

# Happiness at Work

De gebreken van een datagedreven analyse als kwalificatiemethode van medewerkersbevlogenheid

Anouk Meerman, 3942503

19-02-2016

Universiteit Utrecht

MA - New Media & Digital Culture

Begeleider: Ingrid Hoofd

Tweede lezer: Mirko Tobias Schäfer

## Voorwoord

In mijn huidige werk als projectmanager bij *Ynformed*<sup>1</sup> implementeer ik data-oplossingen voor maatschappelijke vraagstukken. Hierbinnen help ik publieke organisaties informatiegestuurd te werken. In mijn werk ervaar ik een sterke drang vanuit de publieke sector om effectiever en efficiënter met interne en externe informatie om te gaan. Echter speel ik vanuit mijn wetenschappelijke achtergrond nog vaak met de gedachte of dit wel de weg is die we zouden moeten inslaan. Deze gedachte wekt de volgende vragen bij mij op: Waar komt deze utopie rond *Big Data* en *Business Intelligence* vandaan? Op welke wijze vormt dit de publieke sector? Hoe gaat de gebruiker om met de data-oplossingen van *Ynformed*? En wat vinden de bronhouders zoals burgers hiervan?

De gemeente Zwolle en mijn collega's bij *Ynformed* hebben mij de mogelijkheid gegeven het datagedreven meten van menselijke gedragingen, vanuit een nieuw media perspectief kritisch te onderzoeken. Mijn dank gaat uit naar de ruimte en de support die ik hiervoor heb gekregen. Als afstudeerder stond ik vaak op het punt de spreekwoordelijke handdoek in de ring te gooien. Gelukkig gaf de positieve, maar kritische blik van mijn begeleider Ingrid Hoofd mij telkens weer de moed om door te pakken. Dankzij haar opmerkingen en vragen heb ik mijn thesis tot in de eindfase weten te perfectioneren. Graag maak ik van de gelegenheid gebruik om haar expliciet te bedanken voor haar inzet en geduld.

Met deze thesis zet ik trots een punt achter enkele bewogen, maar voornamelijk bevlogen studie jaren binnen mijn master *Nieuwe Media & Digitale Cultuur* aan de Universiteit Utrecht.

---

<sup>1</sup> Ynformed is een Utrechts adviesbureau gericht op het implementeren van data-oplossingen met publieke waarde. Door slim gebruik te maken van data en technologie creëert dit unieke kansen voor de publieke sector. Kansen om maatschappelijke vraagstukken op innovatieve wijze effectiever aan te pakken. [www.ynformed.nl](http://www.ynformed.nl)

# Inhoudsopgave

1 Inleiding	4
1.1 Assumpties van een datagedreven onderzoeksmethode	5
1.2 Kritische analyse van een datagedreven onderzoeksmethode	6
2 JD-R model	8
3 Multistep flow of information	11
4 Data en methode	13
4.1 Procedure	13
4.2 Participanten en data	13
4.3 Meetinstrument	14
4.3.1 Frequent mailen over een werkgerelateerde onderwerp	15
4.3.2 Positief gestemde mailcommunicatie	16
4.3.3 Formele of informele groepsdynamiek	17
4.3.4 Mailen buiten werktijd	18
4.3.5 Mediatoren	18
5 Bevlogenheid kwalificeren uit maildata	19
5.1 Betrokken bij werkgerelateerde onderwerpen	19
5.2 Positieve betrokkenheid is niet enkel onderwerp gerelateerd	21
5.3 Maillengte, formaliteit en sentiment	24
5.4 Mailen buiten werktijd als indicator van bevlogenheid	25
5.5 Mediatoren binnen het mailnetwerk	26
6 Conclusie en discussie	28
7 Literatuur	31
Bijlage A   Onderwerp label raamwerk	34
Bijlage B   Positief/negatief label raamwerk op basis van steekproef	35
Bijlage C   Formeel/Informeel label raamwerk	36

---

# 1 Inleiding

In zowel het bedrijfsleven als academische wereld neemt de belangstelling voor *Business Intelligence* (BI) toe (Watson 2007). BI wordt gekenmerkt door het omzetten van gegevens in informatie voor het slimmer laten werken van bijvoorbeeld een organisatie. Vanuit de praktijk wordt BI gezien als een kansrijke basis voor innovatie en productiviteit. Zo heeft de gemeente Zwolle laten blijken de mogelijkheden van een datagedreven meetinstrument voor medewerkersbevlogenheid te willen verkennen (Meerman 2015). Bevlogenheid is een gemoedstoestand - gekenmerkt door de kernfactoren vitaliteit en toewijding - dat een positief effect heeft op arbeidsprestaties (Bakker et al. 2006; Salanova et al. 2005; Sonnentag 2003). Ook vanuit een academisch perspectief wordt de rijkdom en het belang van BI-onderzoek steeds duidelijker (Chen et al. 2012). Professor José van Dijck (2014) omschrijft deze ontwikkeling als *datafiction* van het maatschappelijk en academisch paradigma.

*“Datafication as a legitimate means to access, understand and monitor people’s behavior is becoming a leading principle, not just amongst techno-adepts, but also amongst scholars who see datafication as a revolutionary research opportunity to investigate human conduct.”*

(Dijck 2014, 198)

*Datafication* uit zich in een wereldwijd panopticum, waar de mensheid effectieve beslissingen wil maken door het verzamelen van veel gegevens, om daar vervolgens informatie uit te vergaren en die te transformeren naar bruikbare kennis (Gray 2014, 535). Professor Mary Gray (2014) gedefinieerd dit als *dataveillance* systeem. In dit systeem worden sociale acties getransformeerd naar digitaal (online) gekwantificeerde data, dus bruikbaar voor real-time monitoring en voorspellende analyses (Mayer-Schönberger & Cukier 2013).

De gemeente Zwolle is een publieke organisatie die het *dataveillance* systeem omarmt en toepast bij het meten van medewerkersbevlogenheid. Om de medewerkersbevlogenheid te meten wordt momenteel een vijfjaarlijkse vragenlijst binnen het Medewerkerstevredenheid Onderzoek (MTO) uitgevoerd. Deze vragenlijst is opgesteld aan de hand van het veelgebruikte Job Demands-Resources (JD-R) model, een voorspellend theoretisch raamwerk van enerzijds stress factoren en anderzijds motivatieprocessen (Bakker, Demerouti & Schaufeli 1999). Ondanks de aangetoonde werking van het JD-R model stelt de gemeente dat het MTO voor hun organisatie geen gedetailleerd en frequent inzicht biedt in de bevlogenheid van individuele medewerkers, teams, afdelingen en totale organisatie (Meerman 2015). Een datagedreven kwalificatie zou volgens de gemeente een mogelijke oplossing bieden. De *datafiction* van een organisatie als de gemeente Zwolle gaat gebogen onder vele datagedreven kwalificatie assumpties.

## 1.1 Assumpties van een datagedreven onderzoeksmethode

*Datafication* assumpties vormen een klassiek probleem binnen datagedreven onderzoek. Om deze problemen beter te begrijpen is het ideologische gedachtegoed van *datafication*, zoals door danah Boyd en Kate Crawford (2012) en vele andere critici (Kitchen 2014; Lazer et al. 2014; Gitelman 2013), binnen dit nieuwe wetenschappelijke paradigma, gedeconstrueerd. *Datafication* biedt de geesteswetenschap een nieuwe epistemologische benadering voor het begrijpen van onze wereld: “rather than testing a theory by analysing relevant data, new data analytics seek to gain insights ‘born from the data’” (Kitchen 2014, 2). Critici Boyd en Crawford stellen dat deze nieuwe benadering gebogen gaat onder het geloof dat data beschikken over “a higher form of intelligence and knowledge that can generate insights that were previously impossible, with the aura of truth, objectivity, and accuracy” (2012, 663). Professor Rob Kitchen (2014) borduurt hierop voort en stelt dat datasets impliceren een objectieve en betrouwbare ontologie te hebben en daardoor voor zichzelf schijnen te spreken. Wetenschappers David Lazer, Ryan Kennedy, Gary King en Alessandro Vespignani (2014, 1204-1205) bekritisieren deze epistemologische omwenteling en stellen dat we ons niet enkel moeten focussen op de *Big Data Revolution*, maar op de *All Data Revolution*: analyses waarin zowel kleine als grote datasets van traditionele en nieuwe bronnen worden gecombineerd. Datagedreven inzichten hebben volgens hen namelijk niet de macht om voor zichzelf te spreken of een totaal domein te omvatten (Lazer et al. 2014). Data-analyse is hierdoor geen alleenstaande onderzoeksmethode, doordat de menselijke context ontbreekt en het een sample van de werkelijkheid blijft (Kitchen 2014).

Naast de wijze waarop inzichten uit data worden verkregen is er ook kritiek op het ideologische beeld van datavisualisaties. Kitchen (2014, 4) stelt dat de resultaten verkregen uit data impliceren vrij te zijn van menselijke framing of vooronderstellingen. Data zouden context of domein-specifieke kennis overstijgen, waardoor data te interpreteren is door iedereen die verstand heeft van statistiek of datavisualisaties. Media professor Christiane Paul (2007) bekritiseert deze aanname tweezijdig. Enerzijds wijst Paul (2007) op de rol van data, algoritmes en de interpretatie van de gebruiker. Media historicus Lisa Gitelman (2013, 2) vult daarbij aan dat ruwe data een oxymoron zijn, dit is iets wat van nature niet kan bestaan. Paul (2007) en Gitelman (2013) duiden beide op de processen van computers en de methode van analisten om een datapresentatie weer te geven. Mensen maken beslissingen in het registreren, verwerken, analyseren en presenteren van gegevens, waardoor data subjectief zijn. Anderzijds stelt Paul (2007) dat door de afhankelijkheid van de processen binnen een datavisualisatie-methode de weergave suggestief zijn. De lezer interpreteert de visualisatie vanuit zijn/haar referentiekader. Seth et al (2014, 81) omschrijven dit interpretatieproces als een relatie tussen data, informatie, kennis en wijsheid. Het bovenstaande impliceert dat mensen, dus ook analisten, vanuit hun eigen referentiekader waarden toekennen aan nummers.

Dit onderzoek poogt de bovenstaande *datafication* assumpties aan te tonen binnen een kritische analyse van een nieuwe datagedreven kwalificatie methode naar medewerkersbevlogenheid

## 1.2 Kritische analyse van een datagedreven onderzoeksmethode

In deze analyse wordt de medewerkersbevlogenheid definitie van de wetenschappers Wilmar Schaufeli en Arnold Bakker (2001) gehanteerd. Zij stellen dat bevlogenheid een positieve gemoedstoestand is en bestaat uit de kernfactoren vitaliteit en toewijding (Schaufeli & Bakker 2001). In hoofdstuk 2 wordt deze beschrijving verder toegelicht. Tot op heden heeft er geen gedegen wetenschappelijke analyse plaatsgevonden naar de mogelijkheden van een datagedreven kwalificatie van medewerkersbevlogenheid. Swan (2013, 93-94) beschrijft het kwalificeren van een kwantiteit, zoals een bevlogen gemoedstoestand, als een ontwikkeling die in de nabije toekomst zal plaatsvinden. Dit onderzoek biedt een eerste wetenschappelijke analyse van de wijze en mate waarop een datagedreven onderzoeksmethode patronen en indicatoren van een bevlogen identiteit kan aantonen. Hierbinnen ligt de focus op het kritisch analyseren van de limitaties en assumpties binnen de kwalificatiemethode en niet op het creëren van een nieuwe bevlogenheidschaal of meetinstrument.

Deze kritische analyse beantwoordt de vraag: *In welke mate en op welke wijze toont een datagedreven onderzoeksmethode patronen en indicatoren van medewerkersbevlogenheid?* Voor het datagedreven kwalificeren van medewerkersbevlogenheid fungeert binnen dit onderzoek de mailbox van 8 vrijwillige participanten (ambtenaren van de gemeente Zwolle) als databron. De corpus van deze analyse is een dataset van geanonimiseerde mailteksten. Uit mijn voorgaand onderzoek blijkt de mail het meest gebruikte digitale platform binnen de gemeente Zwolle te zijn en voor dit onderzoek de meest informatierijke databron (Meerman 2015).

Eerdere onderzoeken (Tyler et al. 2003; Wu et al. 2004) pogen informatiestromen in mailnetwerken te analyseren om onderliggende sociale structuren aan te tonen. Wetenschappers Joshua R. Tyler, Dennis M. Wilkinson en Bernardo A. Huberman (2003) bevestigen dat mailinteracties formele en informele gemeenschappen kunnen blootleggen. Sociale structuren die binnen, maar ook buiten organisaties kunnen ontstaan. Bovengenoemd onderzoek beschrijft twee soorten sociale netwerken die voortkomen in mail. Enerzijds formele gemeenschappen, waarbij mailinteracties gebaseerd zijn op bestaande bedrijfsstructuren. Anderzijds informele gemeenschappen, waarbij netwerken ontstaan door overeenkomende persoonlijke interesses of activiteiten. Waar eerder onderzoek (Tyler et al. 2003) zich focust op het blootleggen van netwerken; wie met wie samenwerkt en hoe informatiestromen binnen informele en inofficiële netwerken het bedrijf in en uit gaan, gaat het voorliggend onderzoek verder. Dit onderzoek poogt medewerkersbevlogenheid te kwalificeren door de wijze en mate waarop individuen zich binnen mailgemeenschappen tot elkaar verhouden, op basis van bevlogen mailgedrag ten aanzien van werkgerelateerde onderwerpen.

