

Nationaal verkeerskundecongres 2015

OV-potentie opsporen door datafusie

Een casestudie met OV Chipkaartdata en View.DAT voor Emmen

Martin Elfrink
Gemeente Emmen

Martin Courtz
Provincie Drenthe

Stephan Metz
OV Bureau Groningen Drenthe

Martijn Ebben
Goudappel Coffeng BV

Jeroen Weppner
DAT Mobility

Samenvatting

GSM-data en OV-chipkaartdata geven inzicht in het (onbenut) potentieel voor openbaar vervoer. In een casestudie voor Emmen zijn beide databronnen gecombineerd, wat veel inzichten geeft.

Trefwoorden

Openbaar vervoer, GSM-data, OV-chipkaart, reizigerspotentie

In dun bevolkte, vergrijzende gebieden is het lastig openbaar vervoer rendabel te exploiteren. Terecht werpen OV-autoriteiten, gemeentebesturen en vervoerders zelf kritische blikken op hun OV-systeem. Sluit dit systeem wel optimaal aan bij de behoefte? Het feitelijk OV-*gebruik* wordt al jaar en dag op allerlei manieren steeds beter gemeten. Tijd om de volgende stap te zetten: het in beeld brengen van de OV-*potentie*. Zijn er wellicht onopgemerkte kansen voor extra reizigers? Voor Emmen onderzochten de Gemeente Emmen, de Provincie Drenthe, OV Bureau Groningen Drenthe en Goudappel Coffeng/DAT.Mobility welke *feitelijke* inzichten zijn te ontleen uit de databronnen die de hedendaagse verkeerskundige ter beschikking staan: OV Chipkaart en View.DAT (op basis van GSM-data). Dat blijkt een interessante aanpak.

De onderzoeksvraag

Voor deze pilot zetten we de gemeente Emmen centraal. De gemeente Emmen telt ruim 100.000 inwoners en heeft met het Dierenpark Emmen een belangrijke publiekstrekker. Het vormt dan ook het centrum van de regio Zuidoost-Drenthe. De stad heeft veel ambitie, met de aanstaande verhuizing en uitbreiding van het dierenpark in het vooruitzicht wil de stad meer bezoekers trekken en een aantrekkelijke vestigingslocatie zijn. Kijken we naar de modal split, dan is – zoals in zoveel gebieden met een landelijk karakter - het autogebruik relatief hoog.

Onze onderzoeksvraag luidde: wat kunnen we uit ‘big data’ leren over aan te boren potentieel aan reizigers met bestemming Emmen voor bestaand of wellicht anders vormgegeven openbaar vervoer?

Belangrijk hierbij was onze ambitie zoveel mogelijk ‘te meten’ om daarmee schattingen zo realistisch mogelijk te laten zijn en inzichten/vermoedens van een feitelijke onderbouwing te voorzien. Of met feitelijke inzichten die vermoedens of heersende opinies te weerleggen. Dat is goed voor de kwaliteit van de vervolgbeslissingen. En het biedt procesmatige voordelen: minder improductieve discussie en meer draagvlak.

De onderzoeksopzet

Om de onderzoeksvraag te beantwoorden hebben we het OV-gebruik naar Emmen afgezet tegen het verplaatsingspatroon van mensen naar Emmen. Met andere woorden: we keken naar de OV-verhouding: hoeveel van het totaal aan verplaatsingen wordt gemaakt met het OV (en hoeveel dus niet). Een lage OV-verhouding wijst de weg naar een wellicht nog aan te boren doelgroep.

De databronnen

OV Chipkaart

Voor het feitelijke verplaatsingspatroon met het openbaar vervoer maakten we gebruik van OV Chipkaart-data. We volstonden voor deze eerste verkenning met de gegevens van één vervoerder (QBuzz). QBuzz rijdt namelijk vrijwel alle busvervoer in en rondom Emmen. Alleen de trein naar Zwolle (Vechtdallijnen, zie Courtz en Van Waveren, 2014) en buslijn 305 naar Groningen worden gereden door Arriva. Dit geeft dus een ondermeting op die reisrelaties die bijvoorbeeld ook door QBuzz lijnen 59, 300 en 600 worden bediend. Qbuzz leverde ons – volledig anoniem - van één maand alle gecombineerde ritten (dus inclusief overstappers) per uurblok met herkomst of bestemming Emmen: in totaal 85.619 reizen. Het gros betrof een werkdag (79.287). Op de zaterdagen registreerde Qbuzz in deze pilot 3.579 reizen en op zondagen waren dat er 2.753.

