

10:07 uur: Leve de #koning #Willem-Alexander #troon
#troonswisseling #Beatrix #Maxima

De waarde van 100.000 Twitter-Tweets voor verkeersmanagement



Eric de Kievit
Gemeente Amsterdam,
Dienst Infrastructuur Verkeer en Vervoer,
Beleid & Expertise, Team Expertise
e.de.kievit@ivv.amsterdam.nl

Bijdrage aan het Colloquium Vervoersplanologisch Speurwerk
21 en 22 november 2013, Rotterdam

Samenvatting

K(r)oningsdag 2013; de waarde van Twitter voor verkeersmanagement

Op 30 april 2013 vierde Amsterdam een bijzonder feest. Het was naast Koninginnedag ook de dag van de abdicatie. Een troonswisseling is een uniek evenement en de meeste mensen maken zo'n gebeurtenis maar twee of drie keer mee in hun leven.

Naast extra veiligheidsmaatregelen vraagt een dergelijke dag ook om maatregelen om de verkeersstromen in goede banen te leiden. Team Expertise van DIVV van de Gemeente Amsterdam heeft onderzocht of social media (Twitter) van voorspellende waarde kan zijn voor het managen van grote groepen mensen.

Op basis van 100.000 tweets is onderzocht of er relatie is te leggen tussen Twitter-activiteiten, drukte en verplaatsingen van mensen in de stad.

De volgende vragen stonden centraal:

- Ø Kan een relatie zichtbaar worden gemaakt tussen de epicentra van de gebeurtenissen over de dag en de epicentra van Twitteractiviteiten?
- Ø Is aantal tweets een maat voor drukte op bepaalde tijd op bepaalde plaats?
- Ø Is aantal tweets een maat voor de verplaatsingen van mensen in de stad?
- Ø Wat is de waarde van sociale media (Twitter) voor verkeersmanagement?

Er is gebruik gemaakt van alle Twitter-berichten op 30 april 2013, in een straal van 8km rondom de Dam. Het betrof tweets mét en tweets zónder geo-referentie. Door DTV Consultants¹ is een Twitter database opgebouwd met behulp van de Twitter-API. In een tweetal filmpjes is chronologisch in beeld gebracht wat de relatie is tussen gebeurtenissen op de dag en de Twitter-intensiteit. Vervolgens is de Twitter-intensiteit per uur vergeleken met intensiteit van het verkeer, voornamelijk richting het centrum. Hierbij is gebruik gemaakt van gegevens uit lus-telpunten en VRI's.

Voor het vaststellen van de drukte op een bepaalde locatie zou gebruik worden gemaakt van de speciaal voor deze dag ontwikkelde app: *30April-app*. De vrijdag voor de kroningsdag is besloten om deze functionaliteit van de app niet toe te staan. Als alternatief is voor vaststellen van drukte op de Dam gebruik gemaakt van camera-beelden.

Concluderend

Beide Twitter gegevensverzamelingen - mét en zonder geo-referentie - laten duidelijk verband zien tussen evenementenlocaties en twitteractiviteit. Tweets zónder geo-referentie laten zelfs een sterker verband zien. Dat is verrassend.

Twitter-activiteit als relevante bron voor verkeersmanagement kon niet worden aangetoond. Er kon géén verband worden gelegd met daadwerkelijke drukte op bepaalde evenementenlocaties.

De gemeente Amsterdam blijft onderzoeken of alternatieve databronnen geschikt zijn voor crowd-management, bijvoorbeeld GSM-data. Twitter-data is in ieder geval geen vervanging voor huidige telmethoden die hiervoor gebruikt worden.

¹ DTV Consultants bv. Troonswisseling; de kracht van uitzonderingsgevallen. Breda, 28 mei 2013

1. Waarom Twitter?

Nederlanders behoren in Europa tot de meest actieve sociale media gebruikers. Amsterdam vormt daarop geen uitzondering. Op een normale werkdag versturen Amsterdammers gemiddeld 375 tweets per uur. Het grote voordeel van Twitter is dat het een publieksmedium is. Twitter is een sociaal medium waarop mensen berichten vrijgeven zonder dat deze zijn voorbehouden aan specifieke mensen (vrienden) waarmee zij verbonden zijn. Dat betekent dat het medium toegankelijk is voor breed publiek. Gebruikers van Twitter in het algemeen en in het bijzonder de personen die hun locatie beschikbaar stellen, geven vrijwillig deze gegevens af.

In een aantal gevallen maakt ook een locatierferentie deel uit van deze gegevens en er kan ook een afbeelding of foto worden toegevoegd. Twitter biedt relatief goede analysefaciliteiten. Relatief omdat het uitvoeren van een goede analyse enigszins arbeidsintensief is.

De gemeente Amsterdam past daarnaast webcare toe. Ze reageert actief op uitlatingen die de gemeente aangaan in sociale media, bijvoorbeeld van burgers die vragen of klachten hebben over werk in uitvoering en omleidingen. Op Twitter wordt de hashtag #troonswisseling door de gemeente actief gebruikt tijdens Koninginnedag. Twitteractiviteit wordt daarmee dus ook gestimuleerd. Deze stond op plaats 6 in de meest gebruikte Koninginnedagrelateerde hashtags van de dag in geheel Nederland.



Ten slotte wilde team Expertise van DIVV een extra databron aanboren voor het verkrijgen van inzichten in het verkeer. Het moest hierbij gaan om een databron waar de gemeente Amsterdam nog niet standaard over beschikt en waarvan we zeker weten dat deze op 30 april beschikbaar zal zijn.

1.1 Opbouw van Twitter database

Het binnenhalen van alle Tweets is uitgevoerd door DTV Consultants uit Breda. Daarbij is gebruik gemaakt van de Twitter API (programmeerinterface) en Tweepy, een bibliotheek in de programmeertaal Python waarmee programma's geschreven kunnen worden om de Twitter API te benaderen en op die manier grote hoeveelheden gegevens op te halen.