Voor het opstellen van medewerkersbevlogenheid kenmerken dient in deze analyse het eerder genoemde JD-R model als basis (Bakker, Schaufeli, & Demerouti 1999). Het JD-R model beschrijft indicatoren en processen die een bevlogen gemoedstoestand vormgeven, zoals werkstressoren, energiebronnen en persoonlijke hulpbronnen binnen een werksituatie. In hoofdstuk 2 wordt het model en bijhorende indicatoren nader toegelicht. Vanuit een nieuw media perspectief wordt in deze analyse stilgestaan bij het bestaansrecht van een datagedreven JD-R model en de mogelijke gebreken ervan. Doordat het JD-R model geen inzicht biedt in concrete gedragingen binnen een

digitaal medium als mail, heb ik mij voor het beantwoorden van de hoofdvraag verdiept in de nieuwe media theorieën van Richard Rogers (2013). Hij biedt in zijn boek *Digital Methods* handvatten voor het analyseren van online gedragingen binnen websites, als sociale online platformen (Rogers 2013). Deze handvatten zijn tevens bruikbaar voor het analyseren van mail, omdat dit ook een netwerk is aan digitale sociale interacties. In hoofdstuk 4 worden deze handvatten voor de onderzoeksmethode nader toegelicht.

Om mail interacties, die kunnen duiden op een bevlogen gemoedstoestand beter te begrijpen, biedt de Two-step Flow theorie van Elihu Katz en Paul Felix Lazarsfeld (1955) bruikbare inzichten. Deze theorie stelt dat er een 'multistep flow of information' plaatsvindt. Dit is een interactiepatroon waarbij informatie van massamedia naar opinieliders gaat en vervolgens naar het algemene publiek (Katz en Lazarsfeld 1955, 32). In hoofdstuk 3 wordt de Two-step Flow theorie nader toegelicht. Ondanks dat massamedia in deze analyse geen katalysator zijn, aangezien mailen een één-op-één communicatie is, is deze theorie gebruikt doordat het uitgaat van een netwerk aan interacties en informatiestromen met verschillende beïnvloeding-niveaus. Iets wat in een onderzoek naar bevlogen gedragspatronen binnen een digitaal dynamisch medium als mail relevant wordt geacht (Rogers 2013). Daarnaast hebben, Tyler et al. (2003) aangetoond dat binnen mailgemeenschappen het beïnvloedingsniveau van leiders te identificeren is. Leiders beïnvloeden, los van formele bedrijfsstructuren, opinies over werkgerelateerde onderwerpen vanwege hun grote bereik en hoge mail frequentie. Wetenschappers Fang Wu, Bernardo A. Huberman, Lada A. Adamic en Joshua R. Tyler (2004) borduren hierop voort en tonen informatiestromen binnen deze sociale netwerken aan. Wu et al. (2004) onderbouwen dat deze mailinteracties en informele gemeenschappen ontstaan op basis van betrokkenheid en persoonlijke interesses. Deze gegevens tonen aan dat de informatiestromen en beïnvloedingsrollen, waar de Two-step Flow theorie vanuit gaat, ook binnen mailnetwerken plaatsvinden. Het JD-R model en Two-step Flow theorie dienen samen als kader voor de mailanalyse naar medewerkersbevlogenheid en het framen van de resultaten.

In hoofdstuk 4 wordt de maildatagedreven methodologie voor het kwalificeren van medewerkersbevlogenheid nader toegelicht. Aan de hand van het theoretisch kader zijn vijf hypothesen opgesteld, namelijk een bevlogen medewerker (1) is betrokken bij werkgerelateerde onderwerpen en mailt hier frequent over naar collega's; (2) mailt meer positief over werkgerelateerde onderwerpen, dan negatief; (3) heeft een sterke, informele en positieve relatie met collega's; (4) heeft door zijn/haar vitaliteit en toewijding geen 9-5 mentaliteit en mailt daarom ook buiten werktijd; (5) is een positieve mediator en beïnvloedt door een sterke betrokkenheid en groot bereik veel collega's. In hoofdstuk 4 worden methodologische assumpties en limitaties aangekaart. Het analyseren van mail biedt een onderbouwing of verwerping van de vijf hypothesen. In hoofdstuk 5 zijn de resultaten in de vorm van staafdiagrammen, opsommingen en Gephi netwerkvisualisaties gepresenteerd. Daarbij zijn de assumpties en limitaties die schuil gaan achter deze resultaten kritisch uiteengezet. Ten slotte wordt in hoofdstuk 6 een conclusie gegeven die antwoord geeft op de hoofdvraag, inclusief enkele discussiepunten voor vervolgonderzoek.

---

## 2 JD-R model

Een identiteit wordt gevormd door herkenbare terugkerende narratieve. Het is een proces dat voortdurend beweegt en zich door het heden, verleden en culturele representatie vormt. Een identiteit van zowel een individu als collectief staat volgens Stuart Hall (1990, 223-229) niet vast. Een relevante identiteit voor het huidige onderzoek is een bevlogen medewerker. Bij een dergelijke identiteit loopt een persoon op de werkvloer warm voor bepaalde bedrijfsgerelateerde zaken waar zowel de werknemer als de organisatie baat bij hebben (Schaufeli & Bakker 2001). Deze identiteit is van belang voor de huidige mail data-analyse, welke grip probeert te krijgen op de bevlogen narratieve uit het verleden en heden. Om meerdere kanten van deze mentale staat te belichten is het belangrijk om ook de tegenhanger van bevlogenheid te bekijken. Dit wordt door Maslach en Leiter (1997) kort omschreven als een burn-out, wat duidt op uitputting, distantie en een gebrek aan competentie op het werk. Anders gezegd impliceren Maslach en Leiter (1997) dat wanneer er een tegenovergesteld patroon wordt geconstateerd op het Maslach Burnout Inventory (MBI) of Utrechtse Burn-Out Schaal (UBOS) (Schaufeli & Bakker 2001) – de meest gebruikte schalen voor het meten van een burn-out – er sprake is van bevlogenheid. Hierin wordt duidelijk dat bevlogenheid zich voordoet in een andere situatie dan een burn-out, met haar eigen kenmerken en triggers. Met dit gegeven streven wetenschappers als Schaufeli en Bakker (2001) voor twee aparte dimensies. Zij duiden op de tegenstrijdigheid van de twee waarbij ze aangeven dat beiden onafhankelijk van elkaar gemeten moeten worden. Het zijn geen directe tegenpolen, maar twee totaal andere situaties. Het psychometrisch aangetoonde Utrechtse Bevlogenheid Schaal (UBES), in het Engels verwoord als Utrecht Work Engagement Scale (UWES), biedt hiervoor uitzicht op meer duidelijkheid (Schaufeli & Bakker 2001). Deze schaal geeft aan dat bevlogenheid een complex proces is dat per identiteit kan verschillen.

Schaufeli en Bakker (2001) definiëren bevlogenheid specifiek als een positieve, affectieve en cognitieve toestand van oprechte voldoening. Gekenmerkt door de factoren vitaliteit, toewijding en absorptie.

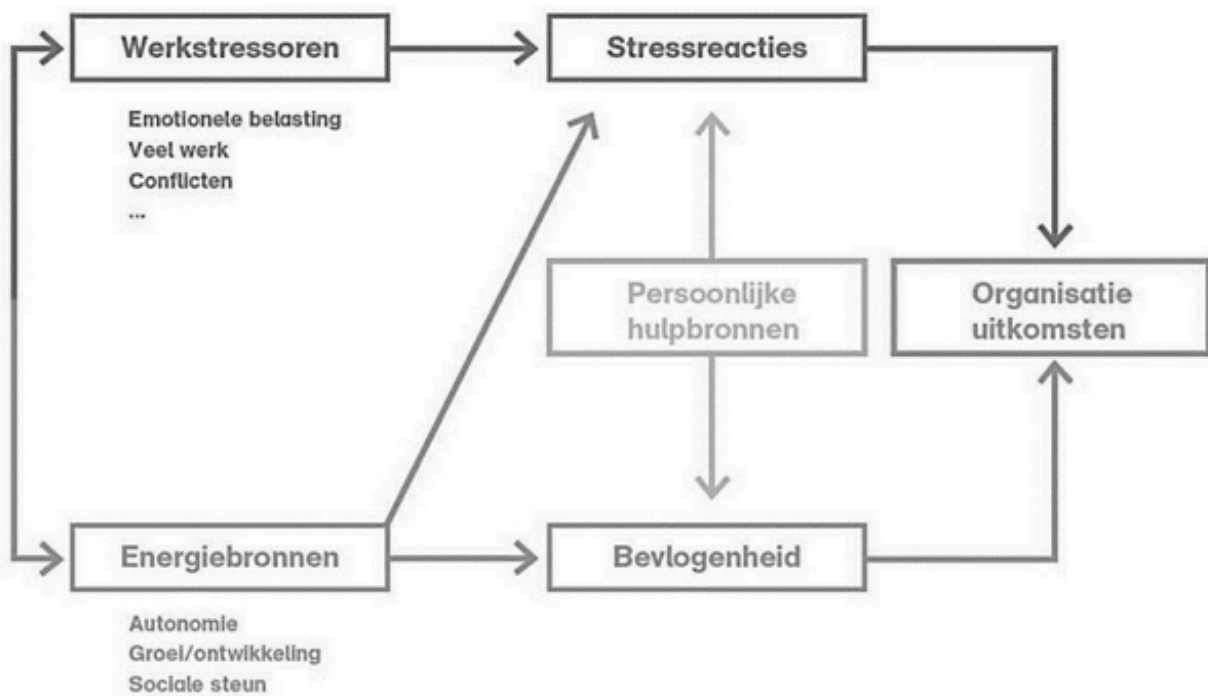
*“Vitaliteit wordt op haar beurt gekenmerkt door bruisen van energie, zich sterk en fit voelen, lang en onvermoeibaar met werken door kunnen gaan en beschikken over grote mentale veerkracht en dito doorzettingsvermogen. Toewijding heeft betrekking op een sterke betrokkenheid bij het werk; het werk wordt als nuttig en zinvol ervaren, is inspirerend en uitdagend, en roept gevoelens van trots en enthousiasme op. Absorptie, ten slotte, heeft betrekking op het op een plezierige wijze helemaal opgaan in het werk, er als het ware mee versmelten waardoor de tijd stil lijkt te staan en het moeilijk is om er zich los van te maken.”*

(Schaufeli & Bakker 2001, 245).

De maildata-analyse borduurt voort op de bovenstaande definitie van bevlogenheid. Tot op heden is niet aangetoond of absorptie daadwerkelijk een kernfactor is of voortvloeit uit vitaliteit en toewijding (Bakker & Demerouti 2008). Om deze reden wordt in deze analyse naar medewerkersbevlogenheid



enkel gefocust op het meten van de bovenstaand omschreven kernfactoren vitaliteit en toewijding. Bij het voorspellen van deze kernfactoren hanteert het UBES het Job Demands-Resources model (JD-R model) van Arnold Bakker, Eva Demerouti en Wilmar Schaufeli (1999), afgebeeld in figuur 1.



Figuur 1. JD-R model (Schaufeli & Bakker 2004)

Werkstressoren, ook wel taakeisen genoemd, energiebronnen en persoonlijke hulpbronnen zijn de voorspellende factoren van een bevlogen identiteit in het JD-R model. Werkstressoren zijn volgens Schaufeli en Bakker (2004) sociale, psychologische, fysieke of organisatorische aspecten van een baan die aanhoudende spanningen met zich meebrengen. Deze factoren impliceren een negatieve relatie te hebben tot de mate van bevlogenheid. Energiebronnen zijn de sociale, psychologische, fysieke of organisatorische aspecten die bijdragen aan werkgerelateerde doelen en persoonlijke groei en/of ontwikkeling (Bakker et al. 1999). Denk hierbij aan extrinsieke en intrinsieke motivaties die gesitueerd zijn in organisatorische processen, interpersoonlijke elementen, werkstructuren en taken (Demerouti & Bakker 2011). Persoonlijke hulpbronnen, als optimisme, weerbaarheid, zelfsturing en persoonlijke effectiviteit, is het laatste element dat Schaufeli en Bakker (2004) hebben toegevoegd aan het JD-R model. Deze hulpbronnen kunnen direct en indirect effect uitoefenen op de bevlogenheid van medewerkers (Llorens et al. 2007).

Uit voorgaande analyses blijkt dat er een aantoonbare positieve relatie is tussen bevlogenheid en arbeidsprestaties (Bakker et al. 2006; Salanova et al. 2005; Sonnentag 2003). Deze maat is multifactorieel en heeft betrekking op de positieve houding ten aanzien van werk en organisatie (Salanova et al. 2005), maar ook op positief gedrag, zoals persoonlijk initiatief, leermotivatie, extra rolgedrag en proactief gedrag (Sonnentag 2003). Voorts zijn er aanwijzingen dat bevlogenheid

gepaard gaat met een positief effect op de gezondheid, als weinig depressieve, stress gerelateerde en psychosomatische klachten (Schaufeli et al. 2004) en aanstekelijkheid van bevlogenheid (Bakker & Demerouti 2008). Binnen deze relatie bevinden zich confounders die op hun beurt bevlogenheid versterken of verzwakken. Confounders zijn achtergrondverhalen die het verband tussen afhankelijke en onafhankelijke variabelen vertroebelen. Denk hierbij aan het effect van een positieve of negatieve groepsdynamiek op individuele bevlogenheid. Bakker en Schaufeli (2001) tonen aan dat de mate van bevlogenheid niet enkel een individueel gegeven is. Tevens blijkt dat bevlogen teams meer hulpbronnen weten te organiseren, hetgeen positief bijdraagt aan het niveau van bevlogenheid van individuele teamleden (Bakker & Schaufeli 2001). Zij stellen dat hoe hoger de bevlogenheid van een team is, des te meer bevlogen de individuele teamleden zijn, wat resulteert in een hogere productiviteit. Wetenschappers Salanova et al. (2003), Schaufeli en Bakker (2001) stellen dat een confounder binnen bevlogen identiteiten een mediërende rol aannemen. Mediatoren hanteren hulpbronnen, zoals digitale technische hulpmiddelen, en beïnvloeden het serviceklimaat en arbeidsprestaties, zoals extra rolgedrag en klanttevredenheid. Met een positieve houding kan een bevlogen individu en/of mediator invloed uitoefenen op groepen als teams en afdelingen.