Van de OV Chipkaart data maakten we een aantal standaard visualisaties (drukke per dag, drukke per uurblok, instappers naar halte etc.). Deze grafieken en kaartbeelden kwamen grotendeels overeen met het verwachte beeld:

- Weekdagen fors drukker dan zaterdagen en zondagen
- Een piek op de maandagen
- Twee hele drukke ochtendspitsuren en een meer gespreide, maar vroege avondspits

- Zaterdagen en zondagen zijn veel gelijkmatiger en de drukte komt veel later op gang
- Veel instappers op Emmen Station, in het centrum, de 'scholierenhaltes' en de overstappunten

Naast bevestiging van onze verwachtingen zaten er ook verrassingen in de grafieken. Zo is het zondag om 19u00 opvallend druk in het OV. Een groot deel van deze reizigers reis naar Groningen: dit zijn waarschijnlijk studenten die na een weekend weer naar hun kamer terugreizen. De tweede opvallende uitkomst is de afwijkende verdeling van reizigers over de dagen in de tweede week. Hierover later meer.

View.DAT

Als tweede databron keken we naar de Herkomst- en Bestemmingspatronen (HB-patronen) zoals deze naar voren komen uit de GSM-data van View.DAT, Deze inzichten zijn gebaseerd op *een zeer groot aantal metingen* van daadwerkelijk gemaakte individuele verplaatsingen.

View.DAT maakt gebruik van de tegenwoordig alom aanwezige mobiele telefoon die zich in de regel meeverplaatst met zijn eigenaar (zie [Mede, 2015]). Apparaten als GSM's, smartphones of iPad's melden zich namelijk 20 – 200 keer per dag, ook als ze niet actief gebruikt worden, aan bij het GSM-netwerk. Het netwerk weet dan waar ze ongeveer zijn (in de buurt van die en die antenne). Voor de grootste netwerkprovider van Nederland, Vodafone, zijn dat circa ... datarecords per dag. View.DAT met partner Mezuro combineert en filtert al deze metingen tot een kennisbron met feitelijke druktepatronen en verplaatsingspatronen. Dat zijn in eerste instantie telefoonverplaatsingen, maar met ophoogheurstieken worden de gegevens vertaald naar *aantallen personen* die zich verplaatsen (o.a. gebaseerd op marktaandeel van Vodafone en data uit het OViN). Het detailniveau is 'per uur' en 'per kern'. Binnen de data is het mogelijk om onderscheid te maken tussen frequente, reguliere en incidentele bezoekers. Hierbij wordt gekeken naar het aantal keer dat de bezoeker in de gekozen maand in Emmen is geweest. Het onderscheid inwoner/bezoeker wordt eveneens 'gemeten': waar een telefoon 's nachts in de regel aanwezig is wordt aangemerkt als zijn woonplaats.

Met View.DAT bekeken we het mobiliteitsgedrag van bezoekers aan Emmen. Een verplaatsing van Emmen naar Groningen wordt dus niet meegeteld. Een verplaatsing Groningen Emmen wel. In onze onderzoeks maand ging het om 1.117.870 bezoeken aan Emmen. Ook hier komt het gros door de week (960.564). Op zaterdagen registreerde View.DAT 94.978 bezoeken en op zondagen 62.328.

