DOOR MAAIKE HOOGEBOOM, DAAN BAKKER
EN WOUTER SLOT

Hulp bij ontrafelen fraude: netwerkanalyse

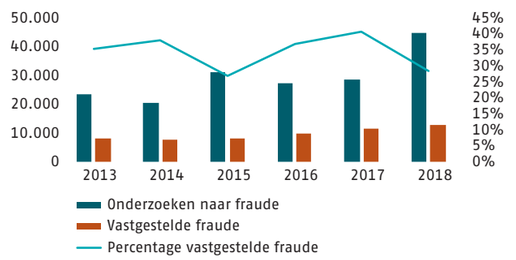
Op 16 december 2019 werd Nederlands meest gezochte crimineel Ridouan T. opgepakt na meer dan een jaar intensief politiewerk. Het in kaart brengen van zijn netwerk heeft hierbij een belangrijke rol gespeeld. Afluisterpraktijken en een infiltratie binnen zijn netwerk hebben er uiteindelijk toe geleid dat hij in Dubai is aangehouden. Ook voor een verzekeringsmaatschappij is het nuttig om netwerken te onderzoeken. Natuurlijk niet om klanten af te luisteren, maar wel om ongewenst gedrag zoals fraude op te sporen.

Fraude opsporen wordt steeds moeilijker. Vanuit schade-experts komen meer signalen over netwerken van klanten, schadeherstellers en andere partijen met frauduleuze praktijken. Bijvoorbeeld een netwerk waarin veelvuldig onderdelen van een auto in een andere auto zijn geplaatst en als gestolen zijn opgegeven of een scooterherstelbedrijf dat standaard klanten aanspoort om ook een kapotte telefoon en jas mee te claimen. Daarnaast worden klanten onzichtbaarder. Vroeger kende de verzekeraar de klant persoonlijk, maar tegenwoordig sluiten meer en meer klanten een verzekering via internet. Als iemand zich onder een valse naam wil verzekeren, is dat hierdoor onder bepaalde omstandig-heden een stuk gemakkelijker.

FRAUDE OPSPOREN

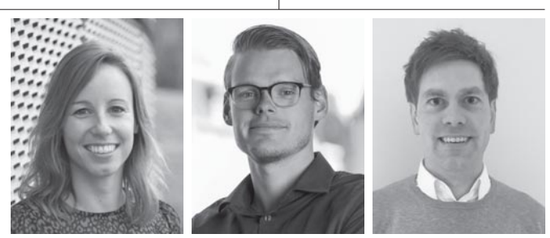
Traditionele (veelal) deterministische fraudedetectietechnieken zijn sterk afhankelijk van menselijke verwachtingen. Fraude-experts gebruiken hun ervaring en hun vermogen om patronen te identificeren. Veel huidige fraudemodellen voorspellen op basis van eerder gevonden fraude. Zo'n 'supervised' model presteert alleen goed als er al veel fraude is bewezen. Dit is vaak niet het geval.

De grotere hoeveelheid beschikbare data, tezamen met de sterk toegenomen computerrekenkracht, maken nieuwe, geavanceerdere en meer datagedreven fraudeopsporingstechnieken mogelijk. 'Unsupervised' learning wordt als toevoeging op de reeds bestaande methoden ingezet. Deze technieken zijn beter in staat om onbekende en complexere patronen te detecteren die niet per se aan bestaande verwachtingen voldoen. De verzekeringssector experimenteert hier volop mee. Dit is waarschijnlijk een belangrijke reden dat het aantal onderzochte claims fors is toegenomen, zie ook onderstaande afbeelding. Echter blijft het aantal vastgestelde fraudeclaims achter en laat het percentage vastgestelde fraude zelfs een daling zien (van 41% in 2017 naar 29% in 2018).



Bron: Verbond van Verzekeraars (https://www.verzekeraars.nl/publicaties/actueel/recordaantal-verzekeringsfraudeurs-opgespoord)

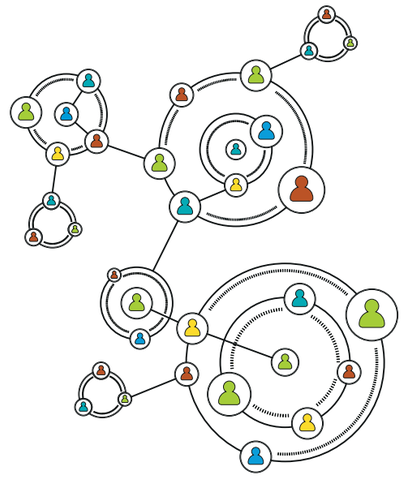
Dr. M. Hoogeboom (links) is Data Scientist, D. Bakker (midden) is Data Scientist en W.W. Slot MSc AAG is Manager Data Science, allen werkzaam bij Achmea.



Het inzetten van netwerkanalyse is naar onze mening een veelbelovende manier om bovenstaande trend weer de juiste kant op te buigen. Met een netwerkalgoritme kunnen verdachte patronen en relaties achterhaald worden. Binnen de bancaire sector is de techniek al succesvol ingezet om verdachte transacties op te sporen[1].

HOE WERKT EEN NETWERKANALYSE?