Voorliggend onderzoek poogt de hierboven omschreven indicatoren van vitaliteit en toewijding, omschreven als werkstressoren, energiebronnen, hulpbronnen, confounders en mediërende rollen, te vormen tot een nieuw media meetinstrument dat toepasbaar is op maildata. Hierbij wordt gebruik gemaakt van de *Digital Methods* handvatten door Rogers (2013) en de communicatie Two-step Flow theorie van Elihu Katz en Paul Felix Lazarsfeld (1955).

---

### 3 Multistep flow of information

Professor M. Lynne Markus (1991) bekritiseert de *Information Riches Theory* (IRT) en stelt dat e-mail communicatie een zeer informatierijk bron is. De IRT stelt dat face-to-face communicatie een veelzijdig medium is door de aanwezigheid van lichaamstaal en directe feedback. Opvolgend worden telefoongesprekken beschreven als middelmatig medium en documenten als mager (Daft & Lengel 1986, 560). Markus (1991) stelt dat mailteksten informatierijk zijn doordat het sociale structuren, regels en gevoelens als afstandelijkheid, zelfstandigheid en gepastheid bevatten. Hierbij is de welbekende uitspraak van Marshall McLuhan (1967) "The medium is the message" van toepassing. Mail als medium heeft invloed op hoe de zender en ontvanger berichten ervaart. De wijze waarop bevolegenheid in mail tot uiting komt kan beïnvloed worden door het digitale medium. Zo kan mailen buiten werktijd, naast bevolegen gedrag, ook een signaal zijn voor overwerkt; Of kan weinig mailen duiden op een lage betrokken of een clean desk policy. De benadering van dit soort tegenstrijdigheden wordt in dit hoofdstuk verder uiteengezet.

De vorm van interacties tussen mensen is volgens het JD-R model een belangrijk gegeven bij het kwalificeren van een bevolegen identiteit (Schaufeli & Bakker 2001). Het JD-R model biedt vanuit een positieve psychologische invalshoek een onderbouwing van een bevolegen identiteit. Echter geeft het geen zicht in de wijze waarop een bevolegen gedrag zich binnen een digitaal medium als mail weerhoudt. Voor het in kaart brengen van dit gedrag (interactiepatronen) binnen mail conversaties wordt in dit onderzoek gebruik gemaakt van de Two-step flow theorie van Elihu Katz en Paul Felix Lazarsfeld (1955). Deze theorie stelt dat er een 'multistep flow of information' plaatsvindt. Een interactiepatroon waarbij informatie van massamedia naar opinieleiders gaat en vervolgens naar het algemene publiek (Katz en Lazarsfeld 1955, 32). Meerdere interactiepatronen vormen samen een sociaal netwerk, met zwakke en sterke niveaus van beïnvloeding. Voorliggend onderzoek impliceert dat één-op-één mailinteracties, net zoals massacommunicatie, een sociaal mailnetwerk vormen. Een gegeven waar de Two-step flow theorie op gebaseerd is (Katz & Lazarsfeld 1955). Tyler et al. (2003) toonde reeds aan dat mailinteracties een sociaal netwerk vormen met formele en informele gemeenschappen. Dit zijn gemeenschappen met verbanden die zich buiten bestaande bedrijfsculturen en -structuren om vormen en gevoed worden door overeenkomende interesses en werkzaamheden. Hiërarchie speelt hierbinnen een toonaangevende rol. Ondanks dat massamedia in dit onderzoek geen katalysator is, biedt de Two-step flow theorie (Katz & Lazarsfeld 1955) en voorgaand onderzoek (Tyler et al. 2003) handvatten bij het analyseren van informatiestromen. Deze analyse poogt vanuit die theorie sociale structuren bloot te leggen die kunnen duiden op een mate van medewerkersbevolegenheid.

De Two-step Flow theorie wordt vaak gehanteerd binnen nieuwe media onderzoeken om de invloed van massamedia op de maatschappelijke opinie en invloedrijke figuren te kwalificeren. Zo heeft voorgaand onderzoek van Shaomei Wu, Jake M. Hofman, Winter A. Mason en Duncan J. Watts (2011) de Two-step Flow theorie toegepast bij het aantonen van mediators binnen Twitter conversaties. Binnen de Two-step Flow theorie zijn opinieleiders mediators die meningen binnen een netwerk verspreiden en vormen. Schaufeli en Bakker (2001) tonen aan dat een rol als mediator past

binnen een bevlogen medewerkersprofiel. Katz en Lazarsfeld (1955) stellen dat er geen eenduidige groep opinieleiders is te formuleren. Dit hangt af van het onderwerp of issue waarover men communiceert en de participanten die hieraan deelnemen. Wu et al. (2004; 2011) bevestigen dat binnen mailgemeenschappen een soortgelijke verhouding is met onderwerp of issue gerelateerde opinieleiders. Informatieverspreiding en gemeenschapsvorming per mail is het effectiefst binnen een netwerk met dezelfde interesses (Wu et al. 2011). Henry Jenkins, Sam Ford en Joshua Green (2013, 1-46) voegen hier nog elementen als een stabiel netwerk, publieke betrokkenheid, participatiemilieu, waarde van beoordeling en transnationale stromingen binnen deze verschijnselen aan toe.

In het voorliggend onderzoek uiten mediators zich in het verspreiden van informatie per mail onder een grote hoeveelheid collega's om zo hun houding ten aanzien van een werkgerelateerde onderwerp te beïnvloeden (Wu et al. 2011). Nancy Fraser (1990) continueert en stelt dat er binnen dit soort netwerken een *weak* en *strong audience* gedefinieerd kan worden. De *weak audience* zijn participanten aanwezig in het mailverkeer, maar maken verder geen deel uit van het beslissingsproces zoals *strong audience* doet. *Strong audience* bezit de macht om netwerken te vormen en zo opinies en groep-dynamieken te beïnvloeden. Het vormen van een netwerk, zoals de *strong audience* poogt te doen, met een positieve houding kan duiden op een indicator van bevlogenheid. Met dit gegeven zal deze analyse zich focussen op de *strong audience* binnen het mailnetwerk. In §4.3 wordt dieper ingegaan op de gehanteerde meetinstrumenten gebaseerd op een combinatie van het JD-R model en Two-step Flow theorie.

---

## 4 Data en methode

Voor het kritisch analyseren van de mate en wijze waarop mail de kernfactoren toewijding en vitaliteit kwalificeert, is een datagedreven empirisch onderzoek uitgevoerd. De analyse focust zich op tekstuele elementen en interactiepatronen uit maildata die kunnen wijzen op indicatoren van een bevlogen identiteit. In dit hoofdstuk wordt toegelicht hoe de maildata zijn benaderd en geanalyseerd, en wat hier de mogelijke gebreken van zijn. In §4.1 wordt de procedure van het onderzoek toegelicht, waarna in §4.2 inhoudelijk wordt stilgestaan bij de participanten en beschikbare data. Tenslotte wordt in §4.3 het meetinstrument geïntroduceerd en per hypothese beschreven hoe deze onderbouwd of verworpen is.

Bij het opstellen van de methode heb ik mij verdiept in het boek *Digital Methods* van Richard Rogers (2013). Rogers focust zich op interactiepatronen en achterliggende drijfveren binnen websites, zoals sociale netwerk platformen. Zijn handvatten zijn bruikbaar voor dit onderzoek doordat mailverkeer een vorm van digitale interactie is. Rogers (2013) impliceert dat het internet een andere vorm van onderzoeken met zich meebrengt, waardoor methodes zich moeten aanpassen aan een nieuw paradigma. Dit houdt in dat dit onderzoek het bestaansrecht van het JD-R model, dat niet over een datagedreven basis beschikt, en de gedateerde Two-step Flow theorie kritisch analyseert.

### 4.1 Procedure

Voor het beantwoorden van de hoofdvraag is een corpusanalyse vormgegeven (Biber et al. 1998). Binnen een corpusanalyse wordt kwalitatief en kwantitatief onderzoek verricht op een grote hoeveelheid feitelijk voorkomend taalmetaal (Biber et al. 1998). Volgens Douglas Biber et al. (1988) wordt deze empirische vorm van onderzoek nog betrouwbaarder door het gebruik van technologie. Het taalmetaal in deze analyse is de mailtekst van Zwolse ambtenaren. Vasthoudend aan de gedachte van Rogers (2013) is een corpusanalyse een gedateerde onderzoeksmethode die zich moet aanpassen aan het digitale medium mail. Om deze reden zijn er hedendaagse onderzoeksmethodes binnen deze mailcorpus uitgevoerd. Zoals het geautomatiseerd verzamelen van bruikbare informatie uit een grote hoeveelheid tekstmetaal met behulp van text mining technieken. Vervolgens zijn er op basis van tekstkenmerken met een clusteranalyse groepen geclassificeerd en mails gerankt. Waarna een geautomatiseerde sentiment bepaling op systematische wijze emoties aan teksten heeft toegekend. Tenslotte zijn netwerken met digitale technieken tot inzichtelijke weergaven gevisualiseerd. In §4.3 worden deze hedendaagse onderzoeksmethode verder uiteengezet en toegelicht hoe ze binnen dit onderzoek zijn uitgevoerd.

### 4.2 Participanten en data

De corpus van dit onderzoek bestaat uit 4.964 unieke mails (teksten) uit de postvak-in van 8 Zwolse ambtenaren, waarin totaal 3.343 unieke personen zijn achterhaald. De mails zijn in het afgelopen jaar verstuurd en/of ontvangen waardoor de participanten niet berekend zijn op het onderzoek en dit de uitkomsten niet beïnvloed. Door de kleine participatiegroep schetsen de bevindingen echter geen representatief beeld van de mate waarop Zwolse ambtenaren bevlogen in hun werk staan.

Binnen dit onderzoek is gekozen voor het analyseren van de postvak-in map, omdat hierin 17.955 interactiepatronen te herkennen zijn. Interacties zijn het direct en indirect (als CC) ontvangen of versturen van mails. In de dataset is geen onderscheid gemaakt in gewicht tussen directe en indirecte mails. De postvak-in bevat een dusdanig grote hoeveelheid interacties doordat deze map mails en de daarop ontvangen reacties beschikt. De omvang van de postvak-in map verschilt per participant, doordat iedereen zijn of haar inbox op een andere wijze beheert. Denk hierbij aan een clean-desk beheer, waarbij werknemers na afhandelen de mails verwijderen. Dit betekent dat de meegenomen data per participant in omvang verschilt. Persoonlijke normen zijn lastig te meten zonder een kwalitatieve reflectie. Door het beperkt aantal participanten en het verschil in postvak-in beheer zijn de uitkomsten van deze analyse niet generaliseerbaar naar de mate van bevlogenheid van de gehele organisatie, ambtenaren of Nederlandse populatie. Door deze beperking kan slechts de wijze van kwalificatie worden uiteengezet. Daarnaast kan de unieke aard van de mails bevestigd worden, omdat bij een interactie tussen twee personen het antwoord op de voorgaande mail wordt meegestuurd. Dit resulteert in langere mailteksten waarvan enkel de bovenstaande interactie daadwerkelijk uniek is. Ondanks dat wordt de totale tekst in deze analyse als een unieke mail gerankt.

In verband met privacy overeenkomsten met de gemeente Zwolle is voorgenomen de dataset te anonimiseren en de hoofdtteksten niet te publiceren. De dataset is geanonimiseerd door alle afzenders, ontvangers en namen in de hoofdttekst automatisch te coderen naar een uniek nummer. Op deze wijze kunnen interactiepatronen achterhaald worden, maar kan de analyse niet ingaan op de persoonlijke achtergrond (zoals geslacht, functie etc.) van de participant. Het anonimiseren van de dataset heeft vele limitaties voor de uiteindelijke poging tot kwalificatie van medewerkersbevlogenheid. Zo is het in dit onderzoek niet meer mogelijk de resultaten te verifiëren met situaties uit de fysieke wereld, zoals bestaande bedrijfscultuur of -structuren. Hierdoor is de verhouding tussen interne en externe mailnetwerken ook niet aantoonbaar. Deze kritieken voor het totaal anonimiseren van de dataset wordt in hoofdstuk 5 verder uiteengezet. Wegens transparantie van dit onderzoek is het van belang de data wel beschikbaar te stellen, waardoor de geanonimiseerde dataset terug te vinden is op Dropbox.<sup>2</sup>

Samengevat is de data bij het verzamelen en opschonen reeds bewerkt voordat de analyse zelf heeft plaatsgevonden. Ook deze manipulaties hebben gevolgen voor de interpretatie van de resultaten uit hoofdstuk 5.

### 4.3 Meetinstrument

In de huidige situatie wordt medewerkersbevlogenheid bij de gemeente Zwolle gemeten met enquêtes die opgesteld zijn volgens het JD-R model (Schaufeli & Bakker 2001). De gevalideerde schalen binnen het JD-R model hangen samen met de vooraf bepaalde enquêtevragen. Dit huidige meetinstrument is voor deze analyse niet toepasbaar, doordat dit onderzoek vertrekt vanuit datagedreven inzichten, waarbij de mogelijke uitkomsten vooraf niet bepaald zijn. Deze benadering is

---

<sup>2</sup> De unieke dataset (n=4.964) en de dataset met interacties (n=17.955) is terug te vinden op Dropbox: [bit.ly/1kThQDi](https://bit.ly/1kThQDi). Wegens privacy afspraken is de online versie de hoofdttekst niet weergegeven.

een nieuwe methode binnen het vaststellen van medewerkersbevlogenheid, waardoor er tot op heden geen schaal ontwikkeld en gevalideerd voor bevlogen indicatoren uit mailinteracties en teksten. De definitie van vitaliteit en toewijding binnen het JD-R model zijn niet specifiek genoeg om één op één uit maildata te kwalificeren. Om de uitkomsten van de maildata-analyse meetbaar te maken zijn er voor dit onderzoek, voortbordurend op de bevlogenheid theorie van Schaufeli en Bakker (2001) en Two-step Flow theorie van Katz en Lazarsfeld (1955), 5 hypothesen opgesteld:

1. Een bevlogen medewerker is betrokken bij werkgerelateerde onderwerpen en mailt hier frequent over naar collega's.
2. Een bevlogen medewerker mailt meer positief over werkgerelateerde onderwerpen, dan negatief.
3. Een bevlogen medewerker heeft een sterke, informele en positieve relatie met collega's.
4. Een bevlogen medewerker heeft door zijn/haar vitaliteit en toewijding geen 9-5 mentaliteit en mailt daarom ook buiten werktijd.
5. Een bevlogen medewerker is een positieve mediator en beïnvloedt door een sterke betrokkenheid en groot bereik veel collega's.