Net als bij de chipkaartdata draaiden we ook op deze databron een aantal standaardvisualisaties (zoals de drukte per dag, bezoekfrequentie en een herkomstenkaart). De data liet het volgende beeld zien:

- een relatief stabiel weekpatroon qua aantallen en verdelingen over soorten bezoekers
- bijna 2/3 van de bezoekers komt frequent in Emmen
- op zaterdag komen bijna net zoveel bezoekers naar Emmen als op weekdays
- op weekenddagen is het aantal incidentele en reguliere bezoekers hoger dan op weekdays.
- Emmen Centrum is verantwoordelijk voor bijna de helft van de bezoekers

Het afwijkende reispatroon in de tweede week zoals de OV Chipkaartdata liet zie, is afwezig in de totale verplaatsing. Dat is dus interessant om nader uit te zoeken. Een verklaring hebben we daarvoor echter nog niet gevonden.

Uiteraard zijn er meer analyses te maken met View.DAT, maar voor dit onderzoek (b)leek het interessanter om te proberen eerst de databronnen met elkaar in verband te brengen.

Intermezzo: hoe zit het met de privacy?

'Big Data' en privacy zijn onlosmakelijk met elkaar verbonden. De onderzoeker – als professional - heeft het liefst alle individuele datarecords vanuit alle mogelijke bronnen. Dat koppelt en analyseert maximaal. Dezelfde onderzoeker – als privépersoon of betrokken burger – zit helemaal niet te wachten op dergelijke databeschikbaarheid die 'big brother' mogelijk maakt. Terecht zijn er zowel voor OV-

Chipkaartdata als voor gsm-data allerlei privacy waarborgen die zijn verankerd in wetgeving, contracten, protocollen én bewust aangebrachte technische beperkingen.

Het gaat de mobiliteitsonderzoeker gelukkig altijd om patronen en groepen, niet om individuele reizen of personen. Daarmee is de wellicht tot personen herleidbare databehoeftte beperkt tot *tijdelijke* verwerkingsslagen die ver 'onder de motorkap' (en dus maximaal afschermbaar) plaats vinden. View.DAT beschikt niet zelf over data van individuele verplaatsingen. View. DAT bevroagt een beveiligde database en krijgt louter aantallen terug (hoeveel reizen zijn er die aan deze criteria voldoen). Het antwoord is altijd minimaal 15, wat herleidbaarheid door het maken van doorsnedes onmogelijk maakt. De beveiligde database zelf (van Mezero) bevat – op een technisch identificerend kenmerk van de gemeten simkaart na - geen gegevens die personen identificeren of beschrijven. Het toestel-id is versleuteld en de encryptie wisselt elke maand. Met andere woorden: gegevens van individuele telefoons uit twee maanden zijn, ook al zou je dat willen en mogen, niet met elkaar in verband te brengen.

Ook de individuele OV-chipkaartdata verlaat de beveiligde omgeving alleen in geaggregeerde en geanonimiseerde vorm. Wij kregen reizen op uurblokniveau, waarbij op geen enkele wijze een relatie is te leggen tussen individuele reizen. Een reis van halte X naar halte Y en een reis van halte Y naar halte X die door twee verschillende reizigers worden gemaakt leveren dezelfde dataset op als een retourreis van een en dezelfde persoon.