Uiteindelijk zijn 95.660 tweets van 30 april 2013 binnen een straal van 8 kilometer van de Dam in analyse meegenomen: 68.520 originele tweets (72%) en 27.140 retweets

(28%). Het betrof 29.554 unieke twitteraars. Gemiddeld verstuurden zij 3,2 tweets per persoon. Het merendeel heeft 1 tweet verstuurd. Er waren een paar "heavy users".

Top 3 heavy-users	<i>112online</i> : n=453 tweets
	<i>bugisradio</i> : n=247 tweets
	<i>247robby</i> : n=298 tweets

1.2 Opvallend weinig tweets mét geo-referentie

Van de totale populatie bleek uiteindelijk slechts 4% van de tweets met GPS coördinaten te zijn verstuurd. Dat gegeven leidde tot enige teleurstelling, immers de verwachting was dat hoe meer twitteraars met GPS-locatie konden worden gevolgd, hoe betrouwbaarder de uitspraken over hun verplaatsingen.

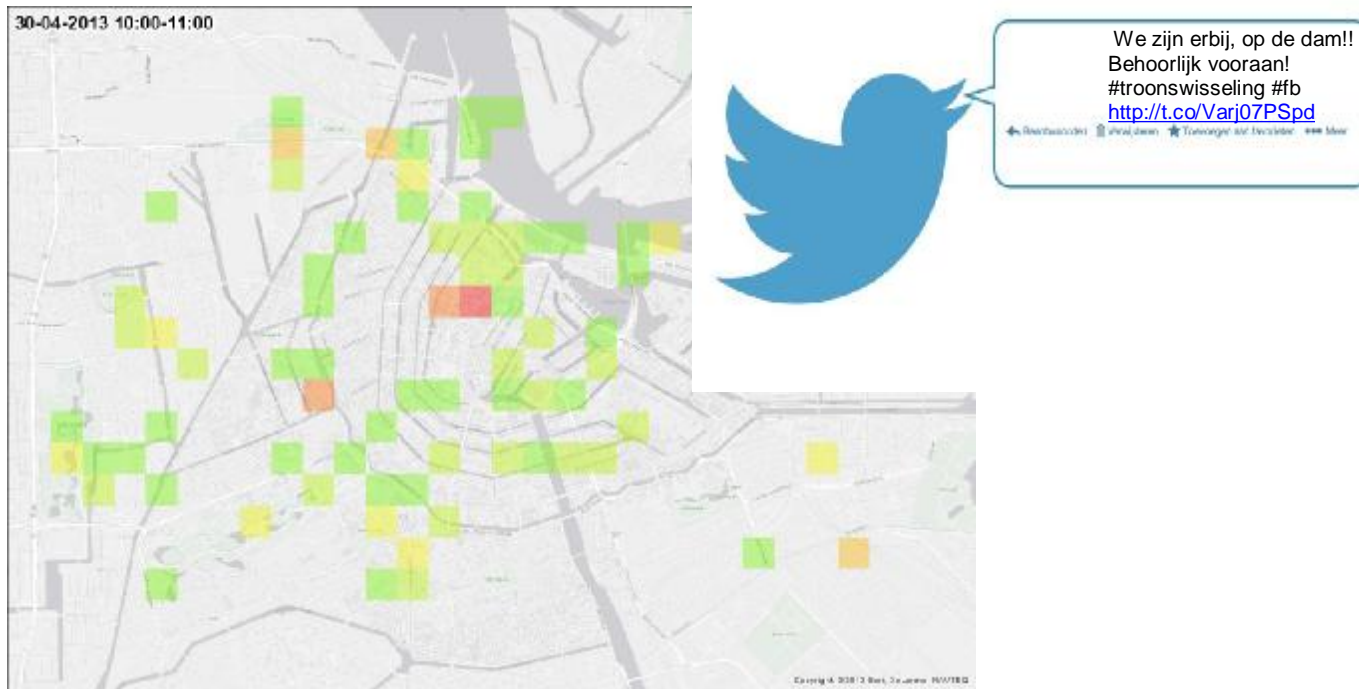
Verdeling tweets mét en zonder geo-referentie:

Aantal tweets	absoluut	relatief
Met geo-referentie	3.785	4%
Zonder geo-referentie	91.875	96%
Totaal	95.660	100%

Een tweet die in de dataset terugkomt hoeft niet per definitie op die locatie gedaan te zijn. Andersom kan het ook zijn dat een tweet die wel in dat gebied is gedaan, niet in de dataset terugkomt. Dit heeft te maken met het feit dat bij sommige tweets de locatie uit het profiel van het twitteraccount gehaald wordt en daarmee een vaste locatie aan de tweets geplakt wordt, ongeacht waarvandaan deze verzonden worden. De tweets die voorzien zijn van geo-referentie (lengtegraad, breedtegraad) worden altijd precies toegekend aan de locatie waarvandaan ze verzonden zijn.

1.3 Visualisatie tweets mét geo-referentie.

In afbeelding 1.3a is de Twitterintensiteit te zien op het moment van de abdicatie. Het verloop is van groen (weinig) via geel en oranje naar rood (druk). Een vierkantje representeert een gebied van 250 x 250 meter. Indien er geen kleur zichtbaar is, zijn er geen tweets verzonden in een vierkant.