Uit de resultaten in hoofdstuk 5 zal blijken of deze hypothesen een afdoende inzicht bieden in de kernfactoren van medewerkersbevlogenheid en wat de limitaties zijn. Om tot dit gegeven te komen zijn er binnen de corpusanalyse verschillende onderzoeksmethodes uitgevoerd en gecombineerd. In de komende subparagrafen wordt toegelicht hoe deze hypothesen uit het JD-R model zijn voortgekomen en op welke wijze de mailanalyse deze poogt aan te tonen.

#### *4.3.1 Frequent mailen over een werkgerelateerde onderwerp*

Schaufeli en Bakker (2001, 245) stellen dat toewijding betrekking heeft op een sterke betrokkenheid bij het werk. Dit kan zich binnen het medium mail uiten in het vaak verzenden, ontvangen en doorsturen van mails over een bepaald werkgerelateerde onderwerp. Deze interactie kan plaatsvinden tussen twee of meerdere personen, waaruit aangenomen mag worden dat deze personen een bepaalde vorm van betrokkenheid met het onderwerp voelen. Hoe betrokkenheid zich buiten het medium mail vormt of per persoon verschillend tot uiting komt, wordt in deze analyse niet meegenomen.

Bij het toekennen van onderwerpen aan mails wordt gebruik gemaakt van text mining met een bijpassend label raamwerk. In bijlage A is dit raamwerk terug te vinden. Het toekennen van labels is geautomatiseerd uitgevoerd met de statistische analyse software Rstudio.<sup>3</sup> Dit raamwerk is opgesteld volgens het vierfasemodel van Fred Wester en Vincent Peters (2004). Dit model omvat een exploratie-, specificatie-, reductie- en integratiefase. In de exploratiefase is het raamwerk met bijhorende labels opgesteld. Binnen de specificatiefase wisselt de dataverzameling en analyse elkaar af, waardoor een continue reflectie op het raamwerk plaatsvindt. De opeenvolgende fases vinden later

---

<sup>3</sup> Hiermee is een custom-script geschreven die termen uit het raamwerk detecteren in mails en vervolgens de voorinstelde labels toekennen aan de tekst. Wegens bedrijfsredenen is het script voor niet openbaar beschikbaar gesteld.

in deze analyse plaats.<sup>4</sup> In de exploratie fase zijn de titels voor het toekennen van onderwerplabels van alle mails geanalyseerd op meest frequent voorkomende termen. Hiervoor is de top 50 van de meest voorkomende termen geanalyseerd, waarvan uiteindelijk 35 termen handmatig onderverdeeld zijn in 9 onderwerplabels.<sup>5</sup> Vervolgens zijn 8.996 mails geautomatiseerd gelabeld op onderwerp. Uit een steekproef van 200 mails is gebleken dat de mailtitel een representatieve weerspiegeling biedt van de totale mail.<sup>6</sup> De steekproef is op een aselechte wijze, onafhankelijk en gelijk verdeeld verzameld. Bij mails waar twee onderwerplabels in voorkomen is het eerst geregistreerde label aangehouden. Uit dezelfde steekproef blijkt dat in 90% van de gevallen dit het juiste label is. Ondanks deze hoge nauwkeurigheidsgraad is de methode niet waterdicht en de uitkomsten dus geen feitelijke gegevens.

Om de onderlinge verhoudingen op basis van onderwerpen binnen het mailnetwerk te achterhalen is een netwerkanalyse uitgevoerd. Uit deze analyse wordt duidelijk welke individuen met elkaar communiceren over welk werkgerelateerd onderwerp en met welke frequentie. Om de resultaten inzichtelijk te maken is het netwerk gevisualiseerd met het programma Gephi.<sup>7</sup> In §5.1 wordt kritisch gereflecteerd op deze datagedreven kwalificatie.

#### *4.3.2 Positief gestemde mailcommunicatie*

Voorgaand onderzoek toont aan dat een positieve houding op de werkvloer een belangrijke indicatie is van toewijding (Salanova et al. 2005; Sonnentag 2003). Een positieve houding houdt in dat iemand zijn werk doet om het bedrijf, maar ook zichzelf, te verbeteren (Schaufeli & Bakker 2001). Deze houding heeft volgens Schaufeli en Bakker (2001) niet enkel voor een individu voordelige gevolgen, maar ook voor de groepsdynamiek binnen een organisatie. Een positieve sociale steun wordt in het JD-R model omschreven als energiebron binnen bevlogenheid. Om deze sociale steun te achterhalen binnen mailinteracties is een sentiment analyse uitgevoerd, waarbij gefocust is op een positief sentiment.

Wegens de grote omvang van de dataset worden de teksten niet handmatig gelabeld op sentiment, maar wordt gebruikt gemaakt van de tekst processing API van Mashape.<sup>8</sup> Mashape is een JSON applicatie gebaseerd op Nederlandse boekreviews die teksten screent op termen die een positieve of negatieve stemming bevatten. In dit onderzoek is gekozen voor de Mashape API, omdat dit een van de weinige openbaar beschikbare Nederlandstalige sentiment analyses op grote teksten is. Open source ontwikkelingen vinden momenteel voornamelijk plaats op basis van de Engelse taal. Hoe vaker de Mashape API termen in een tekst herkent, hoe hoger de tekst positief of negatief wordt

---

<sup>4</sup> In de reductiefase is vastgesteld welke label het meest frequent voorkomt en het belangrijkste is in deze analyse. Tenslotte is in de Integratiefase de theorie gekoppeld aan de bevindingen om zo de onderzoeksvragen te beantwoorden.

<sup>5</sup> De overige 15 zijn niet gebruikt wegens de nietszeggende aard. Denk hierbij aan termen als 'weinig' en 'samen'.

<sup>6</sup> In de steekproef (n=200) staat de willekeurigheid van de data centraal. De mails en toebedeelde labels zijn handmatig gecontroleerd om de methode te valideren. De dataset van de steekproef is terug te vinden op Dropbox: <http://bit.ly/1kTi4dy>

<sup>7</sup> Gephi is een interactieve visualisatie software voor het in kaart brengen van interacties. Online: <https://gephi.org>

<sup>8</sup> Online: <https://www.mashape.com/japerk/text-processing>



gerankt.<sup>9</sup> Bing Lui (2012) omschrijft dit als een sentiment analyse op aspect niveau, waarin een stuk tekst een sentiment en een target bevat. De volgende zin fungeert als illustratie: 'Wat een leuke jas!', waarbij 'leuke' het sentiment toekent aan de target 'jas'. Bij het ontbreken van een sentiment of het gelijk oplopen van positief en negatief wordt de mail gerankt als neutraal.

Mashape is geen open source API waardoor de toebedeelde labels in de steekproef (n=200) handmatig geanalyseerd zijn om de representatie van de methode te valideren.<sup>10</sup> Van 200 mails is de hoofdtekst onderzocht op voorkomende positieve of negatieve gestemde termen. Deze termen zijn vergeleken met de door de API toebedeelde label. In bijlage B is dit raamwerk terug te vinden. Uit de steekproef analyse bleek dat de API op de mail dataset een nauwkeurighedsgraad heeft van 45% en 2 keer zo vaak een positief label aangeeft als de handmatige geanalyseerde steekproef (n=200) uitwijst. Positief gestemde woorden als 'vriendelijke' blijken de nauwkeurighedsgraad te verlagen, doordat deze vaak voorkomen in een automatische handtekening als 'Met vriendelijke groet,' en niks te maken hebben met de inhoud van de mail. De limitatie van de nauwkeurighedsgraad van de sentiment bepaling geven een vertekend beeld van de werkelijkheid. Met deze reden is voor het vaststellen van het algemene sentiment de handmatig geanalyseerde steekproef gehanteerd, voor diepgaandere bevindingen is de algehele dataset genomen. Tenslotte zijn de bevindingen samengevoegd in een netwerkanalyse. In §5.2 zijn de resultaten en kritiek op deze kwalificatie verder uiteengezet.

#### 4.3.3 *Formele of informele groepsdynamiek*

De wijze waarop collega's met elkaar acteren kan meer zicht bieden op onderlinge verhoudingen en een mogelijke bevrogen identiteit. Door meer kennis te hebben van de formele/informele en positieve/negatieve verhouding kunnen deze onderlinge mailinteracties beter begrepen worden. Een formele communicatie impliceert meer afstand tussen individuen als informeel contact. Een positieve onderlinge informele houding kan duiden op een goede groepsdynamiek, zoals team of afdeling. Schaufeli en Bakker (2001) stellen dat groepsdynamieken invloed hebben op individuele bevrogenheid en andersom. Tyler et al. (2003) tonen aan dat formele of informele houding binnen mailgedrag kan voortkomen uit bestaande bedrijfscultuur en -structuren. Door het anonimiseren van de dataset zijn deze hiërarchieën niet meer te achterhalen. Onderlinge verhoudingen zijn hierdoor dus niet meegenomen. In §5.3 is uiteengezet wat voor gevolgen dit heeft op de interpretatie van de uitkomsten.

Voor het achterhalen van formele of informele houding binnen mail is de dataset, net zoals bij de bovenstaande indicatoren, gelabeld op basis van text mining. In dit geval is gekeken naar formele of informele aanhef van een mail. In bijlage C is dit raamwerk terug te vinden. Voor dit onderzoek is gefocust op de aanhef doordat deze unieker van aard is dan de afsluiter.<sup>11</sup> Enkel bij het ontbreken van

---

<sup>9</sup> Voor een voorbeeld hoe sentiment analyses met tekst categorisering precies te werk gaat verwijst ik door naar de website: <http://text-processing.com/demo/sentiment/>

<sup>10</sup> Bedrijven als Coosto, Buzzcapture en OBI4wan behalen na menig ontwikkeljaren een nauwkeurighedsgraad tussen de 70% en 80% (Berg 2013). Het toekennen van stemmingen binnen teksten kan nooit 100% accuraat zijn, doordat menselijke elementen als sarcasme of humor niet getraceerd kunnen worden.

<sup>11</sup> De afsluiting bevat vaak een automatische handtekening, in de trant van "Met vriendelijke groet,". Dit vertroebeld de representatie van de resultaten.

de aanhef is gelabeld op de afsluiting van een mail. De bevindingen uit deze analyse zijn vervolgens gekoppeld aan de inzichten uit de sentiment analyse. Tenslotte is de lengte van een mail vastgesteld op basis van het aantal woorden en de resultaten gebundeld in de hoop een correlatie te constateren.

#### *4.3.4 Mailen buiten werktijd*

Op de kernfactor absorptie wordt binnen deze analyse niet gefocust, omdat het mogelijk voortvloeit uit vitaliteit en toewijding. Schaufeli en Bakker (2001, 245) stellen dat absorptie betrekking heeft op een plezierige wijze van opgaan in je werk en daarmee de tijd te verliezen. Vitaliteit is volgens hen onvermoeibaar met werk door kunnen gaan. Kenmerken van absorptie en vitaliteit kunnen binnen mail herleid worden door mensen te traceren die buiten werktijd nog actief aan het mailen zijn, doordat ze zich positief betrokken voelen bij hun werk. Voor het achterhalen van deze mailindicator op basis van tijdsaanduiding en de positieve toon van mails is een soortgelijke text mining clustermethode toegepast als hierboven. Enkel is verschilt het raamwerk voor het toekennen van het label 'buiten werktijd'. Dit label is bepaald op basis van de tekst achter de aanduiding "verzonden:", zoals "dinsdag 28 april 2015 9:25". Eerst zijn de berichten gelabeld als 'doordeweeks' of 'weekend' op basis van de vermelde dagen. Vervolgens zijn de berichten gelabeld als 'buiten werktijd' wanneer ze doordeweeks vóór 8:30 uur of na 18:00 uur en in het weekend zijn verzonden. Tenslotte zijn deze bevindingen aan de resultaten uit de text mining, netwerk- en sentiment analyse in een poging correlaties te achterhalen.

Doordat dit onderzoek is gestart vanuit een positieve psychologische invalshoek, is in de methode geen rekening gehouden met het tegenovergestelde van bevoegenheid, namelijk burn-out. Het mailen buiten werktijd kan een indicator zijn voor een te hoge werkdruk, wat volgens het UBOS na langdurig aanhouden kan leiden tot een burn-out (Schaufeli & Bakker 2001). Deze beperking is in §5.4 meegenomen in de kritische reflectie van de resultaten.

