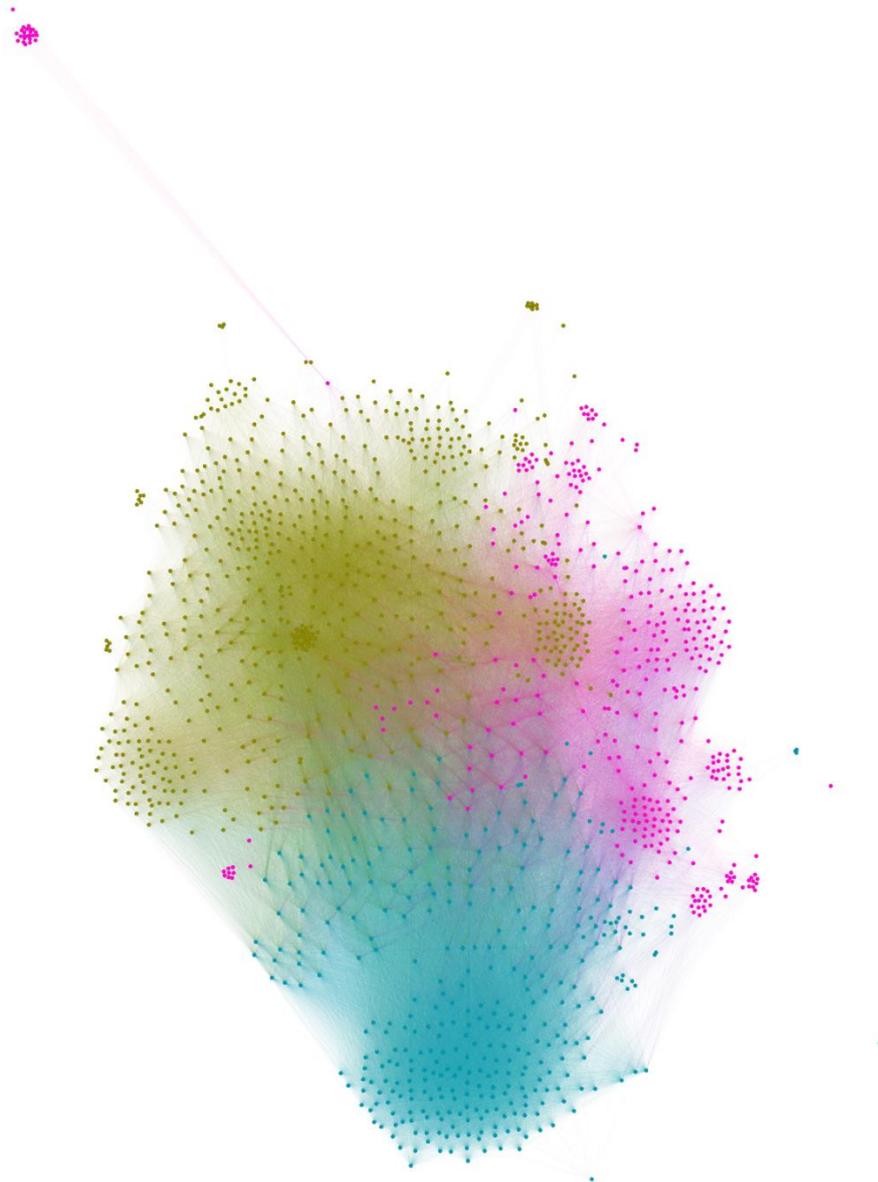


# The monads of Facebook

A comparative study of two digital methods approaches



**Author:** Asger Gehrt Olesen

**Study number:** 20161496

**Supervisor:** Anders Kristian Munk

**University:** Aalborg University Copenhagen

**Programme:** Techno-Anthropology

**Type:** Master thesis

**Pages:** 70

# Abstract

Det sociale er i højere og højere grad udgjort af ikke bare offline, men også online interaktioner med andre. Vi efterlader os et utal af digitale spor, som gør det både relevant og muligt at studere det sociale under nye præmisser.

Dette speciale undersøger, hvordan to forskellige metodiske tilgange til at studere det sociale på digitale platforme empirisk adskiller sig fra hinanden. Den ene tilgang, groft oversat til den *emneorienterede*), benytter sig af såkaldte kontroverser til at forstå, hvordan aktører danner alliancer og bygger netværk ved at engagere sig i de samme kontroversrelaterede problemer. Den anden tilgang, kaldet den *medieorienterede* (*medium centric*), benytter sig af de digitale platformes indbyggede måde at organisere det sociale på til at forstå, hvordan netværk opstår.

Begge tilgange er funderet i 'digitale metoder'. Digitale metoder er en studiegren, der dels har rødder i aktør-netværks-teori der ligger indenfor videnskabs- og teknologistudier og dels i mediestudier. Groft sagt kan man sige at den emneorienterede tilgang udspringer af aktør-netværks-teorien og den medieorienterede af mediestudier.

Det problematiske ved studier, der indtil i dag har undersøgt det sociale via digitale metoder er, at de ofte bruger hhv. den emneorienterede og den medieorienterede tilgang i en sammenblanding. Det har dermed ikke været muligt at identificere, hvad tilgangene hver især bidrager med, hvor der er ligheder og forskelle, og hvilke empiriske implikationer der er af at bruge den ene tilgang kontra den anden i repræsentation af sociale interaktioner.

Igennem Gabriel Tardes begreb *monader* dykker dette speciale ned i denne problemstilling. Specialet tager afsæt i en antagelse om, at det sociale ikke kan findes som *sui generis* påvirkninger af individet. Det sociale er defineret som en lang række af sociale forbindelser imellem individer, der ikke i sig selv er sociale. Monadebegrebet er ikke forbeholdt humane aktører, men omfatter alt, der kan knyttes en forbindelse mellem.

Specialet argumenterer for, at man kan betragte Facebook som et netværk af monader. I det perspektiv vil alle brugere af Facebook, der er venner med hinanden, indgå i hinandens monade. Brugerne vil samtidig være en del af de monader, der opstår ved, at de etablerer forbindelse med grupper og pages via blandt andet kommentarer, likes og delinger.

Specialets problemformulering lyder: *Hvilke empiriske forskelle kan observeres i måderne, hvormed hhv. den emneorienterede tilgang og den medieorienterede tilgang inden for digitale metoder tilføjer optegnelser til de monadiske lister? Og hvilke implikationer har det for socialteori?*

Problemformuleringen besvares gennem et komparativt studie af en konkret case på Facebook, nemlig en uenighed om hvorvidt HPV-vaccinen er skyld i en række bivirkninger. Studie af HPV-kontroversen baserer sig på en kombination af to datakilder. Den ene et helt nyt datasæt med opslag fra 72.000 offentligt tilgængelige danske Facebook-sider. Det andet et datasæt, hvor Facebook-opslag er udvalgt manuelt ud fra deres relation til HPV-kontroversen. Begge datasæt er høstet gennem Facebooks API.

Specialets første bidrag er metodisk. Specialet opstiller for første gang en distinkt operationalisering af hhv. den emneorienterede og den medieorienterede tilgang, som muliggør en transparent sammenligning og vurdering af de to tilgange. Med udgangspunkt i operationaliseringen konstrueres to sammenlignelige netværk, der repræsenterer de to tilgange.