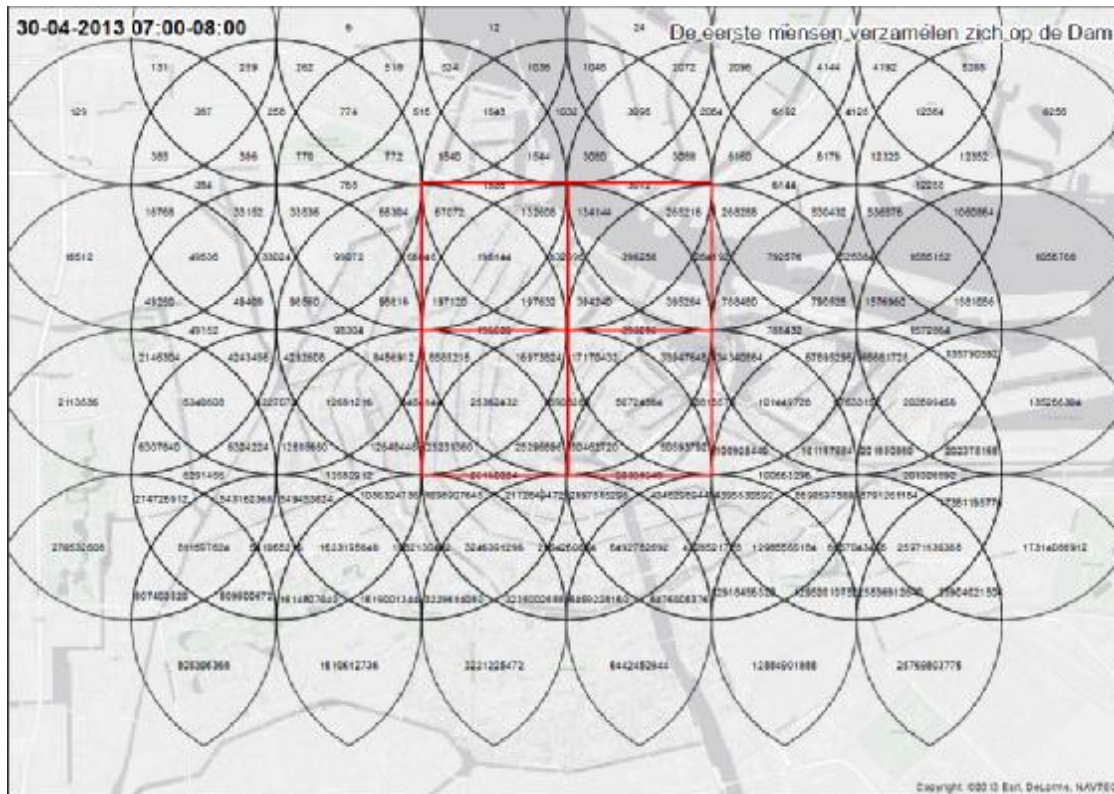
Afbeelding 1.3a: Twitterintensiteit tijdens de abdicatie

Duidelijk zichtbaar is de twitter-intensiteit op de Dam en het Museumplein. Het Museumplein was één van de locaties waar de gebeurtenissen op grote beeldschermen konden worden gevolgd.

Door alle beelden achter elkaar te plakken ontstaat een filmpje van twitter-activiteit over de dag. In de presentatie op het congres zal dit filmpje worden vertoond en dieper op opvallende feiten worden ingegaan.

1.4 Visualisatie tweets zonder geo-referentie.

Om de tweets zónder geo-referentie ook te kunnen visualiseren is door DTV een slimme methode toegepast. In onderstaande afbeelding 1.4a is te zien hoe de tweets zonder plaatsbepaling toch zijn toegedeeld aan een bepaalde locatie.



Afbeelding 1.4a: toedeling van tweets zonder geo-referentie

Met de Twitter API zijn alle tweets opgevraagd in een straal van 1 kilometer in ongeveer 10 overlappende cirkels op de visualisatiekaart, zodanig dat een cirkel die niet aan de rand ligt altijd voor honderd procent overlapt met ten minste één andere cirkel. Dit leidt uiteindelijk tot een visualisatie in vierkanten van 1 bij 1 kilometer.

Alle vlakken die door de overlap van de cirkels ontstaan hebben een nummer toegewezen gekregen. Omdat je met overlappende cirkels geen visualisaties kunt maken zijn de vierkanten die bij benadering ontstaan als figuurtjes door de overlap, als uitgangspunt genomen. Tweets kunnen in 1, 2, 3, of 4 cirkels voorkomen. Eigenlijk wordt gekeken in hoeveel databestanden ze zitten (het aantal opvraging per cirkel). Op deze manier zijn de berichten toegedeeld aan de vierkanten.

In afbeelding 1.4b is te zien wat de twitter-intensiteit op basis van tweets zonder geo-referentie tussen 10.30 – 11.30 is geweest.



Afbeelding 1.4b: Twitterintensiteit tweets zonder geo-referentie

Ook in deze afbeelding is het duidelijk te zien dat er een verband is tussen de Twitteractiviteit in een bepaald gebied en een bepaald programmaonderdeel. Deze beelden zijn ook achter elkaar gezet, zodat een filmpje ontstaat. In de presentatie op het congres zal ook dit filmpje worden vertoond en zal er dieper op opvallende feiten worden ingegaan.