#### *4.3.5 Mediatoren*

Salanova et al. (2003), Schaufeli en Bakker (2001) stellen dat een bevlogen identiteit een mediërende rol op zich neemt. De Two-step Flow theorie van Katz en Lazarsfeld (1955) en eerder onderzoek (Tyler et al. 2003; Wu et al. 2004; 2011) impliceren dat een mediator binnen het medium mail zich uit in het frequent sturen, ontvangen en delen van informatie onder een groot bereik van de Zwolse collega's. Een bevlogen mediator doet dit vanuit een betrokken houding en in positieve relatie tot werktaken en collega's (Schaufeli & Bakker 2001). Voor het achterhalen van mediators binnen het mailverkeer is een netwerkanalyse uitgevoerd. Tyler et al. (2003) hebben aangetoond dat mediators zich voornamelijk in het centrum van het netwerk bevinden. Deze bevindingen zijn vervolgens gekoppeld aan de uitkomsten van de text mining- en sentiment analyse. Hierdoor kunnen mogelijke correlaties achterhaald worden of waarover, wanneer, met welke houding en sentiment een mediator mailt met elkaar in verband staat. Door het anonimiseren van de dataset kan de precieze functie van de Zwolse professional niet achterhaald worden. De assumptie dat mail-mediators een verhoogde kans hebben op een bevlogen identiteit wordt in §5.5 kritisch benaderd.

---

## 5 Bevlogenheid kwalificeren uit maildata

In dit hoofdstuk worden de resultaten van de corpusanalyse gepresenteerd en de hypothesen over mailindicatoren van bevlogenheid aangenomen of verworpen. In dit proces worden de datafication assumpties (§1.2) en beperkingen van de kwalificatie methode (§1.2) afgezet tegen de uitkomsten. Het biedt antwoord op de hoofdvraag of een datagedreven analyse van mailgedrag een bevlogen identiteit, met kenmerken van vitale en toegewijde, kan kwalificeren. §5.1 presenteert de mate van betrokkenheid via mail per werkgerelateerde onderwerp. In §5.2 wordt aangetoond of deze betrokkenheid positief geaard en daadwerkelijk onderwerp gerelateerd is of dat dit ligt aan de relatie met een collega. Vervolgens toont §5.3 aan dat er geen correlatie is tussen de labels mail lengte, formaliteit en sentiment. In §5.4 worden de resultaten over mailen buiten werktijd uiteengezet. Tenslotte komen in §5.5 alle bevindingen samen wanneer ingezoomd wordt op mediators binnen het mailnetwerk.

### 5.1 Betrokken bij werkgerelateerde onderwerpen

Uit de analyse blijkt dat medewerkers van de gemeente Zwolle het meest mailen over bijeenkomsten (30,1%), factureren (19,6%), organiseren (12,3%) en onderzoeken (11,2%). In figuur 2 (pagina 21) is het netwerk van mailinteracties geclusterd op onderwerpen weergegeven.<sup>12</sup> De grootte van de bollen geeft weer hoe vaak iemand mails heeft ontvangen of verstuurd. De dikte van de lijnen geeft weer hoe vaak er gemaaild is tussen twee personen (bollen).

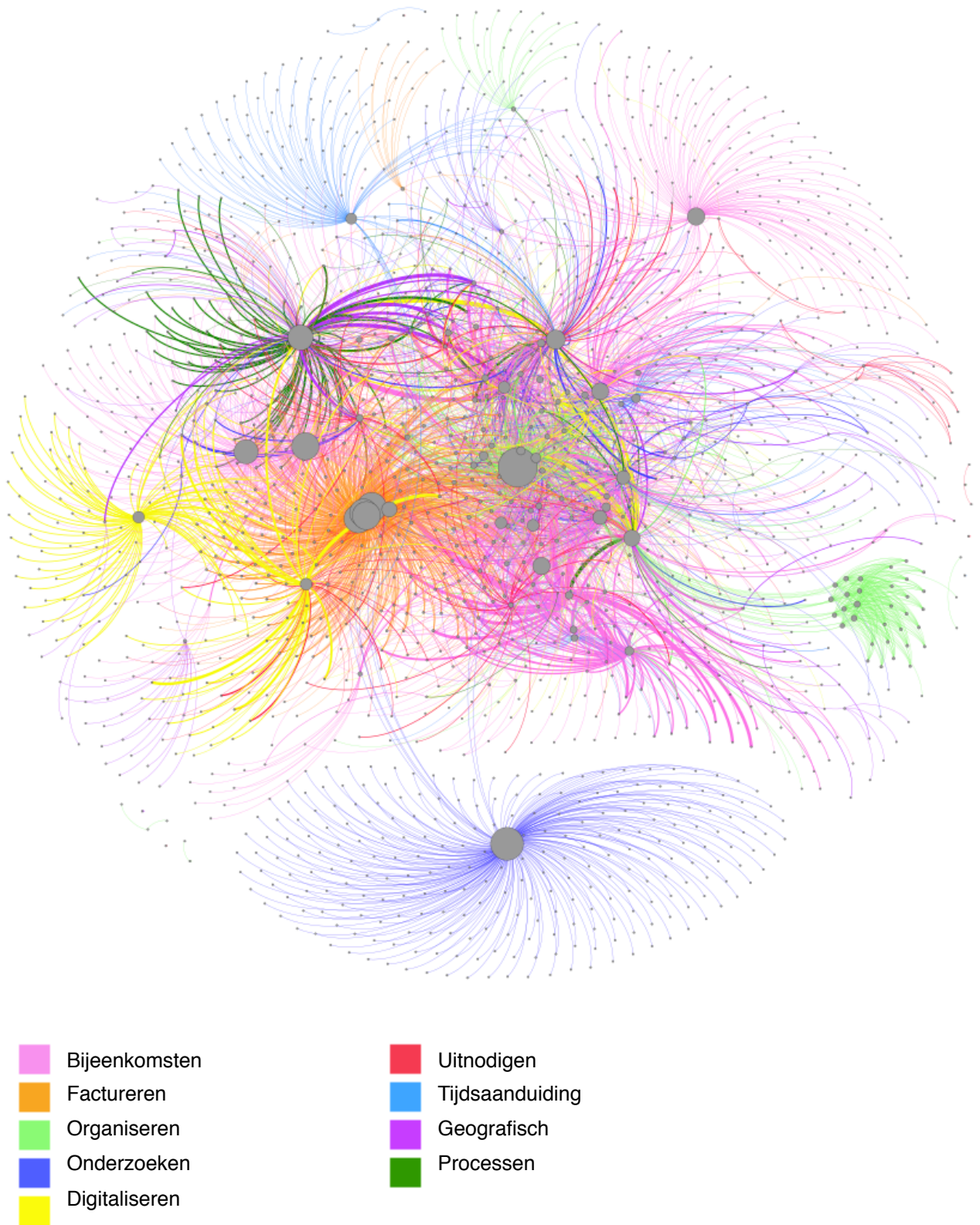
Het netwerk veronderstelt dat het mailverkeer over factureren (oranje) veelal bij vier personen ligt. Dit kan duiden op een betrokkenheid bij factureren vanuit persoonlijke interesse dan wel functieomschrijving. Door beperkingen van de onderzoeksmethode, zoals het anonimiseren van de dataset, kan niet aangetoond worden of deze situatie daadwerkelijk gerelateerd is aan een functie of de bedrijfsstructuren. De context van de situatie ontbreekt, doordat enkel wordt uitgegaan van de data gedreven bevindingen (Kitchen 2014). De situatie kan zijn dat een medewerker veel mails nodig heeft bij het overbrengen van de boodschap of juist een clean-desk beleid hanteert, waardoor mails ontbreken.

Uitgaande van de netwerkvisualisatie bevinden zich in de buitenste ring medewerkers voornamelijk mailen over één onderwerp, zoals onderzoeken (donkerblauw) en organiseren (lichtgroen). In de visualisatie uit zich dit in een bol (persoon) met veelal één kleur (onderwerp). De dataset stelt dat zij meer mails verzenden (Outdegree) dan ontvangen (Indegree), wat erop duidt dat hun mails niet altijd worden beantwoord of deze personen veel in de gebiedende wijs mailen. De data-analyse assumeert dat deze medewerkers per mail over bijvoorbeeld onderzoeken een eenzijdige relatie hebben met collega's. Ondanks deze datagedreven bevindingen kan een hoge werkdruk of voorkeur voor face-to-face communicatie een reden zijn van het niet beantwoorden van mails. Critici, waaronder Boyd & Crawford (2012) en Lazer et al. (2014), stellen dat een databron als mail alleen

---

<sup>12</sup> Om de clusters binnen deze visualisatie overzichtelijk weer te geven is in Gephi het algoritme Fruchterman Reingold (Degree) gehanteerd.

geen macht heeft tot spreken. Bovenstaande context gerelateerde uitkomsten kunnen niet in deze data-analyse worden aangetoond, doordat een reflectief en kwalitatief veldonderzoek ontbreekt.



Figuur 2. Netwerkvizualisatie van verdeling onderwerpen per persoon en de onderlinge verbanden (n=17.955).

Waar het labelen op basis van het raamwerk (figuur 2) uitgaat van 9 onderwerp gerelateerde groepen, kent de computer op basis van modularity 21 groepen toe aan het Zwolse mailverkeer.<sup>13</sup> Van deze groepen omvatten 6 in totaal 96,81% van het gehele netwerk. Dit houdt in dat voornamelijk 6 clusters van personen zijn met veel overeenkomende mailcontacten, doordat ze mogelijk dezelfde interesses hebben. De data toont aan dat de modularity verdeling niet overeenkomt met de onderwerp clusters van figuur 2. Uitgaande van Tyler et al. (2003) kunnen dit informele gemeenschappen zijn, maar door het anonimiseren is dit niet aantoonbaar. Bedrijfsstructuren en functieomschrijvingen kunnen niet achterhaald worden, waardoor niet duidelijk is of de 6 gemeenschappen ontstaan zijn door betrokkenheid bij werkgerelateerde onderwerpen en verantwoordelijkheden.

In het midden van het netwerk, afgebeeld in figuur 2, is een cluster van medewerkers weergegeven die frequent lijken te mailen en hierop relatief gezien veel reacties ontvangen (Indegree). In de visualisatie zijn dit de dikkere lijnen met verschillende onderwerp gerelateerde kleuren. Voortbordurend op het JD-R model van Schaufeli en Bakker (2001) en de *datafication* assumptie dat data beschikt over de waarheid (Boyd & Crawford 2012; Kitchen 2014), zou dit betekenen dat deze medewerkers een hogere betrokkenheid bij een meer uiteenlopend interessegebied hebben, dan personen die zich in de buitenring bevinden. De datavisualisatie zou de mate waarop Zwolse medewerkers per mail in relatie staan tot hun collega's aantonen en de mogelijkheid dat zij betrokken zijn bij werkgerelateerde onderwerpen. Door de vele beperkingen van dit onderzoek toont mail op zichzelfstaand echter geen gedegen causaliteit aan of frequent mailen in alle gevallen duidt op een hoge medewerkersbevlogenheid. Betrokkenheid als kenmerk van toewijding (kernfactor van medewerkersbevlogenheid) is een subjectief gegeven (Gitelman 2013). Dit houdt in dat betrokken gedrag per individu en medium anders tot uiting komt (McLuhan 1967). Frequent mailen duidt hierdoor niet per definitie op bevlogenheid.

## 5.2 Positieve betrokkenheid is niet enkel onderwerp gerelateerd

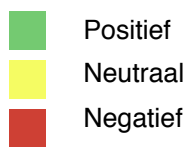
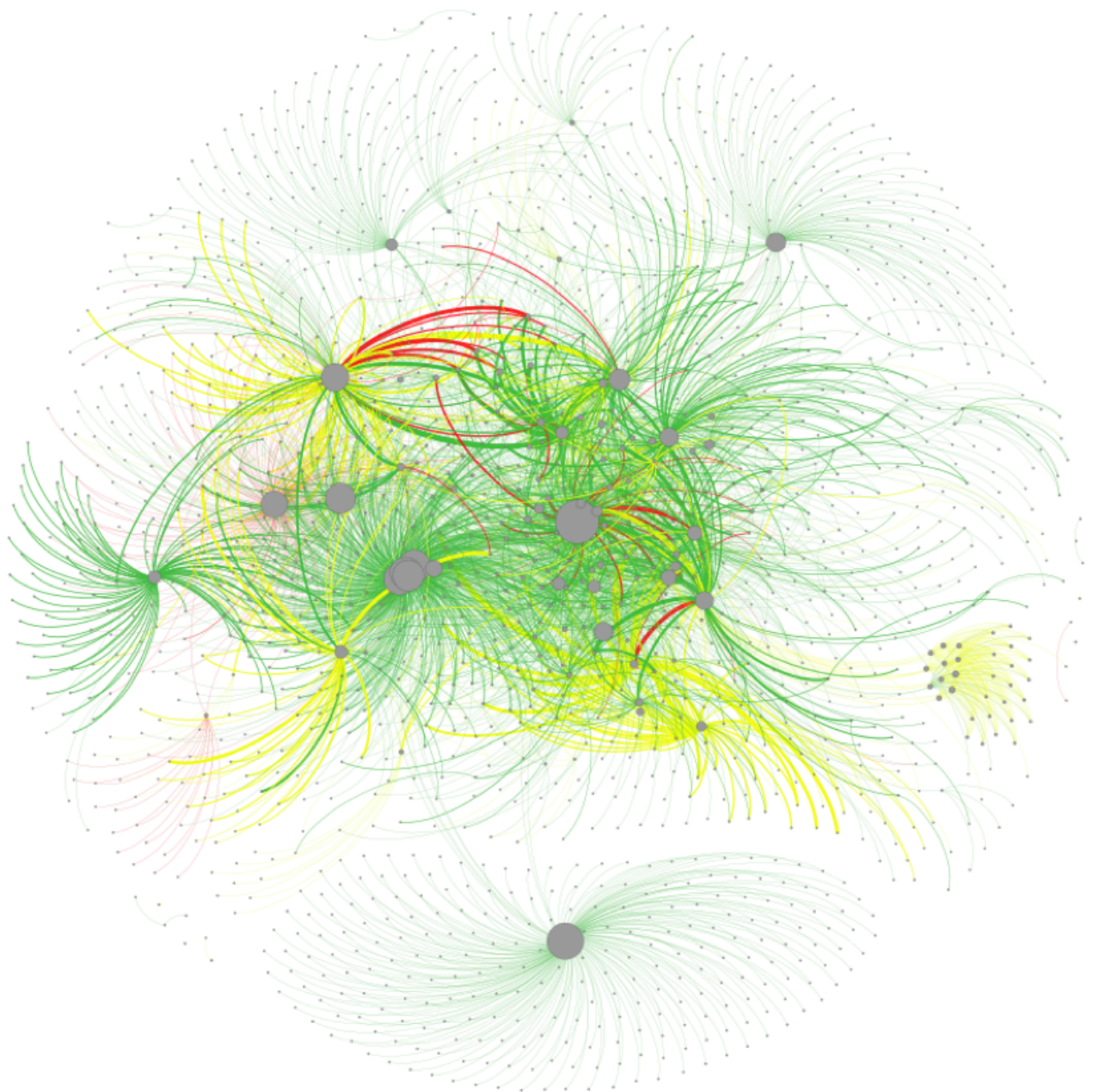
Naast de betrokkenheid op onderwerpniveau is het voor dit onderzoek relevant het netwerk op basis van sentiment in kaart te brengen. Het sentiment geeft aan met welke houding een individu of groep per mail betrokken is bij een bepaald onderwerp, waarbij een positieve annotatie een mailindicator is van bevlogenheid. Uit de steekproef (n=200) blijkt dat Zwolse medewerkers zich over het algemeen per mail 31% positief, 63% neutraal en 5% negatief uiten. Hierbinnen is men het meest positief betrokken bij bijeenkomsten en het negatiefst bij interne digitalisering. In hoeverre betrokken mailgedrag daadwerkelijk duidt op een bevlogen identiteit is een kritisch begrip. Lazer et al. (2014, 1204-1205) omschrijven deze assumptie als de *Big Data Revolution*, waarin de menselijke context achterwegen wordt gelaten, doordat enkel uit wordt gegaan van grote eenduidige datasets. Hetgeen is van toepassing is in de huidige studie, waarbij geanonimiseerde maildata geanalyseerd wordt op indicatoren van medewerkersbevlogenheid. Een kwalitatief reflectieonderzoek voor het bevestigen van de dataresultaten ontbreekt (Kitchen 2014).