Specialets andet bidrag er empirisk. Specialet viser, at der er store forskelle i, hvordan de to tilgange repræsenterer det sociale, når det vedrører HPV-kontroversen. Et tydeligt udtryk herpå er, at overlappet af monader mellem de to netværk er lavt. Det udbygges af, at de tre grupperinger, der identificeres i hvert netværk, i lav grad minder om hinanden. Endeligt er der stor forskel på, hvilke aktører der identificeres som centrale i de to netværk. Monader med mange forbindelser i det ene netværk har meget få i det andet netværk. Det hænger sammen med, hvordan de to tilgange identificerer centrale aktører. Den emneorienterede tilgang belønner monader, der bruger det rette vokabularium, hvorimod den medieorienterede tilgang belønner monader, der samler mange interaktioner fra aktører, der interagerer med andet i netværket.

Specialets tredje bidrag er teoretisk. Tager man de empiriske konklusioner i betragtning er implikationerne for socialteori alvorlige og væsentlige at forholde sig til i fremtidige studier med digitale metoder. Det kan ikke antages, at de to tilgange resulterer i samme udlægning af en kontrovers. Ved at vælge den ene eller den anden metodisk tilgang, træffer man samtidig et teoretisk valg om, hvordan man anskuer en social forbindelse.

## Table of Contents

1	Introduction: In search of the social .....	5
1.1	Origins of digital methods .....	6
1.2	Two theories of the social in Science and Technology Studies.....	7
1.2.1	Durkheim: The social exists independent of the individual.....	8
1.2.2	Tarde: The social is a circulating fluid.....	9
1.3	Two ways to assemble Tarde's monads .....	12
1.3.1	The issue centric approach .....	12
1.3.2	The medium centric approach .....	14
1.3.3	Interwoven methods .....	15
1.4	Problem statement .....	16
1.5	Motivation.....	16
2	Theory: Assembling Tarde's monads with ANT .....	17
2.1	In search of controversies .....	17
2.2	Networks, maps and visualisations.....	21
2.2.1	Panoramas and oligopticons .....	22
2.2.2	Cutting the network .....	24
2.3	Theoretical summary.....	25
3	Methodology.....	25
3.1	The recipe for a comparative digital study .....	26
3.1.1	The data source .....	27
3.1.2	Building a common point of departure .....	30
3.1.3	The operationalisation of monadic network graphs.....	36
3.1.4	Metrics for comparing the two approaches .....	41
3.1.5	Summary of the comparative study .....	44
3.2	Methodological challenges in digital methods projects .....	45
3.2.1	Post-demographic and natively digital data .....	46
3.2.2	Digital bias .....	47
3.2.3	Compliance with GDPR.....	49

3.3	Facebook as an ethnographic field.....	49
3.3.1	Ethnography of the digital.....	51
3.3.2	Gaining and sustaining rapport.....	54
3.3.3	Tools as “armchair harvesting”? .....	57
4	Analysis of the empirical comparison.....	60
4.1	The panoramic comparison .....	61
4.1.1	The medium centric network .....	61
4.1.2	The issue centric network.....	69
4.1.3	Analysing the panoramic similarities and differences.....	75
4.1.4	Summary of the panoramic comparison .....	78
4.2	The oligoptic comparison.....	79
4.2.1	Case 1: Highest degree.....	79
4.2.2	Case 2: Average degree .....	84
4.2.3	Case 3: Low degree .....	89
4.2.4	Oligoptic summary .....	90
5	Discussion.....	91
5.1	Robust methods.....	91
5.1.1	Revisiting the issue dictionary .....	91
5.1.2	Oligoptic comparison based on cases .....	93
5.2	The implications for social theory .....	94
5.2.1	Assessing central actors and stages .....	94
5.2.2	Assembling the monads as representations of issue publics .....	96
6	Conclusion .....	97
7	References.....	99
8	Appendix .....	107
8.1	Appendix A – Table of topics and keywords .....	107

# 1 Introduction: In search of the social

During a lecture in 1998 Bruno Latour proposed that *virtual society* materialises through the information flow of bores and bytes in modems, sockets and cables (Woolgar 2002; Latour 1998). Social associations unfolding on the Internet could thus provide an occasion to let Actor-Network Theory (ANT) study the social in its making by following the digital traces of actors online (Latour 1998).

The aim of the *Virtual Society* programme was to investigate the transformative effects of information and communication technologies on society in its transition to the virtual (Woolgar 2002). As Latour put it in his lecture:

*“Virtual Society thus is not a thing of the future, it’s the materialisation, the traceability of Society. It renders visible because of the obsessive necessity of materialising information into cables, into data that has a size so to speak”* (Latour 1998).

On the receiving end of Latour’s proposal was the young sociologist Richard Rogers who over the following years would develop tools and methods to capture and analyse these digital traces (Rogers 2004, 2009 and 2012; Venturini and Guido 2012).

Rogers argued that digital media platforms afforded an opportunity to describe social life in new ways. The ways in which actors associate, for instance, become traceable in the medium in very specific ways (Rogers and Marres 2002, 342). This is so because users mediate their interactions through practices such as liking, posting or commenting which all leave digital traces.

*“Social relations are established in a tangible form as a material network connection. We take Latour’s claim of the tangibility of the social as a point of departure in our search”* (Rogers and Marres 2002, 342).

In his inauguration speech as professor at the faculty for New Media Studies at the University of Amsterdam, Rogers proposed what he baptised Digital Methods as a novel approach to studying the Internet (Rogers 2009, 1). Rogers proposed that such a research program should, as he put it, “follow the medium” (2009, 1), thus foregrounding the specific ways in which said medium contributes to the ordering of social life.

Digital methods is still a relatively new field of study in Science and Technology Studies (STS). It is influenced by two interwoven traditions: Rogers study of the medium as a device that contribute to the ordering of social life and the Actor-Network Theory based study of controversies as a means to inspect how the social is assembled and reassembled in socio-materiel relations (Latour 2005).

It is not clear in the current digital methods literature how these two traditions contribute to different perceptions of the social. Often the dogma “follow the medium” is collapsed with the dogma “follow the actor” (Marres 2015; Bach et al forthcoming; Marres and Rogers 2005; Public Data Lab 2017; Rogers 2012; Venturini 2012; Marres and Gerlitz 2016, Latour et al 2012). As examples, Marres advocate to follow the actors in controversies, but use the medium specificity of web-links to do so (2015); Public Data Lab study the spread of fake news and misinformation online through the intertwined use of a range of media sources and means to map actors’ positions in the controversy (2017).

Through an empirical test based on the HPV-vaccine controversy I will investigate the consequences of following respectively the actors of a controversy and the medium specific way of ordering social life.

## 1.1 Origins of digital methods

Digital methods scholar Noortje Marres claim that the discipline has been under development as an interdisciplinary endeavour between Science and Technology Studies (STS), Media Studies and Computer Science (2015, 657). Digital methods is as a consequence highly affected by both ANT’s ambition to study the social through controversies (represented by the STS scholars in the interdisciplinary endeavour), by Media Studies’ focus on the online settings that render the social ties visible and by Computer Science methods to engage with the medium.

In STS, controversies have functioned as both objects of sociological inquiry for the branches encompassed in the Sociology of Scientific Knowledge (SSK), and as a methodological move to study the social in its making for scholars of ANT. In the former, the mechanisms through which actors manage to impose *closure* on controversies are studied in order to show how scientific facts are socially constructed (Collins 1983; Pinch and Bijker 1987). In the latter, controversies are seen as momentary opportunities to study the social in its making by *following the actors* and observe how they associate and form alliances (Callon 1986; Latour 2005).