Samengevat: we beschikken vanwege de privacy louter over inzicht in reispatronen

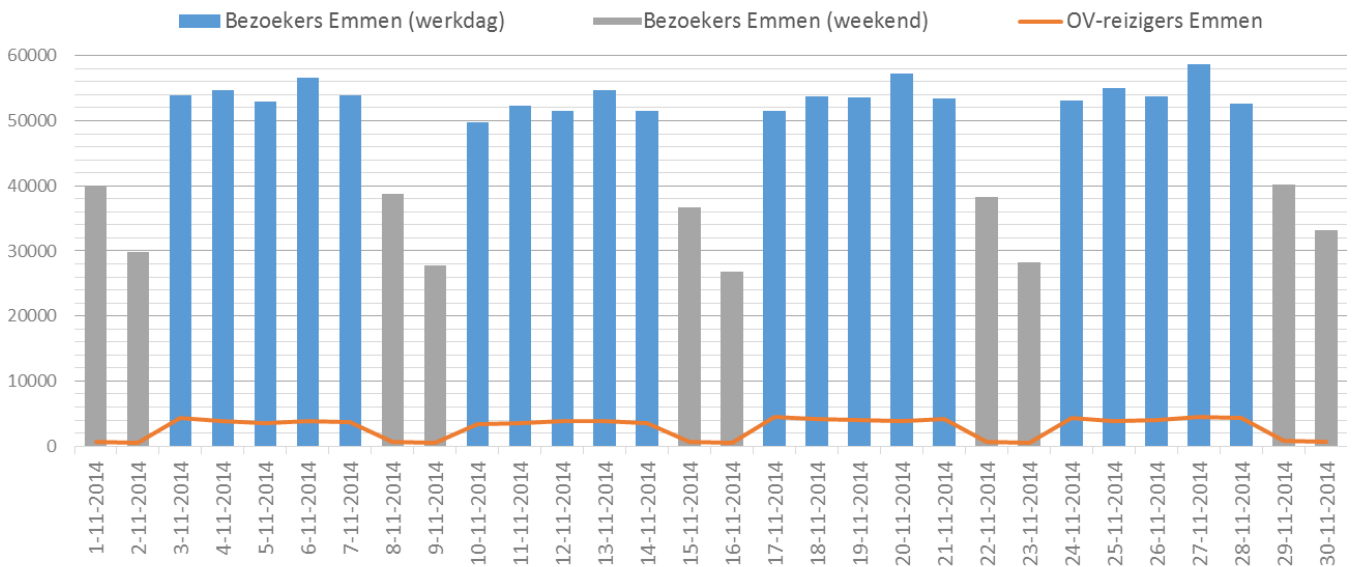
En nu: combineren

Ook zonder inzage in individuele verplaatsingen zijn View.DAT en Chipkaartdata echter met elkaar in verband te brengen op een manier die past bij onze onderzoeksvraag. Onze grootste uitdaging was het gelijk trekken van de beide onderzoekspopulaties. View.DAT betrof bezoekers van Emmen. In de OV Chipkaartdata staan waren alle ritten opgenomen met een herkomst of bestemming in Emmen. Een reiziger van Emmen naar Groningen die in de avond weer terugreist naar Emmen zit wel in de OV Chipkaartdata, maar niet in View.DAT (het is immers ene bezoeker van Groningen en niet van Emmen). We losten dit op door te kijken naar de tijdsperiode van 6 uur 's morgens tot 12 uur (het middaguur) en naar OV reizen met bestemming Emmen. Immers, zo was onze redenering, in deze periode zullen er in de OV-chipkaartdata primair ritten van bezoekers geregistreerd worden en nauwelijks retourritten van inwoners die terugkeren nadat ze die ochtend eerst vanuit Emmen waren vertrokken (deze zijn in de regel in de middag). Een dergelijke aanpak met gerichte aannames en dataselecties lijkt ons schaalbaar naar andere tijdsperioden of onderzoeksvragen.

Met de nu geharmoniseerde databronnen gingen we aan de slag. We deelden per herkomstgebied het aantal OV-verplaatsingen op het totaal aantal verplaatsingen: de OV-verhouding op de betreffende HB-relatie. Zowel de OV-verhouding als het absolute aantal verplaatsingen visualiseerden we geografisch naar herkomst. Die laatste indicator is belangrijk omdat bij lage absolute aantallen de betrouwbaarheid een punt van aandacht is. Reguliere spreiding kan dominantier zijn dan een door de visualisatie gesuggereerd effect. Bovendien zoeken we naar OV-potentie, daarvoor is ook de absolute omvang van de potentie relevant.