---

<sup>13</sup> Statistisch modularity gewicht van 1.5

In figuur 3 (pagina 24) is een soortgelijke netwerkvisualisatie (n=17.955) weergegeven, als op de voorgaande pagina, echter zijn de kleuren hier gebaseerd op sentimenten. In combinatie met de inzichten uit figuur 2, veronderstelt figuur 3 dat medewerkers in de buitenste ring bij mailen veelal één sentiment hanteren, gerelateerd aan één onderwerp. Hierbij moet de relatief lage nauwkeurigheidsgraad van de sentiment bepaling, zoals in §4.3.2 nader is beschreven, kritisch in overweging worden genomen. De data impliceert dat de buitenste ring een eenzijdig mailprofiel heeft, waarin veel berichten op één wijze worden verzonden met weinig effect op het netwerk. Vanwege hun lage beïnvloedingsniveau wordt dit profiel door Fraser (1990) omschreven als *weak audience*. Deze bevindingen tonen niet aan of deze profilering ook geldt voor situaties buiten mailen. Dit is wederom te danken aan de eenduidigheid van de data en het ontbreken van menselijke context (Boyd & Crawford 2012; Lazer et al. 2014; Kitchen 2014). De medewerkers in het centrum van de netwerkvisualisatie mailen over diverse onderwerpen met zowel een positieve, negatieve of neutrale houding. Uit de visualisaties blijkt dat de gecentreerde groep medewerkers in hun mailinteracties geen eenduidig sentiment per onderwerp hanteren. Dit geeft aan dat de positieve of negatieve houding eerder gerelateerd is aan personen binnen het netwerk dan aan het onderwerp. In §5.5 wordt deze groep benoemd tot *strong audience* en verder uiteengezet, omdat ze meer mailkenmerken van bevlogenheid blijken te tonen als andere.

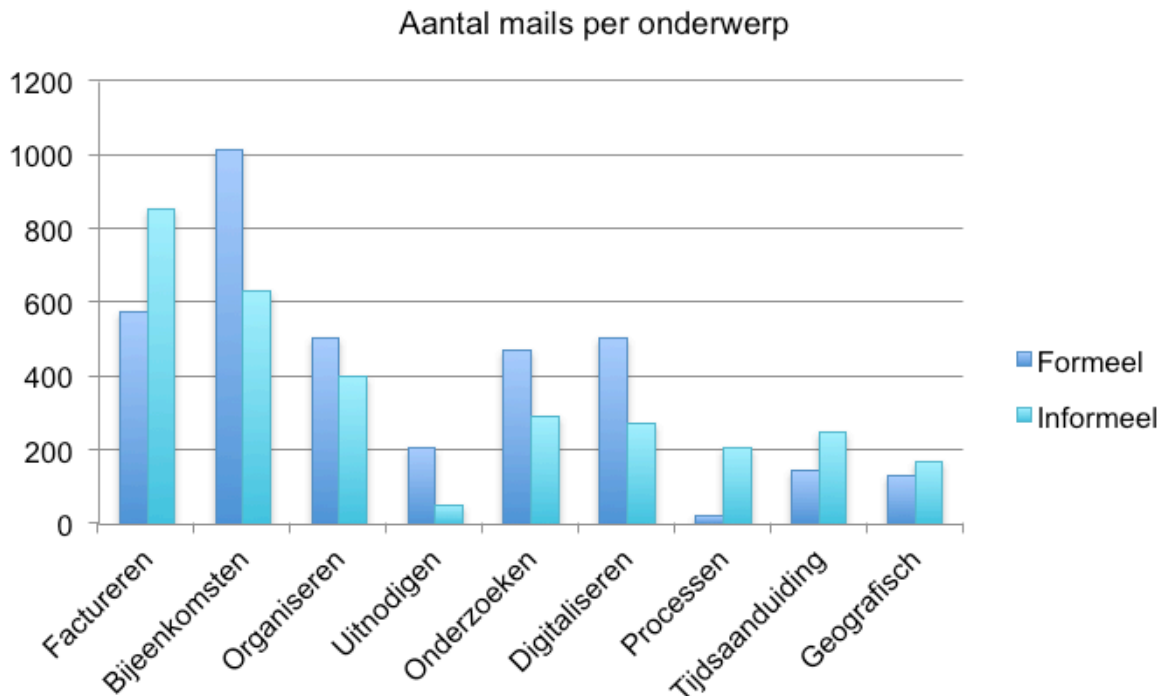
Naast de kritiek op het verkrijgen van de inzichten, zoals de eenduidigheid van de data, het verschil in clean-desk beleid per medewerker, de lage nauwkeurigheidsgraad van de methode en het anonimiseren van de dataset, presenteren ook de netwerkvisualisaties een vertekend beeld. De datavisualisaties zijn volgens critici subject en suggestief van aard (Paul 2007; Gitelman 2014), Enerzijds door het analyseproces van de computer en analist die bepaalde keuzes maken in de wijze waarop de resultaten worden getoond (Paul 2007). De gemaakte keuzes kunnen een lezer sturen bij het interpreteren van de bevindingen in een bepaalde richting. Anderzijds heeft de wijze waarop een gebruiker de resultaten interpreteert effect op de representatie van de resultaten (Gitelman 2013; Seth et al. 2014). De menselijke framing verschilt per persoon en overstijgt geen data (Kitchen 2014). Dit interpretatieproces moet bij het beoordelen van deze analyse in overweging worden genomen.



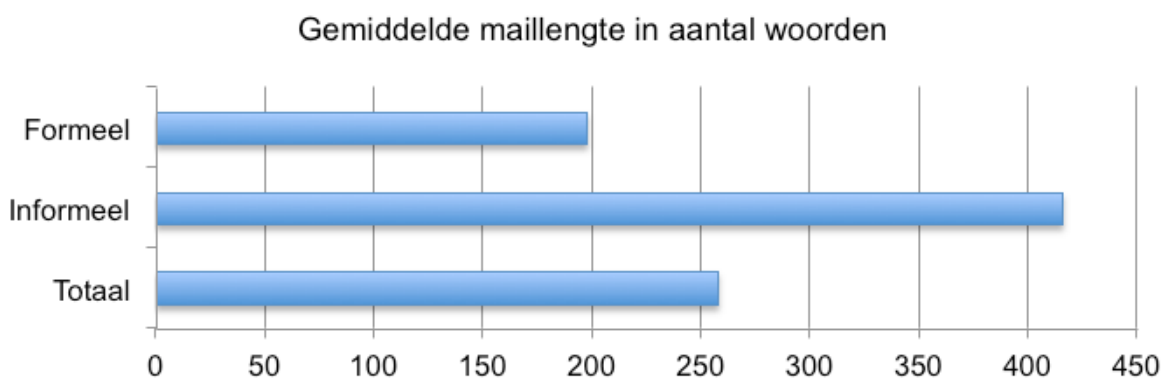
Figuur 3. Netwerkvizualisatie van verdeling sentimenten per persoon en de onderlinge verbanden (n=17.955).

### 5.3 Maillengte, formaliteit en sentiment

Om de mailinteracties en mogelijke indicaties van bevoegenheid nog beter te begrijpen zijn de berichten geanalyseerd op lengte, formaliteit en sentiment. In figuur 4 is het aantal mailberichten per toon gevisualiseerd en in figuur 5 is de gemiddelde maillengte in woorden per toon af te lezen.



Figuur 4. Formele of informele toon binnen mailverkeer gesorteerd op onderwerp (n=4.964).

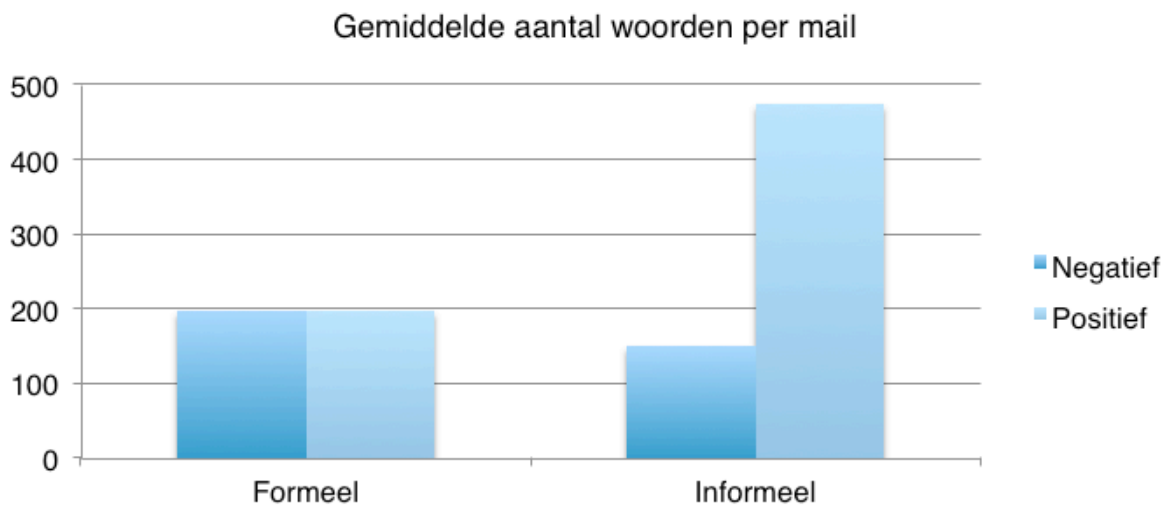


Figuur 5. Gemiddelde lengte van formele, informele en alle mails (n=4.964).

Opmerkelijk is de informele piek bij het label factureren (figuur 4) en het hoge gemiddelde bij informele maillengte (figuur 5). Aangenomen wordt dat de interne mail "Facturen < €1.000 binnen 28 dagen betalen" van 1.643 woorden lang hier invloed op heeft gehad. Drie toonaangevende medewerkers hebben in totaal 836 keer deze mail verspreid. Dit mailgedrag heeft geresulteerd tot een bereik van



210 Zwolse professionals. In §5.5 wordt dieper ingegaan op de mediators binnen het mailnetwerk en hun mate van bevlogenheid. De bovenstaande cijfers geven weer dat wanneer de mail over factureren niet meegerekend wordt, de verhouding tussen formeel/informeel gelijk waardig is.



Figuur 6. Gemiddelde aantal woorden binnen mailverkeer gesorteerd op toon en sentiment (n=4.964)

Figuur 6 geeft de combinatie van de drie labels over de lengte, formaliteit en sentiment van mails weer. Bovenstaande figuren veronderstellen dat medewerkers voornamelijk negatief over interne digitalisering mailen en gebruiken hiervoor gemiddeld de langste berichten, namelijk 336 woorden. De resultaten presenteren een piek bij informele positief gerankte mails met een gemiddelde mail lengte van 473 woorden. De aanleiding hiervan ligt waarschijnlijk bij de eerder omschreven mail over factureren. Om meer te kunnen zeggen over het woorden aantal en zinsopbouw in verbanden met formaliteit en sentiment zal linguïstisch onderzoek moeten plaatsvinden. Taalkundig onderzoek valt buiten de scope van deze analyse, doordat hier gefocust wordt op interactiepatronen.

De data toont aan dat er geen significante correlatie is tussen de gemiddelde lengte van een positieve of negatieve, formele of informele mail. Dit wil zeggen dat de drie factoren samen geen mailindicator of patronen voor het kwalificeren van een bevlogen identiteit tonen, doordat de uitkomsten overeenkomen. De analyse kan niet aantonen of een informele houding voornamelijk wordt gehanteerd binnen positief gestemde mails, wat zou duiden op een samenhang met medewerkersbevlogenheid. De theorie wijst uit dat bestaande verhoudingen binnen een organisatie effect kunnen hebben op de bovenstaande resultaten (Wu et al. 2011). Een belangrijk gegeven voor het kwalificeren van medewerkersbevlogenheid, wat door het anonimiseren niet te achterhalen valt.