ANT's focus on controversies has in digital methods become manifested in the concept of *issue mapping* (Marres 2015). As the name implies, the approach concerns itself with the observation of issues in controversies. As such it does not necessarily confine itself to the medium specific ordering of social life, but instead argues for what I will call an *issue centric* mindset seeking to investigate how publics come into being through their engagement in controversies (Marres 2005).

As such, digital methods arguably encompass two approaches to the study of the social. One is to follow the actors during controversies, *the issue centric approach*, and the other to follow the medium, what I will call the *medium centric approach*.

The two approaches can easily be confused with one another. Both the ambition to follow the actors during controversies and the ambition to follow the medium draw inspiration from ANT. Indeed, they both declare it as their aim to study the social in its making by tracing and mapping the actors associational ties and social links (Rogers 2009; Marres 2015).

The difference then, must be found in their practical consequences. But also here, what separates the two approaches can be difficult to pinpoint. The toolboxes they use are all labelled *Digital Methods* and seem to contain the same brands and types of pliers, crowbars, screwdrivers and equipment for measurement to engage with the digital platforms. The protocols describing how the tools are to be used are often vague and ambiguous.

It might seem like splitting hairs and an effort to drive an artificial wedge in between the two approaches. If the aim and the theoretical realm is the same, why not leave it there? But for a study that claims to be rooted in matters which are *empirically observable*, it seems counterintuitive that the implications of the different representations of the social has not been empirically tested (Latour 2005; Marres 2015; Rogers 2009). This is the locus of techno-anthropological problems and theories where this thesis resides.

## 1.2 Two theories of the social in Science and Technology Studies

*This chapter introduces two social theorists - Emilé Durkheim and Gabriel Tarde - who has inspired different theories, programmes and methods in Science and Technology Studies.*

In the late 1800's and early 1900's Emilé Durkheim (1858-1917) and Gabriel Tarde (1843-1904) fought over the question of what defines the social for sociology (Latour 2018).

Understanding what definition of the social digital methods by proxy of ANT has adapted seems paramount to understand what definition of the social is operationalised by the two approaches - the issue centric and the medium centric.

### 1.2.1 Durkheim: The social exists independent of the individual

*"[Social facts] may be measured, their relative sizes compared, as is done with the intensity of electric currents or luminous foci"* (Durkheim 1951).

The French philosopher Emilé Durkheim argued that the social is a phenomenon, which can be studied empirically through statistics (Durkheim 2000). By observing the relationship between birth rates and suicides, Durkheim demonstrated that social facts exert an external force on the individual and exists independently - *sui generis* - of this individual (Durkheim 1951).

Individuals, Durkheim argued, are aligned by a *collective consciousness*. Social order is regulated by social facts that exert their force on us (Durkheim 2000). The whole of the social organism (society) must thus be thought of as larger than the sum of its parts (the individuals). The social has an existence of its own and is always and already there with us.

In the Durkheimian tradition, then, sociology is the study of the societal body and its efforts to maintain its own equilibrium. Such a state is only reached if necessary pressure is enforced on its constituent parts (Durkheim 2000). As such, Durkheim draws on Herbert Spencer's notion of the social organism (1940).

Durkheim claims that the field of inquiry is best delimited by deciding beforehand what social groups, facts, norms and laws are most suited to reach the aim of a study. All that matters is choosing the property which seems most appropriate for one's purpose (Latour 2018). This furthermore prevent the researcher from floundering in exhaustive intuition (Durkheim 2000, 40-41). Here Durkheim draws on Auguste Comte's positivist framework. Comte argued that society operates as the physical world based on general laws. Positive facts can only be obtained empirically, disregarding metaphysics and intuition.

### 1.2.1.1 Durkheim's influence on Sociology of Scientific Knowledge

With the early social constructivists engaging in the Sociology of Scientific Knowledge (SSK) traits from Durkheim's sociology were introduced to Science and Technology Studies. Concepts such as *causality* called on the researchers to study "*what brings about and sustains a body of knowledge*" leading to an account that is "*not merely descriptive but also explanatory*" (Bloor and Wright 2015).

The focus of SSK frameworks retains to *explaining* how social consensus during controversies impacts the *construction* of scientific facts and technology, as well as to extend these explanations to wider political and societal structures (Bruun et. al. 2007). The sociology practiced in SSK is kept in close proximity to that of Durkheim of sui generis social facts.

### 1.2.2 Tarde: The social is a circulating fluid

Tarde did not believe in the existence of sui generis forces external to the individual, but instead "*considered the social as a circulating fluid that should be followed by new methods and not a specific type of organism*" (Latour 2005, 13). The social as a circulating fluid is "transported" through individuals who *imitate* each other (Berry and Thrift 2007). Imitations are not 1:1 reproductions of an act but entails *refractions* through the individuals (Tarde in Latour 2018, 8). The sum of the refractions "*is the entire reality of a social thing at a given moment; a reality which is constantly changing, just like any other reality, through imperceptible nuances*" (Tarde in Latour 2018, 8). Tarde's definition of the social as empirically observable objects are the associational ties between entities which imitate each other.

Tarde used the concept of monads both to illustrate the circulating fluid of the social and to argue that society is smaller than its individual parts (Latour et al 2012). A monad is defined through a list of items, with which the monad is associated. This means, that if monad A is on the item list of monad B, monad B is a part of monad A's network. As such "... [a] *monad is not a part of a whole, but a point of view on all the other entities taken severally and not as a totality*" (Latour et al 2012).

Tarde dismissed the notions of individuals and society as separable entities, as the monads represent both the parts and the whole in the way they intersect: "*instead of thus explaining the small by the large, and the part by the whole, – I explain collective resemblances of the whole by the massing together of minute elementary acts – the large by the small and the*

*whole by the part*" (Tarde 1999 [1898], 63). The micro is contained in the macro, and the macro is contained in the micro.

Tarde argued, that the individual is not an isolated unit, but an intersection between diverse lines of imitations (Berry and Thrift 2007). Social constitutions like language, grammar and articulation are altered in the passing from individual to individual through the small changes made by each individual.

Tarde propose to use different spheres of life, as the "place" to observe the constitution of monads and the collective resemblances, inspired by the works of the French phonetician Jean-Pierre Rousselot. Rousselot is considered the founder of experimental phonetics, and through his tedious works developed methods for tracing the refractions in language through generations (Latour 2005). Tarde argued, that in order to make sociology a true science, the methods of Rousselot must be generalised and adapted. As such, the social should be observed by skilled and trustworthy sociologists as the refractions made to constitutions in the passing from individual to individual.

*"Let twenty, thirty, or as many as fifty sociologists, from different sections of France or any other country, write out with the greatest care and in the greatest possible detail the succession of minute transformations in the political or industrial world, or some other sphere of life, which it is their privilege to observe in their native town or village, beginning in their own immediate surroundings."* (Tarde 2003 [1899], 198)

Tarde recognised that the study of his envisioned monads was not practically possible at the time (Latour et al 2012; Tarde 1903). In order to empirically investigate monadological structures, the researcher would have to construct a dataset of detailed observations of a huge amount of individuals over a long period of time (Tarde 2003 [1899], 198). This in turn, would demand an army of trained ethnographers, to observe and record the trails of social acts through imitations and refractions, as each individual would need to be closely tailed. In other words, an unimaginable methodological task.

#### 1.2.2.1 Tarde's influence on Actor-Network Theory

Actor-Network Theory (ANT) can be seen as both a showdown with the Durkheimian sociology as represented by SSK and later on a declared allegiance to Tarde's monadology (Latour 2005).