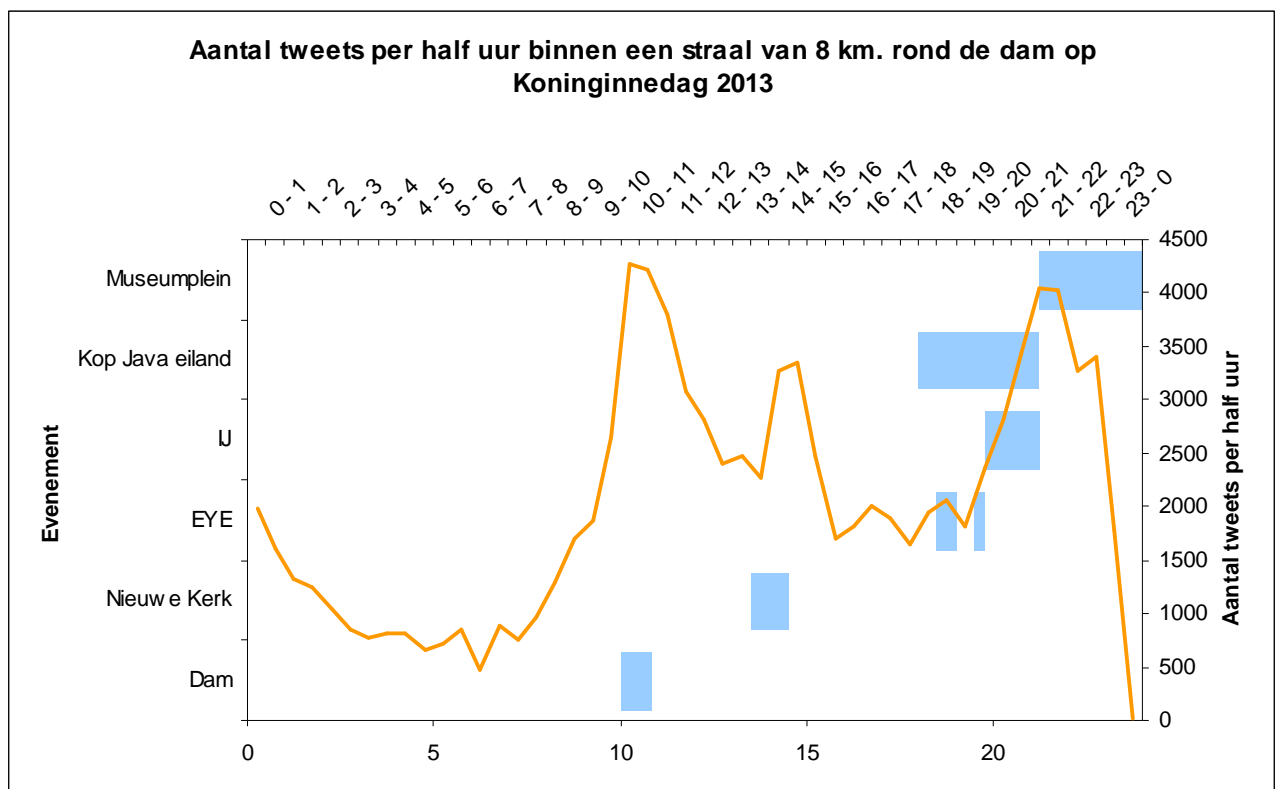
Hoewel niet duidelijk is vast te stellen welke tweets daadwerkelijk vanuit dit gebied zijn gedaan en welke niet, is het toch mogelijk dat er een duidelijk patroon ontstaat van het verloop van de dag. Met name de abdicatie, de balkonscene en de koningsvaart zijn goed te zien maar ook is het inhuldigingsprogramma herkenbaar in de video.

2. Twitter als indicator voor gebeurtenissen van de dag

Eén van de unieke aspecten van de troonswisseling is dat het evenement zich gevolgtijdelijk op meerdere plekken in de stad afspeelt en het programma over een groot deel van de dag verspreid is. Het Paleis op de Dam en de Nieuwe Kerk, het filmmuseum Eye, het Muziekgebouw aan het IJ en diverse evenementlocaties in de stad zoals het Museumplein, zijn de belangrijkste locaties. De genoemde evenementlocaties verschillen van elkaar in publiekswaarde. Enerzijds zijn er locaties met een extreem hoge publiekswaarde maar tegelijkertijd een tijdelijk karakter bijvoorbeeld het filmmuseum, het vertrekpunt van de geplande Koningsvaart. Daarnaast zijn er ook locaties die gedurende de gehele dag in gebruik zijn zoals het Museumplein.

2.1 De koninklijke stofzuiger

In afbeelding 2.1 is de gebeurtenissen van de dag en de twitter-intensiteit in één grafiek weergegeven.

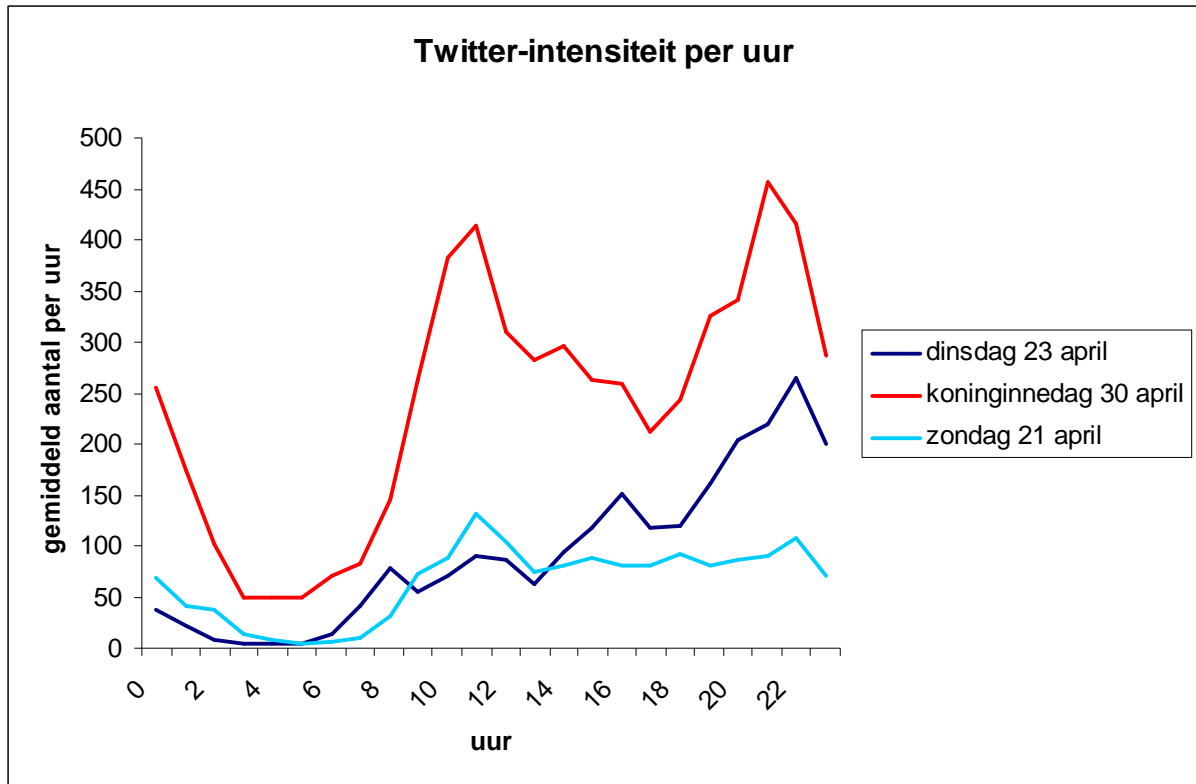


Afbeelding 2.1: relatie twitterintensiteit en gebeurtenissen van de dag

Nogmaals wordt duidelijk dat er een verband bestaat tussen een gebeurtenis tijdens de kroningsdag en het aantal verstuurd tweets. De abdicatie op de Dam, de troonswisseling in de Nieuwe Kerk en de festiviteiten zoals de Koningsvaart zijn goed herkenbaar. Het verschijnen van een persoon op het balkon, zoals een "man met een stofzuiger" werd om 07.00 uur in de ochtend door velen met gejuich ontvangen, wat weer aanleiding was voor extra twitterverkeer.