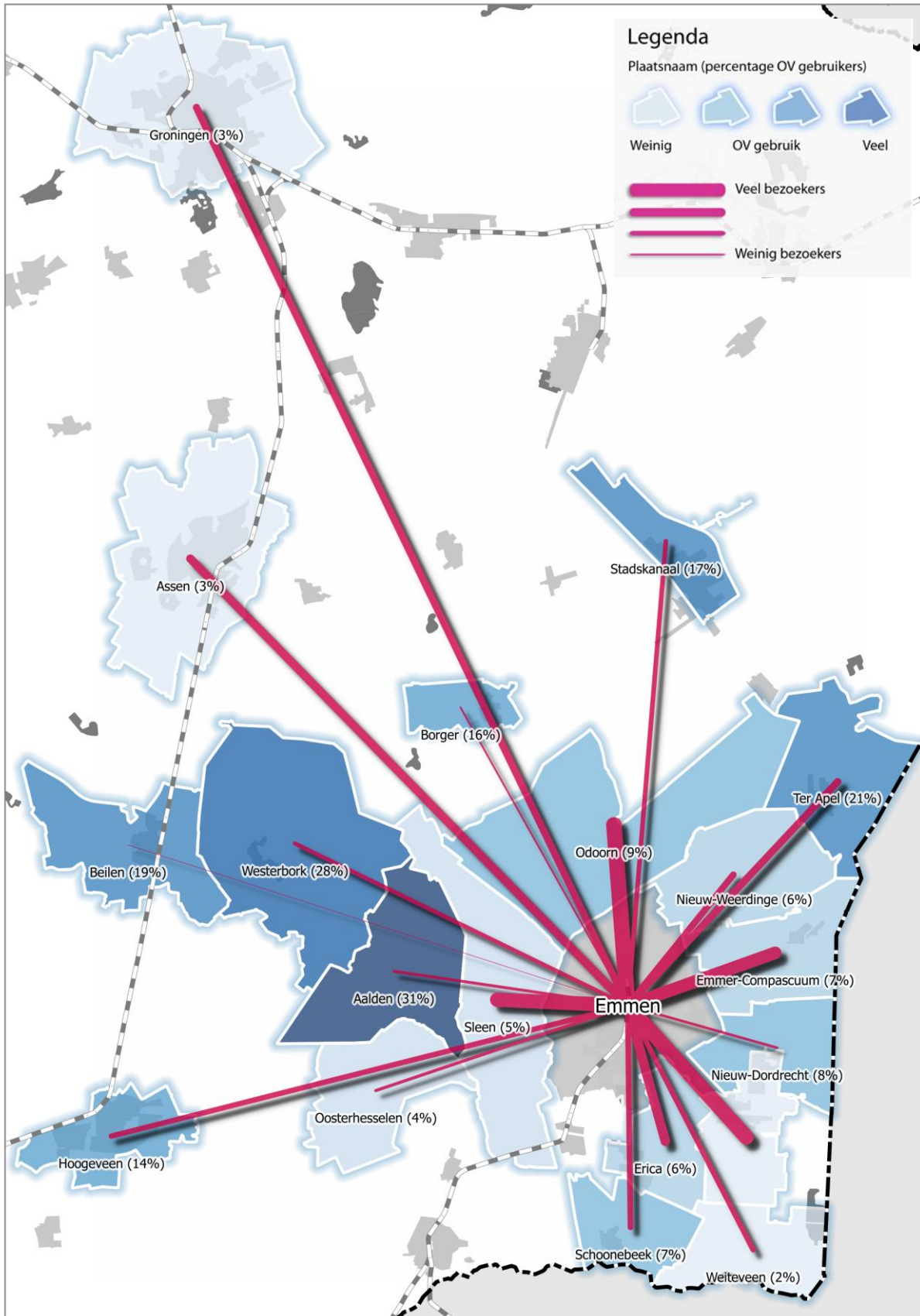
De resultaten

Het eerste wat dan opvalt is de afwijkende OV-verhouding op zaterdagen. Gemiddeld ontvangt Emmen op werkdagen circa 50.000 bezoekers, van incidentele dagtoeristen tot frequente werknemers en scholieren die niet in Emmen woonachtig zijn. Daarvan komen er op werkdagen tussen de 3.500 en 4.500 met het openbaar vervoer. Het OV-aandeel in de modal split is dus circa 8%. Kijken we naar de zaterdagen, dan daalt het aantal bezoekers naar ongeveer 40.000 terwijl het aantal OV-reizigers daalt naar ongeveer 750. Het OV-aandeel bedraagt op zaterdag nog slechts 2% (zie figuur 1).