#### 5.4 Mailen buiten werktijd als indicator van bevlogenheid

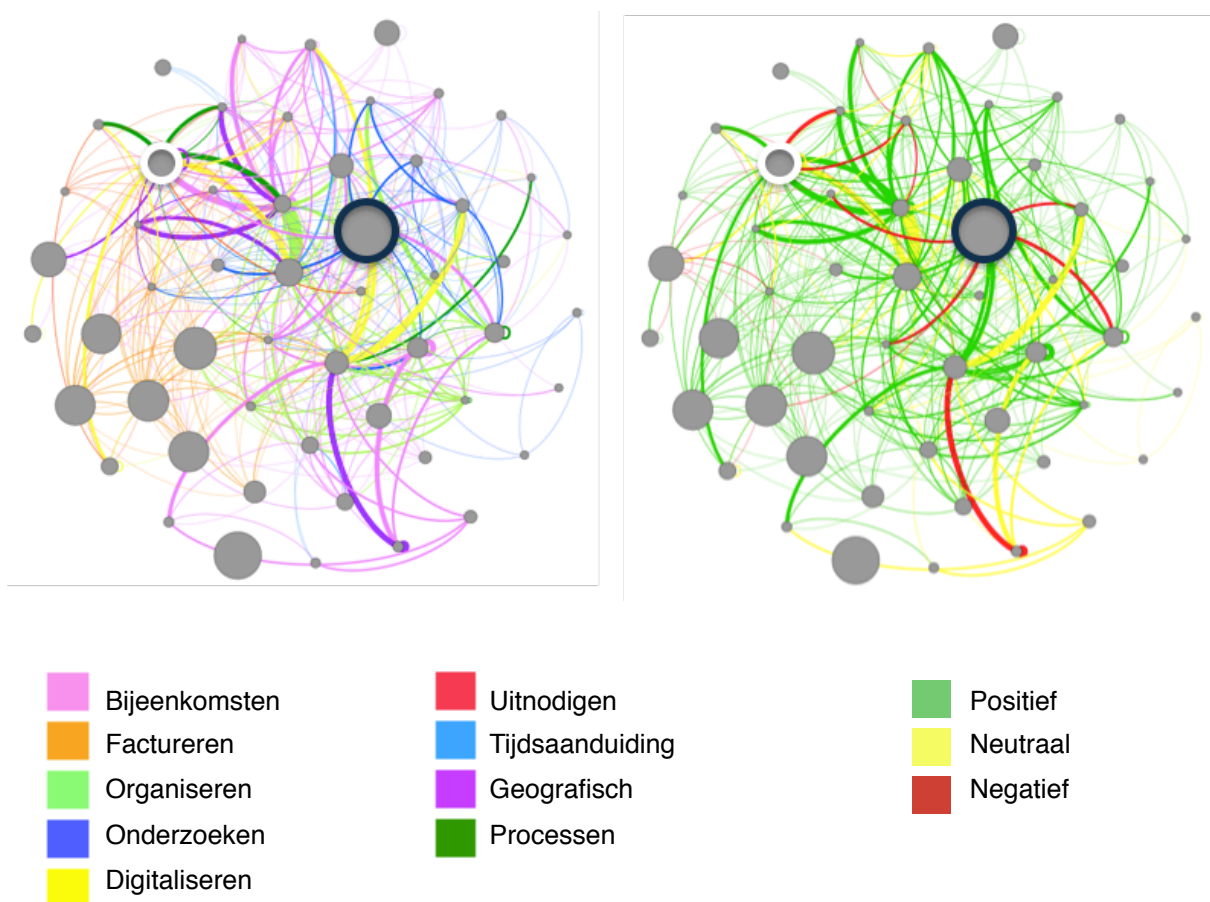
Uit de analyse (n=4.964) blijkt dat 40% van de dataset in de hoofdtekst een verzendtijd van de mail bevat. Van deze mails is het merendeel, namelijk 27% (536 mails), op een donderdag verstuurd. Hiervoor kunnen verschillende redenen zijn, zoals veel afwezig op woensdag en vrijdag of dat door

uitstelgedrag routinematig op donderdag meer mail wordt afgehandeld. Toekomstig kwalitatief onderzoek moet de daadwerkelijk achterliggende reden hiervan aantonen.

In totaal zijn er 459 mails (23%) achterhaald die buiten werktijd, dus voor 8.30 uur, na 18.00 uur en in het weekend, zijn verstuurd. Hierbinnen komen de labels bijeenkomsten (19%) en organiseren (17%) voornamelijk positief in voor. Uit de resultaten gepresenteerd in §5.5 blijkt dat enkele medewerkers die buiten werktijd mailen ook voorkomen in de gecentreerde groep uit figuur 2 en 3. De combinatie van deze mailfactoren lijkt te duiden op een hogere mate van betrokkenheid per mail bij collega's en de onderwerpen bijeenkomsten en organiseren. Ondanks deze bevindingen toont de maildata geen inzicht in de werkelijke reden van mailen buiten werktijd. Naast een positieve bevlogenheid kan de oorzaak ook liggen bij een te hoge werkdruk, wat uiteindelijk kan leiden tot een burn-out (Schaufeli & Bakker 2001).

### 5.5 Mediatoren binnen het mailnetwerk

De datagedreven bevindingen uit §5.1, §5.2 en §5.4 impliceren dat de gecentreerde groep medewerkers sterk betrokken is bij hun werk. In figuur 7 is ingezoomd op dit mailnetwerk, wat door Fraser (1990) omschreven wordt als de *strong audience*.<sup>14</sup>



Figuur 7. Netwerkwisualisatie van meest gemailde personen (n=17.955)

<sup>14</sup> Algoritme: Fruchterman Reingold (Degree). Gefilterd op 30 of meer interacties en toont momenteel 3,81% van de personen uit de totale dataset, n=17955.

Deze visualisaties geven personen weer die 30 mailinteracties of meer hebben gehad met andere figuren binnen het mailnetwerk. Links is het netwerk op onderwerp en rechts hetzelfde netwerk, maar dan op basis van sentiment afgebeeld. De dikke en dunne lijnen in beide visualisaties tonen aan dat een positieve of negatieve mail niet afhankelijk is van frequent of minder vaak mailen tussen twee individuen. Naast de personen die een sterke mailinteractie hebben zijn er ook bollen met dunnere lijnen zichtbaar, wat inhoudt dat zij met een groot bereik veel mails verzenden.

De wit en zwart omcirkelde bollen impliceren de meest toonaangevende participanten in het mailverkeer te zijn, doordat ze het grootste bereik hebben met veel mailinteractie en zo veel personen in hun netwerk kunnen beïnvloeden. De wit omcirkelde persoon heeft van iemand uit de buitenste ring van het netwerk een mail over interne digitalisering ontvangen en vervolgens met een negatieve annotatie onder 146 collega's verspreidt. Het negatieve sentiment verspreid zich slechts één interactie, wat impliceert dat een mediator enkel de stemming van zijn of haar eigen netwerk beïnvloed en daarbuiten geen effect heeft. Verder mailt de wit omcirkelde persoon voornamelijk positief (23%), tegenover 15% negatieve mails, die hij/zij zowel doordeweeks als buiten werktijd over interne processen (32%) en digitalisering (29%) mailt.

Voorgaand analyses (Tyler et al. 2003; Wu et al. 2004) en voorliggend onderzoek tonen aan dat opinieleiders binnen mailverkeer een belangrijke rol innemen voor informatieverbreiding. Zonder reflectief onderzoek is het echter niet mogelijk om vast te stellen of dit gedrag ook duidt op een bevlogen identiteit. Informatiestromen zijn datagedreven te kwalificeren, een subjectieve gemoedstoestand daarentegen niet.

Samengevat bevatten mediators volgens de mailanalyse alle mogelijke aantoonbare indicatoren van medewerkersbevlogenheid. Ze zouden door een frequent mailpatroon een sterke betrokkenheid uiten bij werkgerelateerde onderwerpen of collega's. Echter blijken deze indicaties zwak en patronen niet te herleiden tot een werkelijke bevlogen identiteit, doordat de datagedreven resultaten gebogen gaan onder aantoonbare *datafication* assumpties en limitaties. Critici als Boyd, Crawford (2012) en Kitchen (2014) omschrijven dit als het geloof in een hogere vorm van objectieve intelligentie. Een visualisatie als in figuur 7 impliceert een sterk positief toegewijde en vitale groep Zwolse medewerkers te presenteren, maar voor het aantonen van een causale verband met bevlogen gedrag of gemoedstoestand is nader onderzoek noodzakelijk.

---

## 6 Conclusie en discussie

In dit onderzoek is de mate en wijze waarop een datagedreven kwalificatie van mailgedrag inzicht biedt in de kernfactoren toewijding en vitaliteit van medewerkersbevlogenheid onderzocht. Er is geanalyseerd of frequentie, onderwerpen, sentiment, formaliteit en tijden van mailen indicatoren of patronen van medewerkersbevlogenheid aantonen. De gemeente Zwolle fungeert hierbinnen als casus, omdat zij het *datafication* systeem omarmen en de mogelijkheden van een nieuw meetinstrument verkennen. Concluderend kan gesteld worden dat de een mail-datagedreven analyse op zichzelf staand geen gedegen verklaring biedt voor bevlogen gedragingen, doordat de bevindingen een eenzijdig beeld schetsen. In deze laatste sectie van het onderzoek worden de gebreken van de onderzoeksmethode en gebruikte theorieën uiteengezet. Vervolgens wordt bediscussieert wat deze conclusies voor, zowel de gemeente Zwolle, als academisch onderzoekers betekend en waar vervolgonderzoek zich in de toekomst op moet focussen.

Uit de analyse is gebleken dat, ondanks de vele bevindingen, de onderzoeksmethode ook vele beperkingen kent. Hierdoor is het vaststellen van patronen of indicatoren die wijzen op een mate van medewerkersbevlogenheid niet mogelijk. Ten eerste is de analyse gebaseerd op een kleine participatiegroep van slechts 8 medewerkers van de gemeente Zwolle. Hierdoor schetsen de bevindingen geen representatief beeld van mogelijk bevlogen mailgedrag van alle (Zwolse) ambtenaren. Deze limitatie heeft daarnaast ook effect op het mogelijk ontbreken van o.a. een correlatie tussen de labels lengte, formaliteit en sentiment. Het verder ontwikkelen van deze onderzoeksmethode zal daarom moeten beginnen bij het uitbreiden van de dataset, met meer participanten vanuit verschillende achtergronden, afdelingen en functieniveaus.

Ten tweede is de dataset, die gegenereerd is door de kleine participatiegroep, volledig geanonimiseerd tot unieke nummers. Door het anonimiseren zijn de werkelijke functieomschrijving of bedrijfsstructuren en oorzaken van collegiale verbanden in dit onderzoek niet te achterhalen. Kitchen (2014) wijst ons op de beperking dat data slechts een fractie van de werkelijkheid tonen, doordat de menselijke context ontbreekt. De maildata-analyse toont ons een sample van het werkelijke bevlogen gedrag, doordat de uitkomsten niet herleidbaar zijn tot situaties op de werkvloer en bestaande bedrijfscultuur van de gemeente Zwolle. Door het anonimiseren van de dataset zijn verbanden tussen individuen, maar ook teams of afdelingen binnen de organisatie in dit onderzoek niet te achterhalen.

Het ontbreken van context in data-analyses brengt mij bij de derde beperking van dit onderzoek, namelijk de eenzijdigheid in databronnen. Deze kwalificatie van medewerkersbevlogenheid maakt slechts gebruik van één databron, namelijk mail. Voortbordurend op de uitspraak van McLuhan (1967) "Medium is the message", bepaalt het medium mail in dit onderzoek hoe bevlogenheid vorm krijgt. De interpretatie van mogelijk bevlogen gedragingen zal bij een ander medium, zoals telefoonverkeer, mogelijk anders tot uiting komen. Hieruit kan geconcludeerd worden dat maildata opzichzelfstaand geen recht hebben tot spreken (Kitchen 2014). Data-analyses omvatten nooit een totaal domein, omdat in geen enkele situatie alle informatie beschikbaar is (Lazer et al. 2014). Met deze reden zal toekomstig onderzoek zich moeten focussen op de *All Data Revolution* - het analyseren van verschillende soorten databronnen (Lazer et al. 2014, 1204-1205) - waardoor deze

analyse gevoed wordt met veelzijdiger informatie. Kwalitatief onderzoek zal hierin een grote bijdrage leveren. Concluderend kan gesteld worden dat metingen enkel op deze wijze dicht bij een kwalificatie van de werkelijke medewerkersbevlogenheid komt.

Naast een kleine participatiegroep, het anonimiseren van de data en het analyseren van slechts één databron, kent de uitvoering van deze onderzoeksmethode ook enkele limitaties. Zo heeft tot op heden geen datagedreven benadering van medewerkersbevlogenheid plaatsgevonden, waardoor de methode nog nooit eerder is uitgevoerd of bevestigd. Dit gegeven uit zich in het vierde kritiekpunt van dit onderzoek, namelijk de 45% validatiegraad voor het bepalen van sentimenten. Deze lage waarmerking geeft in de datavisualisatie van figuur 3 een vertekend beeld. Met het voeden van de analyse met veelzijdige informatie en van meerdere participanten kan de sentiment analyse verder ontwikkeld worden tot een hogere validatiegraad. Daarnaast wijzen zowel Paul (2007) als Seth et al. (2014) op de rol van keuzes die gemaakt worden binnen het visualisatie proces, maar ook de wijze waarop gebruikers de datavisualisatie interpreteren. De subjectiviteit van datavisualisaties wordt hiermee benadrukt en de afhankelijkheid van menselijke referentiekaders bevestigd. Dit gegeven verwerpt de assumptie dat datagedreven inzichten vrij zijn van menselijke framing en veronderstellingen (Kitchen 2014, 4).