2.2 Hoe bijzonder is de K(ronings)dag?

Zoals hiervoor al beschreven: een Koninginnedag is al uniek, maar een troonswisseling is nog unieker. Om vast te stellen in welke mate het Twittergedrag afwijkt van een 'normale' Twitterdag is voor de dinsdag een week ervoor (23^{ste}) en de zondag ervoor (21^{ste}) het aantal tweets opgevraagd². In onderstaande afbeelding 2.2 staat het resultaat.



Afbeelding 2.2: twitterintensiteit op koningsdag vergeleken met 'normale' werkdag en zondag

Het betreft tweets met geo-referentie. De data van 30 april wijkt iets af van de DTV-data uit afbeelding 2.1, omdat ze afkomstig zijn van andere leverancier. Het patroon van de Kroningsdag van afbeelding 2.1 is opnieuw duidelijk herkenbaar. Er worden op deze 30 april ruim 2½ keer meer tweets verzonden dan op een normale werkdag. Op zondag is na 09.00 het aantal tweets per uur redelijk constant. Op een werkdag neemt twitter-verkeer toe met het verstrijken van de dag met een piek rond 16.30 uur en 22.00 uur.

2.3 Conclusie

Visualisaties van tweets mét en tweets zónder geo-referentie laten een duidelijk verband zien met de gebeurtenissen op de Kroningsdag.

² met dank aan Coosto bv voor het leveren van de Twitter-data

De tweets met geolocatie zijn heel precies toe te kennen aan een geografische locatie, maar het zijn er relatief weinig. Ondanks dat lijken gebeurtenissen die zich op straat afspelen en aan een klein gebied zijn toe te schrijven, goed op te sporen.

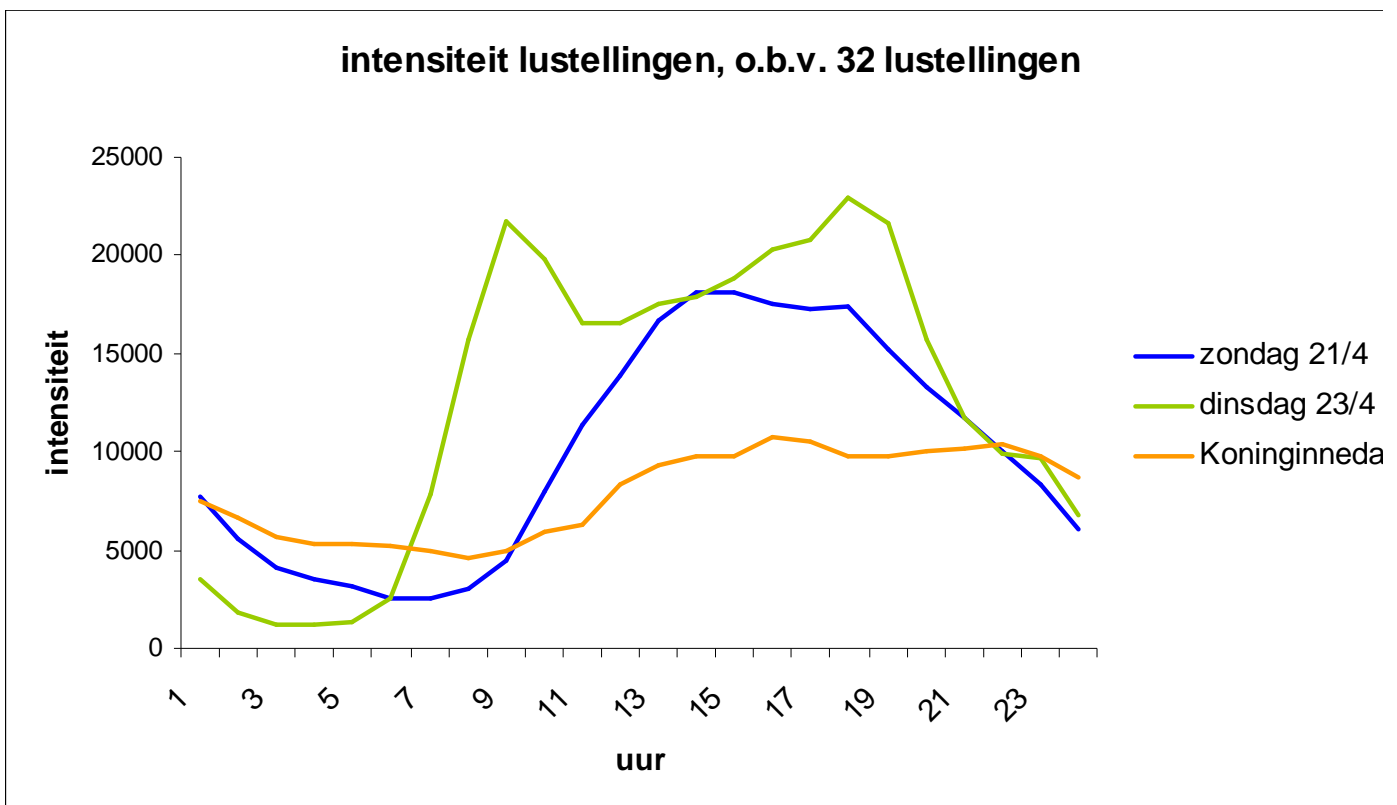
De tweets zonder geolocatie bevatten veel ruis. Er zitten bijvoorbeeld ook tweets bij die vanaf vaste computers verzonden worden, maar omdat ten gevolge van het evenement in het gebied ander gedrag ontstaat dan op een normale dag, blijkt dat het verloop van het evenement toch ook goed gevisualiseerd kan worden door de tweets rücksichtslos aan de geografische locatie te koppelen. Zelfs nog beter dan tweets mét geo-locatie.