In 1962, Thomas Kuhn proposed with his book *The Structure of Scientific Revolutions* that science does not recognise nature as it is but that scientific representations of nature is controlled by certain social and cognitive paradigms (Kuhn 1962). These paradigms have a *priori* thresholds for what is observable and what lines of enquiry can be followed. Kuhn's proposal gave inspirations to both the wave of sociological studies of scientific controversies as described above, but also to a branch that opposed the explanatory grounds of social constructionism. Through detailed ethnographic studies of the everyday life of laboratories the latter showed that scientific facts were results of negotiation processes. Negotiation processes include a long range of non-human actors such as materials, scientists, apparatus and experimental animals (Jensen 2003, 4-5). Among these laboratory studies Bruno Latour and Steve Woolgar's *Laboratory Life* is by many seen as a precursor for ANT.

It was through the descriptive controversy-centred case studies in the 1980's and 1990's, that the ANT vocabulary was sharpened. Concepts such as *free association* (dismissing a priori social categories), *generalised symmetry* (the breakdown of the divisions between social and natural) and dismissing the micro/macro divide can soundly be said to be closely related to Tardes ontology.

In line with Tarde, ANT authors criticise the Durkheimian sociology for being self-designated and self-appointed spokesmen of the people:

*"On the basis of scattered information, replies to questionnaires, anecdotes, statistics and feelings, the sociologist interprets, sounds out, incorporates and states what the actors are (classes, categories, groups, cultures, etc.), what they want, what interests them and how they live."* (Callon and Latour 1981, 297).

Callon and Latour accuse the self-appointed spokesmen for keeping the Social Organism in the form of Hobbes's Leviathan artificially alive in order to speak on its behalf:

*"...they have, for more than a century now, taken over from Hobbes's sovereign: the voice that speaks in the mask is their own"* (Callon and Latour 1981, 297).

The longer lasting social order of the social organism is not a result of sui generis social laws, but a result of the durable associations between actors. This in turn keeps the illusion of an overarching society afloat they argue (Latour 2005).

ANT's allegiance to Tarde became explicit in Latour's introduction to ANT (2005). In line with Tarde, Latour dismisses the idea of something *sui generis* and external to the individual. Instead he argues for the Tarde inspired *sociology of associations* as opposed to the Durkheimian *sociology of the social*.

Heavily inspired by Tarde's monads, Latour claims that the social "[...] *is visible only by the traces it leaves (under trials) when a new association is being produced between elements which themselves are in no way 'social'*" (Latour, 2005). The 'trials' referred to are controversies. During controversies, the self-evidence of established facts, norms and social groups breaks down, and actors form new groups around the controversy specific issues. The group formation processes expose for inspection the social ties between actors. Controversies as such offer an "*empirical occasion for a wider social inquiry*" (Collins and Pinch 1998; Latour 2005; Marres 2015).

### 1.3 Two ways to assemble Tarde's monads

Through the study of the connected web with digital methods, the prerequisites for following Tarde's monads in the search for the social are met. Social acts carried out on digital platforms are logged to detail, indexed and navigated by armies of servers and crawlers.

The ambition of digital methods is to collect and repurpose these social acts and associations for the study of the social. This ambition can be achieved either by following the actors during controversies through the *issue centric approach* or by following the digital platforms on which the actors perform the social through *the medium centric approach*.

#### 1.3.1 The issue centric approach

According to Noortje Marres, different assumptions guide the implementation of controversy analysis as a digital method. These assumptions are reflected in what she calls the *demarcationist*, the *discursivist* and the *empiricist* approach to digital issue mapping:

Demarcationists study knowledge controversies and argue they should be clearly distinguished from non-epistemic debates. Discursivists and empiricists on the other hand are interested in the entanglement of epistemic, public and political dynamics (Marres 2015, 662)

It is through the empiricist implementation of controversy analysis that Marres offers an extension of ANT's focus on controversies to digital settings (Marres 2015, 662). This

approach implies that the researcher source the places and spaces, where the issues of a controversy are debated in order to derive lists of keywords for platform query, webspace crawling and social media listening.

Two main problems occur using the issue centric approach for controversy mapping. Firstly, the mapping exercise will be in favour of the ones who manage to assemble the most durable vocabulary and choose on what stages<sup>1</sup> the controversy should be deployed.

Who are the opponents in a controversy? Who are for instance the opponents of the “anti-vacciners” in the HPV-vaccine controversy? The answer, according to the issue centric approach, would be the ones in direct association to the stages constructed for the controversy and the vocabulary assembled for the occasion.

Further complications to this problem are introduced when considering Marres’ interpretation of the two American pragmatic thinkers Walter Lippmann and John Dewey. In her readings of Dewey and Lippmann the *issue public* is defined to include all those directly and *indirectly* affected by some problem (Marres 2005).

When generating a dataset based on an issue centric approach, the issue public is delimited by the researcher’s ability to imagine where the issues are debated. This is because the machinery put in motion by researchers to retrieve information about controversies, are stage and vocabulary specific. They are like rescue dogs, only trained to find the persons who are at once both in distress *and* related to the specific stage and the specific vocabulary. But how about the parts of the public(s) who does not act on the controversy specific stages and with the controversy specific vocabulary, or label their actions in accordance with the issues derived by the researcher? They might also be worthy of researcher’s attention. Especially if the aim is “*wider social inquiry*”. But by using the issue attuned crawlers, the approach fails to recognise monadic entries and associations which are not expressed through the issue specific fora and with the issue specific vocabulary.

Secondly, the Tardian monad is being limited by the issue centric approach as it mainly focus on actions in direct association with the controversy. Michel Callon argue that the social and political identities are reconfigured during controversies (1987). What is being left out, are all the indirect associations, defined as non-controversy related entities, that make up and might

---

<sup>1</sup> Understood as a physical or virtual place to perform or act

inform “*wider social inquiry*”, the reconfiguration of political and social identities, and the composition of the monads respectively.

A consequence of the permeating focus on issue derived search criteria is that it prevents the qualification of questions that draw on both direct and indirect associations. What spheres of interest does users engage with before and after being engaged in a controversy? Is the participation in the HPV controversy a gateway for specific spheres of interest?

These types of questions cannot be answered based on an issue centric approach alone. The scope of inputs informing the monads is too narrow, as the search criteria does not with certainty allow for the entry of indirect spheres of interest.

### 1.3.2 The medium centric approach

Latour et. al. in their digital test of Tarde’s monads propose that “*it might now be possible to account for longer lasting features of social order*” (2012, 592) by following the medium in the way suggested by Rogers (2009, 1). By taking advantage of the profiles available on many digital platforms, the task of social science is now to learn how to “*navigate through overlapping ‘monads’ instead of alternating between the two levels of individual and aggregate*” (Latour et. al. 2012, 592).

Latour et. al. abide to ANT insofar that the social associations are not only applicable to human agents, but should be extended to all entities. The monads offer the descriptive layer advocated for in ANT: “*the more you wish to pinpoint an actor, the more you have to deploy its actor-network*” (2012, 592).

By following the medium follows an assumption that the digital platforms orders social life in accordance with a monadic point of view. This in turn means, that the methods for exposing and deploying actor-networks are hard-wired into the medium (Latour et. al. 2012).

The medium centric approach offers an opportunity for ANT to diverge from controversy analysis. It also entails a promise to overcome the problems of the issue centric approach in form of stage specific enactment and sole focus on direct associations. But the holy grail is not safe in hands yet, as varieties of those problems are foregrounded.