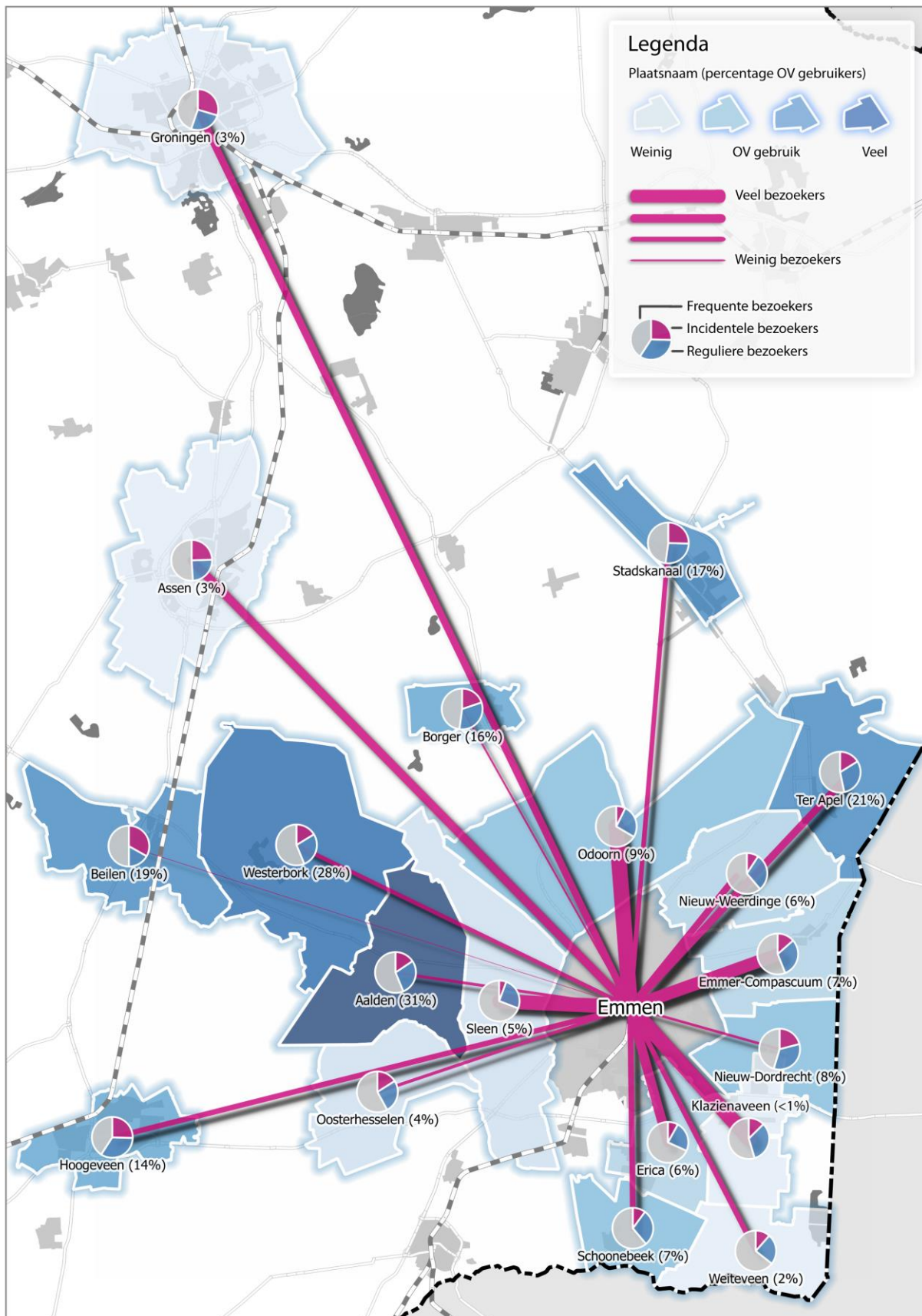
Figuur 1: dagelijkse mobiliteit naar Emmen stad (OV-reizigers QBuzz)