3. Relatie verkeersintensiteit en twitter-activiteit?

In de auto wordt niet getwitterd mag je hopen. Toch is het interessant om te kijken of er een verband bestaat tussen twitter-activiteit en verkeersintensiteit. Het vermoeden is dat mensen zodra ze op plaats van bestemming zijn 'de wereld' laten weten "We zijn er het feest gaat beginnen". In de analyse is gebruik gemaakt van twee data-bronnen; enerzijds de gegevens uit de tellussen en anderzijds de gegevens vanuit de verkeersregelininstallaties (VRI's).

Alvorens tot die vergelijking over te gaan is eerst gekeken in hoeverre de verkeersintensiteit op de Kroningsdag afwijkt van een 'gewone' dinsdag en zondag. Immers op Koninginnedag is vrijwel iedereen vrij en is het eerder vergelijkbaar met een zondag dan een doordeweekse dag.

Afbeelding 3.1 geeft voor 30 april en de referentiedagen 23 april (week ervoor) en 21 april (zondag) de 24-uurs verkeersintensiteit in Amsterdam.

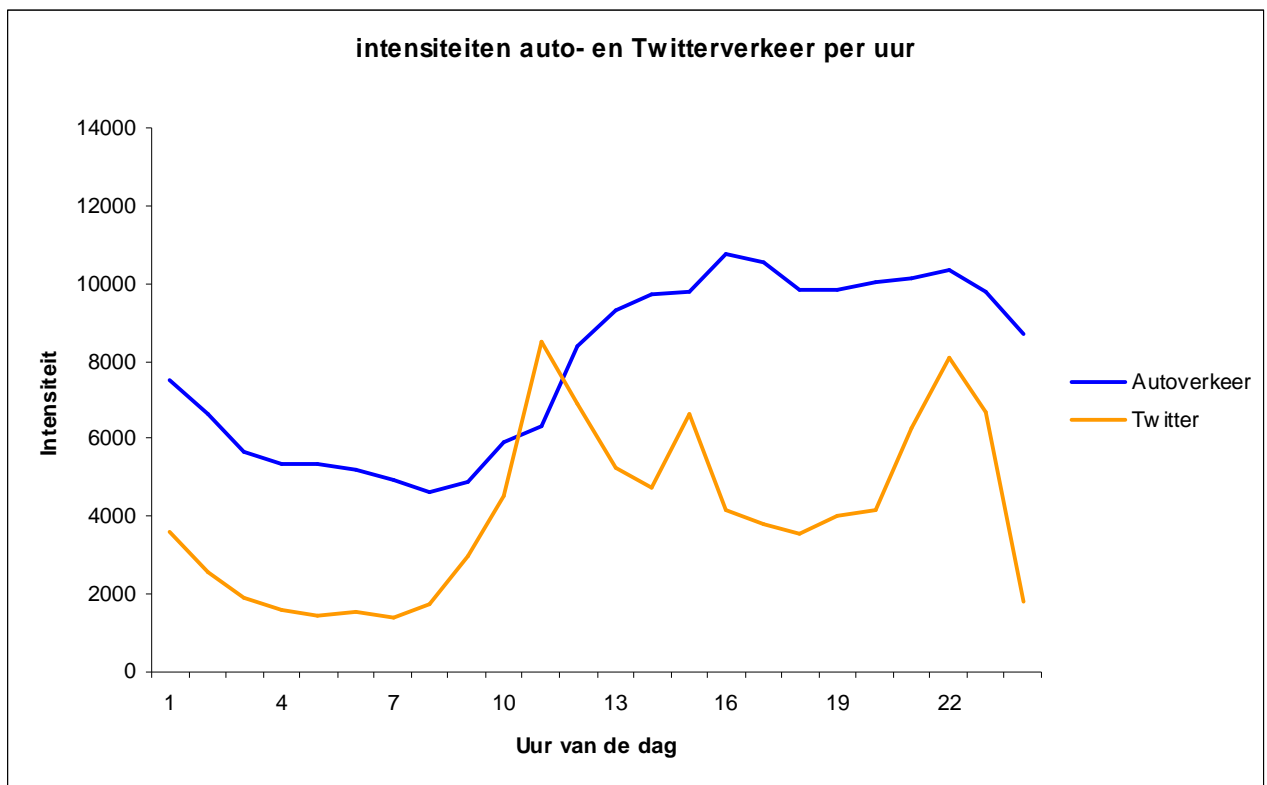


Afbeelding 3.1: verkeersintensiteit op kroningsdag vergeleken met twee 'normale' dagen

Het is duidelijk te zien dat de Kroningsdag aanzienlijk rustiger verloopt dan een normale werkdag en normale zondag. Dit is ook niet zo gek, omdat er een uitvoerig mobiliteitsplan in werking is gesteld met afsluitingen en omleidingen. Bovendien is iedereen geadviseerd om vooral niet met de auto naar Amsterdam te komen. Op een 'normale' werkdag zijn duidelijke pieken rond de spits te zien, de zondag verloopt vlakker. Deze beide dagen kennen aanzienlijk hogere intensiteiten dan 30 april.

3.1 Verkeersintensiteit en Twitter.

Nu we de verkeersintensiteit op Kroningsdag kennen, kunnen we intensiteiten van het Twitterverkeer en het 'normale' verkeer naast elkaar zetten. In afbeelding 3.1a zijn beide databronnen in één figuur gepresenteerd.