Netvizz, a tool for extracting data from Facebook (Rieder 2013), is medium specific. By inputs of unique identifiers of Facebook pages, it is possible to extract all posts and user interactions from a defined period of time. This data is subsequently converted into a network where users and posts are represented as nodes and the interactions represented as links. If a user has *liked* a post, a like-link will be defined. As such, the tool follows the native representation of the social (i.e. associations) of the platform.

How would one go about defining the pages to be included in a study of the social on Facebook? What guidelines does the medium centric approach offer to narrow the search? There is a real danger of flipping to the *sui generis* social theory, and apply predefined social categories to define the pages, in order to solve the problem of how and where to cut the network. If Tarde's monadology is to be taken literal, such categories should emerge from the observations.

### 1.3.3 Interwoven methods

The proposed separation of the two approaches made above is a deliberate and artificial construction. In reality, they are often mixed, blurred and unarticulated, and it can at times be difficult to identify the one from the other. But it comes as no surprise. Many of the researchers working with digital methods has either collaborated or been apprentices of one another. As an example, Noortje Marres, who proposed the empiricist implementation of controversy analysis in digital settings, is an apprentice of both Bruno Latour and Richard Rogers. This is evident in her research, where she makes use of both the issue centric and medium centric approaches.

Although a theoretical division between the issue centric and the medium centric approaches can be inferred, caricatured by Marres' "*empiricist implementation of controversy analysis*" and Rogers' "*follow the medium*", it is unclear what *empirically* separates the two approaches. How does the approaches render the study of the social possible and how is the social represented? How do they e.g. attribute centrality to actors and create groupings of monads? This might also be a consequence of researcher's preference to *black box* studies and describe them as straight lines going from a to b, from problem statement to final conclusion.

The lack of transparency obscures the fact that in order to go from a to b, detours has to be made to c, d and e. The detours are for example the blind alleys met when it turns out that

the platform chosen for a study does not offer the data points required. When these detours are redefined to be matters of fact as straight lines from a to b, the methodological considerations, finesses and choices become empirically muted. As such, there is no way to infer whether inclusion and exclusion of entities are results of the configuration of the issue or results of a medium specific ordering of social life.

## 1.4 Problem statement

The issue centric and medium centric approaches presented here are in essence two different ways of assembling the social in a way that is consistent with both Tarde and Actor-Network Theory (ANT). One approach follows the issues the other the medium. Both approaches use certain and at times deliberately mixed up methodological grips.

The main problem addressed in this thesis is, that no actual test of the implications of the different ways to represent the social has been carried out. This is a problem, as it from an empirical point of view obscures what the one approach has to offer compared with the other. The artificial divide between the two approaches will naturally have to be maintained in order to perform such an experiment.

I set forth to devise and execute a study, with the aim of render visible the empirical differences and similarities between the two approaches:

Conducting a comparative study on the same dataset delimited to contain entries based on a controversial search term, I will examine *how the addition of monadic entries differ between issue centric and medium centric approaches to digital methods? What are the implications for social theory?*

This thesis is divided in five chapters. After the introduction follows a deeper dive into the theoretical concepts used in the study. Chapter 3 lay out the methodological considerations, both in regards to the empirical comparison. Chapter 4 analyse the results of the empirical comparison, followed by a discussion of the results and a conclusion of the study found in chapter 5 and 6 respectively.

## 1.5 Motivation

With this thesis I am dedicated to solve a problem that has for long been budding in digital methods. Perhaps it has been overshadowed by an eagerness to develop cutting edge

methods and tools to study the social in digital settings, or perhaps it has fallen between two (or more) stools as a consequence of the interdisciplinary nature of digital methods.

The aim of the game is to avoid jumping head over heels from method to tool to conclusion. Skipping the intermediate steps and the possibilities to learn from those the circumstances under which tools and methods skew data and conclusions.

This thesis as such is highly driven by a motivation to understand and account for the traditions within digital methods that has become manifested in tools and methods. What logic lies behind the operationalisation and translation of ANT into methods into tools into conclusions?

## 2 Theory: Assembling Tarde's monads with ANT

*At this point it should be clear that the developments in social theory, digitalisation of social acts and the methods to trace these acts has created an opportunity to resurrect Tarde's notion of monads. Two main approaches - the issue centric and the medium centric - has been identified and kept artificially separated, in order to conduct an empirical comparison between the two. Before diving into the methodological considerations of this comparison, some theoretical points need clarification. This chapter aims to establish a deeper understanding of the turn towards the digital, and to shine a light on what conceptions of the public and networks is operationalised in ANT.*

### 2.1 In search of controversies

According to Latour there are two ontologies in sociology, in line with the Durkheim/Tarde debate outline in the introduction (2005). The first he labels *Sociology of the social*. Here attention to predefined social groups and categories is key. Researchers explain why the informants of the predefined groups act as they do given the existence of a societal framework and sui generis social laws. The explanations are applied as generalisations to a higher societal context and confirms the existence of a societal level above the individuals (Latour 2005).

The second, *Sociology of associations* or ANT, dismiss assumptions about a priori social groups and norms, as well as how these affect the outcome of controversies (Latour 2005). There is no a priori macrostructure to explain the actions of the micro. There is no global in which the local is situated. There is no context in which actions are situated besides their

own. Actors may contextualise and frame, but that is not a task for the researcher. Instead, what should be (micro)scoped is how actors explain contexts and make frames through the associations they make during group formations (Latour 2005, 184-185).

The difference between the two ontological trenches is summarized by Bruno Latour to be “[...] *on the question of knowing how appropriate it is to autonomise the specific facts that sociology is concerned with but also on the question of their exteriority and, in sum, on the strength with which this world imposes upon us*” (Latour 2018).

Although *generalised symmetry* in ANT is usually used to describe the breakdown of the division between human and non-human actors, symmetry is also applied to the dichotomy between the micro and the macro: “*Organizations, individuals, institutions, nations, corporations – if these are bigger or more powerful than other actor-networks we do not start with this difference and work from there; rather, it is this difference that needs to be explained.*” (Cressman 2009).

A main tenet of ANT is to “*follow the actors themselves*” in an effort to “*catch up with their often wild innovations in order to learn from them what the collective existence has become in their hands, which methods they have elaborated to make it fit together, which accounts could best define the new associations that they have been forced to establish*” (Latour 2005, 12).

According to Michel Callon, another notable figure in ANT, an actor-network is “*reducible neither to an actor alone nor to a network*” and furthermore “*is simultaneously an actor whose activity is networking heterogeneous elements and a network that is able to redefine and transform what it is made of*” (1987, 93).

By allowing the actors to speak on their own behalf without explaining, interpreting, generalising and reducing what comes to the foreground is the empirically observed actor-network interpreted through the actors’ own definitions and not by the predefined categories of the observer. The rather opaque language of ANT as *programme, actor, translation, intermediary, mediator* etc. help the observer to refrain from aggregating the actors own social explanations into a predefined vocabulary.

By definition, an actor can be both human and non-human:

*“If action is limited a priori to what ‘intentional’, ‘meaningful’ humans do, it is hard to see how a hammer, a basket, a door closer, a cat, a rug, a mug, a list, or a tag could act. They might exist in the domain of ‘material’ ‘causal’ relations, but not in the ‘reflexive’ ‘symbolic’ domain of social relations. By contrast, if we stick to our decision to start from the controversies about actors and agencies, then any thing that does modify a state of affairs by making a difference is an actor—or, if it has no figuration yet, an actant.” (Latour 2005, 71)*

The breakdown of the division between human and non-human actors in ANT is as mentioned earlier related to the principle of *generalised symmetry* (Callon 1986). Generalised symmetry refers to the practise of sticking to the same vocabulary when describing heterogeneous actors (Callon 1986, 4).

