)

[2] [http://mathworld.wolfram.com/GuthriesProblem.html](http://mathworld.wolfram.com/GuthriesProblem.html" \t "_blank)

<http://heuristieken.nl/wiki/index.php?title=Kaartkleuren>