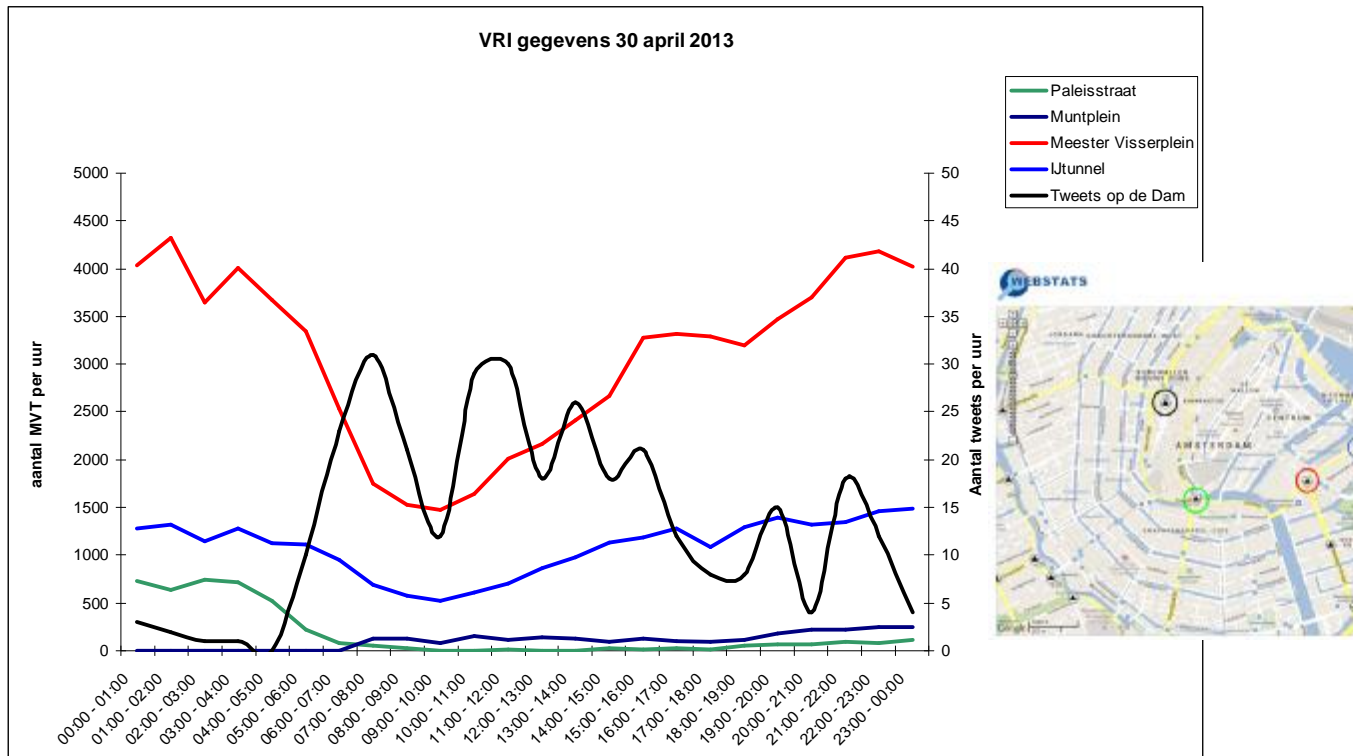


Afbeelding 3.1a: intensiteiten auto- en Twitter verkeer gedurende de dag

De Twitter-lijn laat een duidelijk verband zien met de gebeurtenissen van de dag, daarentegen verloopt de intensiteit van het autoverkeer gelijkmatig met een piek rond 17.00 uur. Mogelijk houdt dit verband met vertrek van bezoekers die genoeg gezien hebben of verplaatsingen naar Kop van Java-eiland in verband met het avondprogramma aldaar (o.a. optreden van Armin van Buuren).

Er is ook nog meer specifiek voor het verkeer rond de Dam gekeken naar het verloop van de verkeersintensiteit en tweets vanaf de Dam. In afbeelding 3.1b is van vier toegangswegen richting het centrum de verkeersintensiteit uitgezet tegen de Twitter-frequentie van tweets met geo-referentie.

Het Muntplein was vanaf 04:00 uur en de Paleisstraat vanaf 07:30 uur afgesloten voor verkeer. De tellingen van de Paleisstraat betreft hoogwaardigheidsbekleders die met speciaal vervoer naar Nieuwe Kerk werden vervoerd.



Afbeelding 3.1b: vergelijking Twitter-intensiteit met verkeersintensiteit uit VRI gegevens