As the dichotomy between human and non-human actors is broken down in the symmetrical treatment of heterogeneous networks, focus is diverted to the containers or black boxes that carry meaning (Latour 2005, 39). Such containers can be either *intermediaries* or *mediators*. What characterises the former is that the output can be inferred from the input. No transformation, translation, distortion or modification has happened to the object they carry.

What characterises a mediator is that the output cannot be inferred by the input, as the object they carry is modified, transformed, translated or distorted in the process.

Actors can be difficult to follow and the black boxes that carry meaning difficult to inspect. What the case studies of the 80’s and 90’s showed was that controversies function well as methodological moves to both follow and inspect.

According to digital methods scholar Tommaso Venturini *“controversies are situations where actors disagree (or better, agree on their disagreement). The notion of disagreement is to be taken in the widest sense: controversies begin when actors discover that they cannot ignore each other and controversies end when actors manage to work out a solid compromise to live together. Anything between these two extremes can be called a controversy.”* (Venturini 2010, 261)

During controversies, established facts and routines which were previously black boxed, are pried open and become accessible for inspection, and actors’ group formation revolve around the various programmes and anti-programmes deployed to impact the construction of the new facts, norms and routines (Latour 2005).

In the example of Durkheim and Tarde, Durkheim's programme is the theory of generalisations based on sui generis social laws, and Tarde's anti-programme is the theory of refractions through imitations. As Latour concludes, Durkheim's program - sui generis social facts- managed to "*muster on the spot the largest number of well aligned and faithful allies*" (Latour 1986) and therefore won the right to define facts, norms and routines for sociology. Allies is anything which acts, or to which activity is granted by others (Latour 1996, 7).

Latour in his later works and Marres in general draws heavily on the two American pragmatists Walter Lippmann and John Dewey, and their definition of the public. Lippmann argue that the public is a phantom to be defined on ad hoc basis by the people interested in a topic (Lippmann 1993). According to Dewey the public is a variable entity made up by the actors who by direct or indirect association is affected by a transaction (Dewey 2012, 58). If the consequences of the transaction are not taken care of by the appropriate institutions, the public will take action in its stead. In this view, studies of controversies are studies of different publics.

Steven Epstein's study of AIDS activism is an exemplary description of a Deweyan public (1996). This public was constituted by the actors directly and indirectly affected by AIDS research and treatment strategies and methodologies. Prior to the controversy, AIDS research and treatment was shaped and defined almost exclusively by scientific experts. These experts had certain interests and motives concerning the development, testing and monitoring of new treatments. Epstein showed how the dissatisfaction of those affected negatively by the interests and motives of the experts assembled in a public that would eventually influence the scientific decision making of AIDS research (Epstein 1995). The public was constituted by people from all walks of life united by the consequences of how AIDS research was carried out at the time.

The issue public is not limited to only consist of lay people (Marres 2015). In controversies the gap between lay people and experts is momentarily bridged and established conventions and regimes annulled (Munk forthcoming). This means that the privileged position experts normally speak from when distributing knowledge, facts and paradigms from the scientific realm is short circuited, allowing others to take that position.

The relatively recent focus on the issue publics is associated with an interventionist turn within ANT and STS (Marres 2005; Marres 2015; Munk and Abrahamsson 2012). The new ambition is to intervene and not just to analyse and describe. As a result, the controversy

analysis deployed as a digital method today, is much more a *controversy detection* enterprise, as opposed to previous endeavours of STS to analyse relations between actors (Marres 2015).

## 2.2 Networks, maps and visualisations

*This chapter address the typical outputs of digital methods projects and how they fit into the descriptive type of reporting attuned to Actor-Network Theory.*

The modus operandi of *sociology of associations* does not offer a list of predefined categories or actor-networks to study. Instead, it demands of the researcher to be rigorously descriptive, to keep possibilities open and reject explanatory frameworks (Munk and Abrahamsson 2012). It prescribes to follow the actors and deploy their networks as monads and associational ties linking these together in an attempt to meet the standard of “*the uniquely adequate account of a given situation*” (Latour 2005, 144)

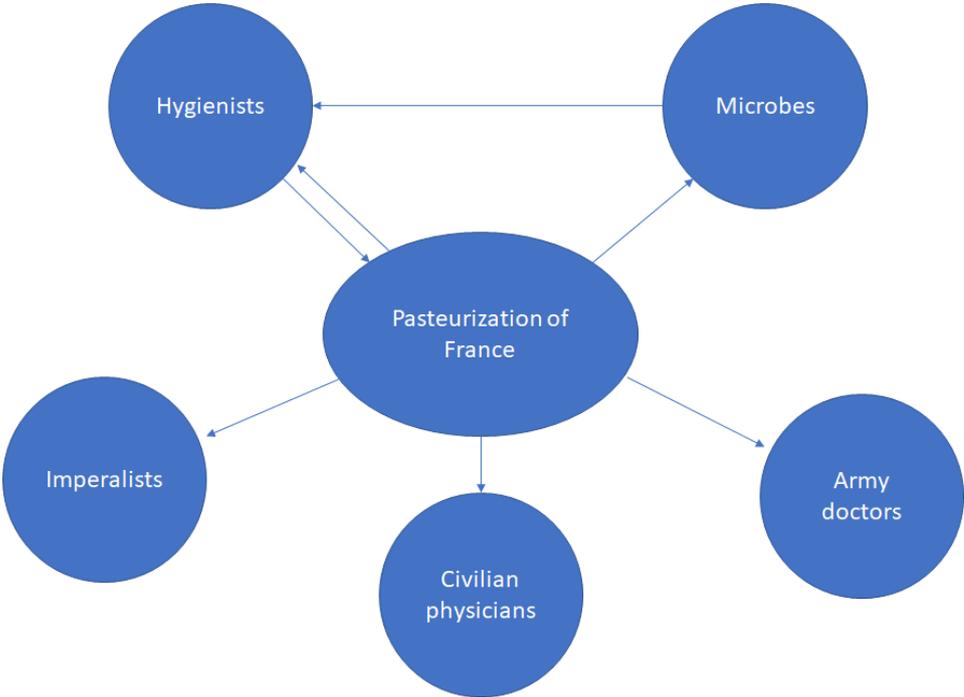
But when has the actor-network been described well enough? When has the associations, the translations, the mediators, the intermediators been sufficiently strongly deployed for the account of the actors to be able stand on its own feet, and give view to the social? “*You stop when you have written your 50,000 words or whatever is the format here*” Latour flippantly remarks (2005, 148). As Venturini and Guido notes “*while the notion of actor-network has gradually been accepted as a description of collective dynamics, ANT is still criticized for providing little help with operationalization*” (Venturini and Guido 2012).

The ambition of this thesis is not to produce uniquely adequate accounts, but to render visible the collective dynamics as literal network graphs. These networks in turn enables the construction of such accounts as “*a network is not what is represented in the text, but what readies the text to take the relay of actors as mediators*” (Latour 2005, 131). To make a comparison, take the example of a map of the World. It is difficult to imagine, that anybody would grasp how territorial borders has been divided or know the conception of the history of the countries and their inhabitants only by looking at the map. The same goes for network graphs.

The network graph, which is a standard output of digital methods projects, is not to be confused with a uniquely adequate actor-network account. Instead it readies a list of questions to be asked about the actors’ networks. How are they assembled? What are their ties made of? What makes them durable?