Dit onderzoek is vertrokken vanuit de gedachte dat een datagedreven kwalificatie een allesomvattend, accuraat en frequent inzicht biedt in medewerkersbevlogenheid (Meerman 2015). Critici Boyd, Crawford (2012) en Kitchen (2014) beschrijven dit als de veronderstelling dat data-analyses beschikken over een hogere vorm van intelligentie, doordat de vermeende objectieve en betrouwbare ontologie. Doordat de onderzoeksmethode vanuit deze aanname is vormgegeven, bevat de methode vele limitaties en assumpties, waardoor gedegen causale verbanden of correlaties voor bevlogen gedragingen met maildata niet aantoonbaar zijn. Beginnende met de interpretatieve aard van het gehanteerde JD-R model. Het model focust zich op het schalen van subjectieve gevoelens of belevingen van een gemoedstoestand die per individu anders is of anders tot uiting komt, en gaat niet in op concrete gedragingen. Hierdoor kent het JD-R model geen kwantitatieve analyse en biedt het een beperkt handvat voor datagedreven analyses. Met behulp van de Two-step Flow theorie en bevindingen uit het boek *Digital Methods* is gepoogd deze limitaties te beperken. Een nieuwe wetenschappelijke en toepasbare benadering van medewerkersbevlogenheid is hiermee nog niet ontstaan, echter is wel aangetoond dat kwantitatieve inzichten kunnen bijdragen aan een kwalitatieve benadering van medewerkersbevlogenheid. Zoals Rogers (2013) aangeeft moeten onderzoeksmethodes zich aanpassen aan het nieuwe medium waar de analyse op uitgevoerd wordt. De Two-step Flow theorie blijkt voor een datagedreven benadering van medewerkersbevlogenheid een goede aanvulling te zijn op het zwakke JD-R model. Uit de analyse blijkt dat binnen interne mailnetwerken een 'multistep flow of information' plaatsvindt, waarbij mail-mediators personen zijn met een verhoogde kans op een bevlogen identiteit. In hoeverre deze mediators buiten mail ook dit gedrag vertonen is niet vastgesteld.

De nadruk binnen deze discussie ligt op de bijdrage die datagedreven kwalificatie biedt bij onderzoek naar een nieuw meetinstrument voor medewerkersbevlogenheid. Maildatagedreven

inzichten bieden ons nieuwe kennis over bevlogen digitaal gedrag, die we met het blote oog niet kunnen waarnemen (Kitchen 2010). Ondanks de drang naar kwalificatie en het meetbaar maken van ons gedrag, bieden data op zichzelf geen gedegen representatie van de werkelijkheid. Hierin volg ik Lazer et al. (2014) die stellen dat we ons niet blind moeten staren op data-analyses, maar het beste onderzoek diverse, groot of klein, traditionele of nieuwe bronnen bevat. Dit onderzoek legt de basis voor een nieuw digitaal meetinstrument van medewerkersbevlogenheid. Voor een validatie van deze bevindingen zal vervolgonderzoek de mailanalyse moeten aanvullen met meer context vanuit etnografische inzichten of gedetailleerde resultaten van het huidige MTO. De gemeente Zwolle beschikt met deze kritische analyse niet over een nieuw meetinstrument, maar is wel op weg het *datafication* paradigma binnen de publieke sector op de kaart te zetten. Vanuit academisch perspectief is het van belang om kritisch stil te blijven staan bij de ethische gevolgen van een datagedreven kwalificatie van medewerkersbevlogenheid. De eindgebruiker, medewerker of manager, moet zich realiseren dat een data-analyse een verhoogde kans op bevlogenheid presenteert en niet de daadwerkelijke waarden. Men zal op deze wijze de resultaten moeten interpreteren en ernaar handelen. Enkel zo levert een datagedreven meetinstrument in de toekomst een positieve bijdrage binnen ons *gedataficeerde* paradigma.

---

## 7 Literatuur

Bakker, Arnold B., W.B. Schaufeli en E. Demerouti. 1999. "Werkstressoren, energiebronnen en burnout: het WEB Model." in J. Winnubst, F. Schuur en J. Dam. *Praktijk gezond werken* (II 3.2): 1-19. Maarssen: Elsevier.

Bakker, Arnold B., J.H. Gierveld en K. van Rijswijk. 2006. *Succesfactoren bij vrouwelijke schoolleiders in het primair onderwijs: Een onderzoek naar burnout, bevlogenheid en prestaties*. Diemen: Right Management Consultants.

Bakker, Arnold B., en E. Demerouti. 2008. "Towards a model of work engagement." *Career Development International*, 13(3): 209-223.

Berg, Odette van der. 2013. "Is sentiment-analyse belangrijk bij ontwikkelen van je content?" *De kracht van Content*. <http://dekrachtvancontent.nl/social-media/sentiment-analyse-belangrijk-bij-content/>

Biber, Douglas, Susan Conrad en Randi Reppen. 1998. *Corpus Linguistics: Investigating Language Structure and Use*. Cambridge: Cambridge University Press.

Boyd, danah en Kate Crawford. 2012. "Critical Questions for Big Data" *Information, Communication & Society*, 15(5): 662-679.

Chen, H., R.H.L. Chiang en V.C. Storey. 2012. *Business Intelligence and Analytics: from Big Data to Big Impact*. *MIS Quarterly*, 36(4): 1165-1188.

Demerouti, E en A.B. Bakker. 2011. "The Job Demands-Resources model: Challenges for future research." *SA Journal of Industrial Psychology/SA Tydskrif vid Bedryfsielkunde*, 37(2).

Dijck, José van. 2014. "Datafication, dataism and dataveillance: Big Data between scientific paradigm and secular belief." *Surveillance & Society*, 12(2): 197-208.

Fraser, Nancy. 1990. *Rethinking the Public Sphere: A Contribution to the Critique of Actually Existing Democracy*. *Social Tekst*, 25/26: 56-80.

Gitelman, Lisa. 2013. *Raw Data Is an Oxymoron*. London: MIT Press.

Gray, Mary L. 2014. "Critiquing Big Data: Politics, Ethics, Epistemology." *International Journal of Communication*, 8: 1663-1672.

- Hall, Stuart. 1990. "Cultural identity and diaspora" in *Identity*. London: Lawrence & Wishart: 222-237.
- Jenkins, Henry, Sam Ford en Joshua Green 2013. *Spreadable Media: Creating Value and Meaning in a Networked Culture*. New York: NYU Press.
- Katz, Elihu en Paul Felix Lazarsfeld. 1955. *Personal Influence: the Part Played by People in the Flow of Mass Communications*. New York: Springer.
- Kitchen, Rob. 2014. "Big Data, new epistemologies and paradigm shifts" *Big Data & Society* 1.
- Lazer, David, Ryan Kennedy, Gary King en Alessandro Vespignani. 2014. "The Parable of Google Flu: Traps in Big Data Analysis." *Science*, 343: 1203-1205.
- Liu, Bing. 2012. "Sentiment Analysis and Opinion Mining." *Synthesis Lectures on Human Language Technologies*. University of Illinois at Chicago.
- Llorens, S., W.B. Schaufeli, A.B. Bakker en M. Salanova. 2007. "Does a positive gain spiral of resources, efficacy beliefs and engagement exist?" *Computers in Human Behaviour*, 23: 825-884.
- Markus, M.L. 2011. "Is Information Richness Theory Rich Enough? Or How Managers Using Email Cope with Lack of Richness" *Working paper Anderson Graduate School of Management*.
- Mayer-Schönberger, Victor en K. Cukier. 2013. *Big Data: A Revolution That Will Transform How We Live, Work and Think*. New York: HMHPC.
- Maslach, C. en M.P. Leiter. 1997. *The truth about burnout*. San Francisco: Jossey-Bass.
- McLuhan, Marshall. 1967. *The medium is the message*. London: MIT Press.
- Meerman, Anouk D. 2015. "Happiness at Work 2.0. De ethiek van een data-gedreven kwalificatie van Zwolse medewerkersbevlogenheid." Universiteit van Utrecht.
- Paul, C. 2007. "The Database as System and Cultural Form: Anatomies of Cultural Narratives." *Database Aesthetics: Art in the Age of Information Overflow*. Minneapolis, MN: University of Minnesota Press: 1-16.
- Rogers, Richard. 2013. *Digital Methods*. Cambridge: MIT Press.



Salanova, M., S. Agut en J.M. Peiró. 2005. "Linking organizational resources and work engagement to employee performance and customer loyalty: The mediation of service climate." *Journal of Applied Psychology*, 90: 1217-1227.

Schaufeli, W.B. en A.B. Bakker. 2001. "Werk en welbevinding: naar een positieve benadering van Arbeids- en Gezondheidspsychologie" *Gedrag & Organisatie*, 14: 229-253.

Schaufeli, W.B. en A.B. Bakker. 2004. "Job demands, job resources, and their relationship with burnout and engagement: a multi-sample study." *Journal of Organizational Behaviour*, 25: 293-315.

Seth, Anil, Remco van den Bosch, Steffen Mieske, Holger Baumgardt,, Mark den Brok, Jay Strader, Nadine Neumayer, Igor Chilingarian, Michael Hilker, Richard McDermid, Lee Spitler, Jean Brodie, Matthias Frank en Jonelle L. Walsh. 2014. "A supermassive Black Hole In An Ultra-Compact Dwarf Galaxy" *Nature*, 513(7518): 398-400.

Sonnentag, S. 2003. "Recovery, work engagement, and proactive behaviour: A new look at the interface between non-work and work." *Journal of Applied Psychology*, 88: 518-528.

Swan, M. 2012. "Sensor Mania! The Internet of Things, Wearable Computing, Objective Metrics, and the Quantified Self 2.0" *JSAN*, 1(3): 217-253.

Tyler, Joshua R., Dennis M. Wilkinson en Bernardo A. Huberman. 2003. "Email as Spectroscopy: Automated Discovery of Community Structure within Organizations"

Watson, H.J. 2007. "The Current State of Business Intelligence" *Computer*, 40(9): 96 - 99.

Wester, Fred en Vincent Peters. 2004. *Kwalitatieve analyse. Uitgangspunten en procedures*. Bussum: Coutinho.

Wu, Fang, Bernardo A. Huberman, Lada A. Adamic en Joshua R. Tyler. 2004. "Information flow in social groups" *Physica A: Statistical Mechanics and its Applications*, 337 (1): 327-335.

Wu, Shaomei, Jake M. Hofman, Winter A. Mason en Duncan J. Watts. 2011. "Who Says What to Whom on Twitter" in *Proceedings of the 20th international conference on World wide web (WWW '11)*. New York: ACM, 705-714.

## Bijlage A | Onderwerp label raamwerk

Title	Keywords	Frequent y	Description	Example
<b>Factureren (Label 1)</b>	Factureren Betalen	1400 1271	Mails over facturen en de betalingen hiervan	“Facturen betreffende 2014”
<b>Bijeenkomst (Label 2)</b>	Afscheid Bijeenkomst PFO Programma Presentatie Symposium	570 466 382 283 253 231	Mails gerelateerd aan een evenement of bijeenkomst en de omschrijving hiervan	“Check voortgangslijst PFO Brink”
<b>Organisatie (Label 3)</b>	Beleid Organisatie- dialogen Sturing Organisatie Organisatieplan Topkring	412 373 352 312 266 286	Mails over organisatorische processen als beleid en sturing	“150401 rapportage sturing beleid” (ID
<b>Uitnodigen (Label 4)</b>	Uitnodiging Reminder	924 468	Mails over het uitnodigen van bijeenkomsten	“uitnodiging 7 april” (ID 1863)
<b>Onderzoek (Label 5)</b>	Data Onderzoek	408 397	Mails over onderzoek gerelateerde projecten	
<b>Digitaal (Label 6)</b>	Fatclients Sharepoint Eigentijds Digitalisering	454 352 261 240	Mails over digitale processen binnen de gemeente Zwolle	
<b>Proces (Label 7)</b>	Vorbereid Check Updates Voortgangslijst Procesbegeleider	480 393 388 374 236	Mails over processen binnen projecten	
<b>Tijds- aanduiding (Label 8)</b>	Dagen Februari Januari Vanavond November Maart	1357 661 332 272 240 228	Mails waarin tijden, dagen en maanden centraal staan	
<b>Geografisch (Label 9)</b>	Zwolle Cities Nederland	559 282 277	Mails waarin een plaatsaanduiding centraal staat	

## Bijlage B | Positief/negatief label raamwerk op basis van steekproef

Title	Keywords	Description	Example
<b>Positief</b>	goed, goede, mooi(e), leuk, benieuwd, actieve, mooist, (heel) interessant, een must, gerust, van harte, graag, lekker(s), liefst, bedankt, fijne, prettige, enthousiast, aanrader, luchtige, genoeg, harte welkom, vriendelijk, positieve, blij, gave, rijzende ster.	Bij de steekproef gehanteerde termen die aanduiden dat een tekst positief gestemd is.	"mooie concurrentie en zeker en aanrader voor iedereen om eens te zien hoe anderen... "
<b>Negatief</b>	probleem, helaas, niet kan, echter, onverwacht gevel, griep, koorts, verontwaardigd, veronderstelling, excuses, ziek, jammer, helaas	Bij de steekproef gehanteerde termen die aanduiden dat een tekst negatief gestemd is.	"de uitrol van de updates van diverse applicaties gaat wederom niet door helaas..."

## Bijlage C | Formeel/Informeel label raamwerk

Toon	Aanhef	Afsluiting
<b>Informeel (Label i)</b>	He Hé Hi Hey Hai Hay Hoi Dag Ha Lieve	Doei Gr Groeten Groetjes Grtz Yo Mvg Xxx Liefs
<b>Formeel (Label f)</b>	Beste Geachte Goedemiddag Goedemorgen Goedenavond	Hoogachtend Met vriendelijke groeten Met vriendelijke groet Tot ziens Met hartelijke groeten Met hartelijke groet Hoogachtend