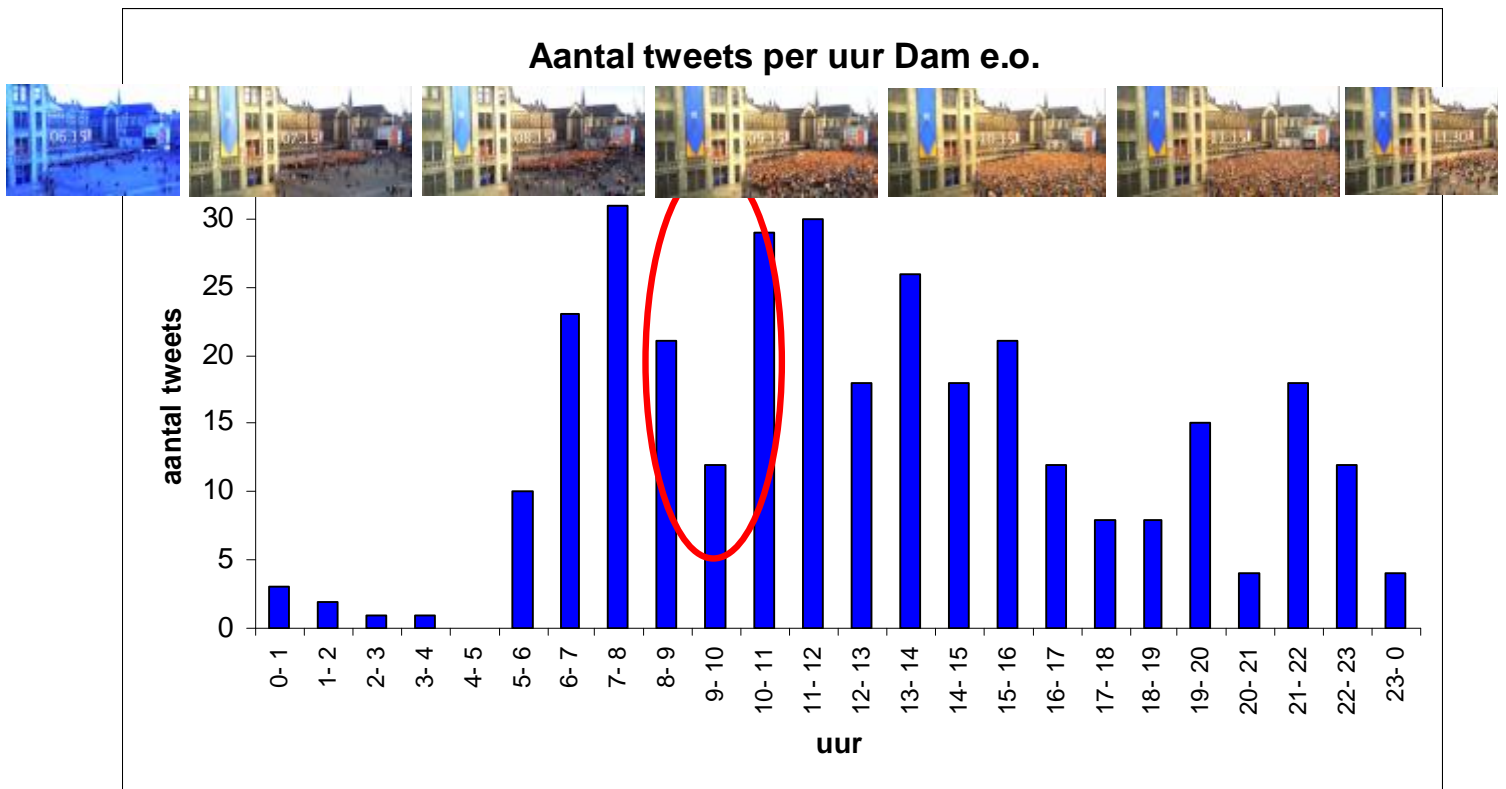
Uit de grafiek blijkt dat aantal tweets op de Dam fluctueert met de gebeurtenissen van de dag. Er is geen verband met de verkeersintensiteiten van en/of naar het centrum gevonden.

4. Is de Twitter-intensiteit een maat voor de drukte?

Bij de voorbereidingen van de Kroningsdag was het aanvankelijk de bedoeling om via de gratis *30april*-app mensen actief te berichten over drukte op bepaalde plaatsen in de stad. Tevens konden hier reisadviezen aan worden gekoppeld. Op 30 april is de app wel ingezet, maar de functionaliteit voor het bepalen van de drukte is vlak voor Kroningsdag ingetrokken.

In de analyse kon daardoor geen gebruik gemaakt worden van informatie over de drukte. Dat is jammer, omdat er een verwacht verband bestaat tussen de drukte op een bepaalde plaats en mate van twitter-activiteit. Dit kan relevant zijn voor crowd-management; onderdeel van de voorbereiding van dit grote publieksevenement, dat zich erop richt grote mensenmassa's en mensenstromen in veilige banen te leiden.

Tijdens de analyse kwamen wel beelden beschikbaar van de drukte op verschillende tijdstippen van de dag op de Dam. Deze beelden zijn langs de tijdlijn van twitterberichten op de Dam geplaatst. Afbeelding 4.1 toont het resultaat.



Afbeelding 4.1: relatie drukte op de Dam en Twitter-intensiteit

Tussen 08.00 uur en 10.00 uur is er een dip in de Twitter-activiteit onder bezoekers van de Dam zichtbaar. Op basis van de camera-beelden neemt de drukte op datzelfde tijdstip echter aanzienlijk toe. Tijdens de abdicatie om 10.00 uur is de Dam vol. Er is in dit geval dus geen één-op-één-relatie gevonden tussen de mate van drukte op bepaalde plaats en intensiteit van Twitter-verkeer.

4. Conclusies en aanbevelingen.

Op basis van analyses van bijna 100.000 tweets in een straal van 8 km om de Dam is relatie zichtbaar gemaakt tussen de epicentra van de gebeurtenissen over de dag en de epicentra van Twitteractiviteiten op 30 april. Uit de analyse blijkt verder dat Twitteractiviteit zich net als het inhuldigings-programma door de stad verplaatst. In de dataset zonder geo-referentie is de samenhang met het programma sterker dan in de dataset met geo-referentie. Dit was een enigszins verrassend resultaat. In beide gevallen gaat het erom dat de visualisatie mogelijk is doordat wat er gebeurt, afwijkt van wat er op een reguliere dag zou gebeuren. Het gaat niet om de totale dataverzameling, maar juist om de extra tweets die verzonden zijn door mensen die getuige waren van deze historische dag.

Er is in dit onderzoek geen verband tussen de drukte op verschillende plekken gedurende de troonswisseling en de Twitterintensiteit gevonden. Aanvankelijk zouden deze gegevens ontleend worden aan de app *30April* die helaas kort voor het evenement werd teruggetrokken. Twitter-activiteit als maat voor (dynamisch) verkeersmanagement kon daarmee (vooralsnog) niet worden aangetoond. Wij hebben geen verband gevonden tussen Twitter-activiteit en de verkeersintensiteit.

Het onderzoek naar het gebruik van Twitterdata ten behoeve van sturen van grote groepen mensen had voor het team Expertise van de Gemeente Amsterdam tot doel om kennis te ontwikkelen over de waarde van dit soort social media data. Wat we hebben geleerd is dat het in ieder geval géén vervanging kan zijn van onze huidige telmethoden en technieken.

Gebruiken van OV-chipkaart of GSM gegevens?

DIVV blijft zoeken naar alternatieve databronnen als middel voor het monitoren van het verkeer. De gedachten gaan daarbij uit naar het gebruik van GSM-data en OV-chipkaart gegevens.