To illustrate the difference between a descriptive actor-network and the conception of networks operationalised in graphs take the example of Bruno Latour’s book *The Pasteurization of France* (1993). In this book, Latour describes how Louis Pasteur succeeded in persuading the French to believe in his definition of the relationship between microbes and diseases instead of the definitions of his opponents. Through his description Latour shows how the success did not depend on Pasteur alone, but on an entire network of forces including allegiances between Napoleon III and the catholic church, herds of sheep and petri dishes, journalists and hospital architects (Latour 1993; Munk forthcoming). Latour would obviously not have been able to capture this mobilisation of Pasteur’s network, by solely showing a graph visualisation as the one depicted in Figure 1.

Figure 1 - This graph shows a simple version of the network activated by Pasteur in the pasteurization of France



### 2.2.1 Panoramas and oligopticons

As described earlier both Latour and Tarde disregard the division of society into micro and macro structures. The long lastingness of macro structures is in fact associations made durable between monads. Instead of the micro/macro divide, Latour propose two types of vision (Latour 1986). Firstly, the *oligopticon* is designated as a site for the manufacture of social structures such as culture, opinion, consensus and knowledge (Latour 2005B). As

opposed to the all-seeing *panopticon*<sup>2</sup>, the oligopticon sees very little, but what they see, they see very well. The oligopticon is the monad's inventory list. From this view can be inspected what the monad is made of and what it is producing. Contrary to oligoptica, *panoramas* see everything. "*But they also see nothing since they simply show an image painted (or projected) on the tiny wall of a room fully closed to the outside.*" (Latour 2005, 187). Panoramas could be compared to how the monades look if linked together in a network by the associational ties of their inventory list. But they should not be misconceived as being a representation of something larger in size and "*they are misleading if taken as a description of what is the common world*" (Latour 2005, 189).

Imagine standing on the top most stands of a stadium converted to a music concert venue. From this elevated position, you can overview all the concert participants on the floor level of the venue. If you squint your eyes you might be able to identify tendencies of people clustering in groups. And by the use of binoculars you might even be able to count the size of the group, assess who each person directs their attention towards and perhaps identify the most central person in the group. You know that at least one thing binds all the people together at the venue at this specific time: the concert that is about to start. But your squinting eyes and your binoculars cannot penetrate the makeup of all the other associations shifting through time from this *panoramic* point of vision. In order to gain such insight, you would have to move within proximal distance of all of the people. Through your narrow magnifying glass, you might be able to inspect the paperthin and unstable traces left as the groups are formed and dissolved. And in order not to get confused by the many other traces crossing the one you follow, you would have to equip the magnifying glass with blinders. You would also have to keep records on each individual in the crowd with detailed descriptions of the makeup of their associations with other individuals. But you would not be able to see far from this *oligoptic* point of vision as a result of your equipment that is easily blinded.

The ability to overview data and make sense of the overlapping monads is offered through digital methods preference for network graph visualisations. By visualising the monads in networks, a zoom function between the panorama and the oligoptica "behind" it is offered to the spectator (Venturini 2012). The panoramic navigational layer leaves every association of a monadic entity traceable, and at the same time functions as a proxy to all other information contained about the monad in the oligiptic layer.

---

<sup>2</sup> The panopticon was a prison system designed by Jeremy Bentham. It was designed for total surveillance, to induce total paranoia in the inmates and megalomania in the guards (Latour 2005, 181)

## 2.2.2 Cutting the network

*This section aims to shine a light on how to cut the network. Tarde's proposed monadic structure can theoretically expand endlessly in one way or the other and Latour's uniquely adequate account for ANT does not offer much help to make the decision.*

The strategy for cutting the network impacts the realm of possibilities for the following account. As exemplified with the HPV controversy: if filter A is applied, some pro-HPV actors will be excluded, but if filter B is applied, some con-HPV actors will be excluded. The entire monadic structure is dependent on what filters are applied.

In her article *Cutting the network*, Marilyn Strathern explains how the debates of cutting networks in anthropology dates back to the 1950's. The debates addressed the problem of certain kinship classes that could potentially lead to endless networks of relations, as they have no built-in stop-function. As she puts it: "[...] *indigenes may tell themselves that they are all related - trace far enough back and everyone share substance with everyone else*" (Strathern 1996, 529). She proposes that "*factors from diverse domains can affect the reach of an otherwise homogeneous network based on blood or family*" (1996, 530) and that such factors function as a method for both defining, expanding and cutting network.

Through an illustrative example, taken from the realms of ANT inspired science studies, Strathern show how two different types of networks can be assembled around the invention of a blood test for detecting hepatitis C virus. The first - a social network as Strathern calls it - forms a homogeneous network composed of the scientists working on the invention. The extent of the network "*appears to be bounded by the definition of who belongs to it*". In other words, if an actor recognise herself as a part of the network, she is a part of it. However, the network was cut in size as the corporation responsible for the research that led to the invention filed for a patent: "*Forty names to a scientific article became six names to a patent*" (Strathern 1996, 524). Strathern argue, that once the patent was introduced and the invention exposed to the world of commerce (another domain), it also introduced the question over what area the network spread, the answer being the ones who participated in the final spurt. The factor from the commercial world affecting the network was that of ownership as "*owners exclude those who do not belong*" (Strathern 1996, 524).

Strathern concludes that "*[ANT scholars ed.] will not have to look far in order to determine network length; they have always known that belonging divides and property disowns. So, where technology might enlarge networks, proprietorship can be guaranteed to cut them*

*down to size*" (Strathern 1996, 531). Understood figuratively, the technology referred to here, might be the proliferation of Tarde's monads through digital platforms. If the detailed information on user interactions on Facebook enables a continuous expansion of the network, then perhaps cutting the network is simply to follow the ownership signaled by the demarcation of the actors issuing the interactions as belonging to controversy.

## 2.3 Theoretical summary

It should now be clear that there are two types of networks in play in digital methods: the literal network graph and the descriptive actor-network. These two are not to be confused with one another. The network graph may be seen as a visual representation of Tarde's monads from *panoramic* and *oligoptic* point of views. The issue centric and medium centric approaches typically outputs network graphs. The actor-network may be seen as a description of the composition of the network seeking to answer questions such as what the ties are made of, what makes them durable and how they enter the item lists of the monads?

As such, the objects of comparison in this study, are the network graphs representing Tarde's monads.

Furthermore, it has been established that heterogeneous networks are to be treated symmetrically. This means that the distinction between human and non-human is broken down, treating all under the same label as actors. Actors can be intermediaries or mediators. Intermediaries can be seen as containers that carry meaning, without translating or modifying. The input is equal the output. Mediators also carry meaning, but here the output cannot be inferred by the input.

The study of controversies is also the study of issue publics. Issue publics are sparked into being, when public institutions are not able to handle the consequences of a transaction. As such, the groupings of monads in the network graphs depicting a controversy may be seen as different publics assembling around an issue such as AIDS research. In this definition of the public, these groupings will take on the task to change the state of affairs by impacting e.g. decision making and scientific knowledge production.

## 3 Methodology

The problem statement of this study ask *how the addition of monadic entries differs between the issue centric and the medium centric approaches*. This is investigated by designing and

conducting a comparative study based on a dataset representing a controversy, which will be addressed in the first part of this chapter.

The second part shine a light on how tool and method affect both the composition of as well as the conclusions drawn from the study.

Lastly the third part introduce the notion of ethnographic fieldwork in relation to harvesting data for the comparative study.