In figuur 2 is voor bestemming Emmen Centrum het OV-aandeel per herkomst weergegeven. Kernen die net iets verder liggen dan Emmen laten een relatief sterk 'OV-profiel' zien, zoals Ter Apel, Borger en Westerbork. De kernen liggen dicht genoeg bij Emmen om een grote stroom bezoekers te genereren, maar te ver weg om met de fiets te komen. Klazienaveen is een uitzondering hierop: een erg laag aandeel OV in combinatie met een groot aantal verplaatsingen. Een verklaring ligt mogelijk in de goede fietsverbindingen met Emmen. Schoonoord, Sleen en Erm vallen ook op: hun aandeel OV ligt lager dan veel gebieden die op dezelfde afstand liggen. Gezien het aantal verplaatsingen zijn deze relaties interessant om nader te onderzoeken, hier ligt immers OV-potentieel!



Figuur 2: aantal bezoekers van Emmen Centrum per werkdag en aandeel openbaar vervoer

Geen eenduidig beeld, maar wel interessante onderzoeksinformatie levert de confrontatie van het OV-aandeel met het bezoekersprofiel vanuit een gebied. Voor iedere kern is in figuur 3 een verdeling van de bezoekers naar frequentie van het bezoek opgenomen. Frequente bezoekers komen meer dan 10 maal per maand in Emmen, incidentele bezoekers minder dan 3 maal. Een interessant vervolgonderzoek lijkt ons de frequentie van de OV-gebruikers hiermee te confronteren. Hebben deze een totaal ander patroon? Het lijkt er wel op gegeven het achterblijvende OV-gebruik op zaterdagen, maar nadere analyse van de OV-chipkaartdata kan hierin inzicht geven.



Figuur 3: Verdeling bezoekers naar intensiteit bezoek in november en aandeel openbaar vervoer

Conclusies

Uit Big Data is veel te halen, zeker omdat we aan het begin van de curve staan qua ontwikkeling. De datarijkeid neemt alleen maar toe. Zo is de gebiedsindeling van View.DAT na afronding van deze gepresenteerde casestudie al weer verder verfijnd, waardoor inzichten aangescherpt kunnen worden.

De pilot Emmen laat daarnaast zien dat 'big data'-bronnen zich binnen de verkeerskunde prima laten combineren ookal verschillen zij qua onderzoekspopulatie en differentiatieniveaus. En wat nog belangrijker is: dat combineren kan zonder dat individuele reisbewegingen met elkaar gecombineerd hoeven worden. Dat maakt gecombineerde analyses eerder praktisch haalbaar (en het waarborgt daarnaast de privacy maximaal).

View.DAT en OV Chipkaart gecombineerd geeft ons voor het eerst (feitelijk) inzicht in de OV-potentie van geografische gebieden. Dat is een grote stap voorwaarts voor vervoerbedrijven en opdrachtgevers in het openbaar vervoer. Zij kunnen aan de slag met bijvoorbeeld een slimmer tarievenaanbod of andere vormen van mobiliteitsbeleid. De ingezette verdere verfijning van de gebiedsindeling van View.DAT maakt deze analyses steeds beter.

Tot slot leert de casestudie Emmen ons welke nieuwe datavragen relevant zijn, zoals bijvoorbeeld het inzicht in de bezoekersfrequentie in het openbaar vervoer. Hoe verhoudt die zich met de bezoekersfrequentie van alle bezoekers? En zou er nog toegevoegde waarde zitten in combinatie met tellingen uit de VRI's?

'Big Data' in de verkeerskunde is pionieren, maar duidelijk is wel dat we door de databerg te ontginnen kunnen komen tot een verdere verbetering van het openbaar vervoer.

Literatuur:

- Mede, Peter van der Mede, 'GSM-data niet meer weg te denken voor mobiliteitstoepassingen', in: GIS-Magazine, maart 2015.
- Courtz, Martin & Henk Doeke van Waveren, 'Databoogjes geven inzicht', in: OV-magazine, mei 2014.