Netwerken komen in veel gebieden voor, denk bijvoorbeeld aan sociale netwerken waarbij vriendschappen tussen personen weergeven worden of transportnetwerken van alle vluchten of treintrajecten. Een netwerk is een verzameling van objecten of personen waartussen een connectie bestaat. Deze objecten en personen worden 'nodes' genoemd en worden weergeven als punten. De connecties tussen de nodes worden weergeven door een lijn, dit noemen we een 'edge'.



Bij het opsporen van fraude in een netwerk zijn personen, bedrijven of objecten weergegeven als nodes. Connecties tussen deze nodes komen tot stand doordat ze in dezelfde claim voorkomen, maar een koppeling kan ook plaatsvinden op basis van andere relaties zoals adres, IBAN of telefoonnummer. Uiteraard fungeert de AVG als leidraad voor het gebruiken van deze persoonlijke gegevens.

Geavanceerde netwerk algoritmes kunnen worden gebruikt binnen het fraudeonderzoek. In een netwerkanalyse wordt gekeken naar statistische eigenschappen van grootschalige netwerken. De techniek maakt deel uit van het onderzoeksgebied rondom de graaftheorie. Grote techbedrijven als Facebook, LinkedIn en Google ontwikkelen en gebruiken deze netwerkmodellen. Page Rank is een algoritme van Google gebaseerd op netwerkanalyse om te bepalen hoe belangrijk een website is. De Page Rank van een pagina is de kans dat een persoon die random surft op het internet uiteindelijk uitkomt bij deze pagina. Bij het opsporen van fraude kan men gebruik maken van variant hierop genaamd Personalized Page Rank[2]. In dit algoritme staat de set met bewezen frauduleuze spelers centraal. Een random walk begint in een random fraude node uit deze set van frauduleuze spelers. Elke iteratie kiest de random walk met kans alpha een uitgaande edge naar een volgende node. Met kans 1-alpha springt de random walk naar een random node in de set met frauduleuze nodes. Elke iteratie wordt voor elke node bepaald wat de kans is dat het algoritme zich in deze node bevindt, dit heet de Page Rank. Het algoritme stopt als er convergentie optreedt, waarbij de nieuwe Page Rank waarden dicht bij de voorgaande waarden liggen. Deze Page Rank geeft de kans aan dat een node ook frauduleus is.

'Zit Donald Trump in jouw netwerk? Alle profielen op Facebook zijn in 5 stappen met elkaar te verbinden'
Bron: STAtOR, maart 2016

MEER FRAUDE OPSPOREN MET EEN NETWERKANALYSE

Het uitvoeren van een netwerkanalyse kan veel bijdragen aan het onderzoek naar fraude bij verzekeraars. Door een netwerk in kaart te brengen kan in eerste instantie gekeken worden naar bewezen fraude die voorkomt in een netwerk op basis van de Personalized Page Rank. Als er veel fraude voorkomt bij een netwerk van klanten, dan kan dat een aanleiding zijn voor een gedetailleerder onderzoek.

Maar zoals gezegd gaat een netwerkanalyse verder dan fraude opsporen op basis van al gevonden fraude. Het modelleren van verdachte gedragingen binnen een netwerkstructuur is een aanvullende manier om fraude op te sporen. Een significante afwijking van bepaalde claimoorzaken in een netwerkstructuur zou bijvoorbeeld een aanleiding voor een onderzoek kunnen zijn. Denk hierbij aan een netwerkstructuur met onevenredig veel inbraakschades. Een netwerk met veel claims binnen een bepaalde tijd is een ander voorbeeld. Een significant hogere claimfrequentie zou een indicator kunnen zijn voor fraude binnen een netwerk.

CONCLUSIE: DE NETWERKANALYSE ALS WAARDEVOLLE TOEVOEGING

Wij adviseren om de netwerkanalyse als middel toe te voegen aan het fraudedetectieproces. Het toepassingsgebied is tweeledig. Enerzijds kunnen afwijkende gedragingen (anomalies) binnen het netwerk een goede voorspelling/indicatie zijn voor fraude. Anderzijds ondersteunt een visueel aantrekkelijk en inzichtelijk gemaakt netwerk de fraude expert bij het afhandelen van lopende onderzoeken. Hopelijk zal hierdoor het percentage vastgestelde fraude stijgen en kan het aantal 'false positives' worden gereduceerd. Een fraude onderzoek is immers vaak een tijdrovende klus.

Tot slot genereert het toepassen van deze techniek vaak een soort sneeuwbaleffect. Als er eenmaal fraude is gevonden, kan dat leiden tot meer nuttige onderzoeken naar andere spelers in het netwerk. In de zaak tegen Ridouan T. zie je dit ook terug. Niet alleen hij maar ook vele anderen uit het netwerk zijn inmiddels opgepakt. ■

1 – Molloy I. et al. (2017) Graph Analytics for Real-Time Scoring of Cross-Channel Transactional Fraud. In: Grossklags J., Preneel B. (eds) Financial Cryptography and Data Security. FC 2016. Lecture Notes in Computer Science, vol 9603. Springer, Berlin, Heidelberg

2 – https://www.sicara.ai/blog/2019-01-09-fraud-detection-personalized-page-rank