### 3.1 The recipe for a comparative digital study

*This section describes how this study operationalise the assembly of the monads through the issue centric and the medium centric approaches, respectively. Furthermore, the metrics for comparing the two monadic network graphs will be presented.*

A monad is understood as an entity with an item list. Each entry on the item list represents another monad, that the first monad is linked to. It may also include information about the quality of the link, what it is made of and how strong it is. This means, that at least two things need to be established in order to assemble the monad: what constitutes as as monad, and what constitutes as a link i.e. how does other monads enter the item list.

The issue centric approach builds the item lists by means of co-occurrences of issue specific keywords or phrases. If the issue specific keyword 'Aluminium' is found in the text of two Facebook posts, these two posts will be connected to each other. The medium centric approach builds the item list based on the way the platform natively orders social life i.e. by the means available to interact. If the same person has interacted with two posts on Facebook, then a link will be generated between the two posts, and they will be in each other's item list.

No digital platform offers options for researchers to extract neatly assembled issue or medium centric item lists to assemble the monads. And no research previously conducted in digital methods communities offer detailed protocols to follow.

The problem addressed in this chapter is therefore how the two novel and theoretical concepts of the issue centric and medium centric approaches are operationalised as practical research protocols for the assembly of Tarde's monads.

The task of operationalising the novel concepts into practical protocols is broken up in four sub-tasks and questions:

1. Data source: How and from where is data collected?
2. Building a common point of departure: What functions as delimiters for the dataset and how is it composed?
3. Networks as basis for comparison: How are the specificities of the two approaches translated into concrete manipulations of the dataset manifested in networks of associations?
4. The rules for comparison: What metrics are used to compare the two approaches?

Each of these sub-tasks will be addressed in the following sections.

### 3.1.1 The data source

This study identifies three requirements to the data and the data source put forward by the two approaches:

1. The data has to be accessible *as close to the source as possible*. This means, that as many data brokers, tools and other mediators as possible has to be eliminated. This is important to foreground the specificity with which the medium orders social life.
2. All actors have to be *uniquely identifiable* and the identifies has to be left behind as traces every time an actor acts in order to build the monadic structures of the medium.
3. The traces left by the actors must be in the form of *textual entries or thematic labels* in order to assemble the monadic structures based on the issue centric approach. This means that messages or tags must be available to identify controversy specific issues.

Several platforms meet these requirements. Twitter for instance allows researchers to harvest data “live” and close to the source; actors are uniquely identifiable, and entries are tagged with thematic labels. The problem with Twitter is that it does not allow harvest of entries older than six days.

Facebook also meet the three requirements listed above. All entries are uniquely identifiable, which means that the monadic networks can be compiled both by following the identifiers throughout the dataset (the medium centric approach), or by thematic labels on textual entries made by the users (the issue centric approach). Furthermore, Facebook offers

extraction of data through what is known as an Application Programming Interface (API) i.e. as close to the source as possible. It also allows harvest of historical data, dating back to when the platform was introduced in 2006. Therefore, Facebook will function as the medium from where data is harvested to assemble the monads in this study.

This study chooses to use Facebook posts as monads in the network graphs. The reasons why are that posts are uniquely identifiable, contains information on users' interactions as well as textual entries. This means that both the medium centric approach and the issue centric approach are able to assemble the monadic network graphs.

An API is a construction set up to safeguard a repository of data. It is mainly setup to make data available for developers creating functions to the platform such as games and quizzes, but also for embedding platform content externally e.g. in news articles. Developers can access the repository of data through the API if they meet certain criteria. In general, these criteria concern the right level of authentication and the use of the right language and vocabulary.

Ultimo November 2017, Facebook announced, that as of February 6th, 2018 restrictions would be made on certain data points of their API (Facebook 2017). Consequently, it is no longer be possible to harvest unique identifiers of the users who act on public posts published by pages through the API. The same changes were announced to apply to public groups and events as of April 28th with the 2.12 version of the API (Facebook 2018B).

These changes negatively impact research into the ontologies embedded in Facebook's way of ordering social interactions. Associations can no longer be traced to their originator, and as a consequence the medium centric version of the Tardian monad can no longer be assembled.

However, in light of the looming perspectives this loss of access would imply for digital methods studies, the author of this thesis in collaboration with his supervisor started harvesting the Danish public Facebook conversations between 2012 and 2017. As of February 5th, 2018, the day before the API changes went into effect, 72.000 pages had been harvested, encompassing roughly 25 million posts, 600 million reactions, 125 million comments generated by around 12 million unique users.

Facebook's data guardian is called the Graph API. The Graph API is a representation of the information on the platform and is named after the idea of a "social graph" (Facebook

2018A). Facebook's social graph is composed of *nodes* - representing individual objects, such as a user, a Photo or a page and *edges* - representing connections between a collection of objects (Facebook 2018A). Edges can be both in the form of other objects, but also actions such as Reactions, Tags and Shares. E.g. if a user tagged in an object such as a Photo or a post, an edge will be made between the user and the object.

Nodes can have *fields*, which are attributes attaining to the node such as an *About section* or *Location* of a page or the *Gender* of a user. These attributes can be either private or public, which is determined by the privacy settings of each object i.e. the specificity of the medium. In this thesis, all attributes concerning users are characterised as private whereas those of a page are characterised as public<sup>3</sup>.

Access to the API is obtained by creating an app via the developer section of Facebook<sup>4</sup>.

Authentication comes in form of access tokens. Access tokens act as keys to specific data points in the repository safeguarded by the API. If the key does not match the access level of the data point, access is denied. If the key match, the data is previewed in the form of a file which can then be downloaded and manipulated at will.

There are three different types of access tokens (user, App and page) for unlocking different views into Facebook as represented through the API. The type used for the study of this thesis is the App token. Unlike the page and user token the App token does not need permission from users to access data through the API. It can thus be said to preview only the public content of the platform. Or at least what the platform defines as public content.

In order to extract data through the API, the right language and vocabulary must be obtained. Several programming languages are supported through different Software Development Kits (SDK). For this thesis the programming language Python was chosen, partly because the author was already familiar with the language and partly because it has a reputation for being able to handle calculations on large amounts of data.

A query to the API is called a request. For each request a package containing the access token as well as information about the objects that should be returned is send via Hypertext Transfer Protocol<sup>5</sup> (HTTP) to the API.

---

<sup>3</sup> This is of course a point which needs further development and will be discussed in the section *Post-demographic and natively digital data*.

<sup>4</sup> developers.facebook.com

The answer from the API to a request comes in the form of a JSON (JavaScript Object Notation) file. JSON files contain text which has been formatted in a tree-like structure as seen in

Figure 2. The text contains information about the objects inquired upon.

Figure 2 - This figure shows how the data in the JSON files is formatted in a tree like structure.

```
{
  "post_id": "294878953399_10152129482201081",
  "post_time": "2013-12-11T13:35:21+0000",
  "post_message": "\"Christmas comes early!\" or \"I've been good this year\" (my birthday on friday.)",
  "post_author": {
    "name": "Øyvind Dyrnes",
    "id": "10156004026591081"
  },
  "post_picture": "https://scontent.xx.fbcdn.net/v/t1.0-0/s130x130/1456497_10152129482201081_1326119659_n.jpg?oh=6aa6a37607d5646715e9cc5c5f856b26&oe=5B1E99E5",
  "post_link": "https://www.facebook.com/photo.php?fbid=10152129482201081&set=o.294878953399&type=3",
  "post_reactions": [
    {
      "id": "10157084354452571",
      "type": "LIKE"
    },
    {
      "id": "294878953399",
      "type": "LIKE"
    }
  ],
  "comments": []
}
```

All this communication back and forth, to and from the API is automatically handled by Python scripts written specifically for the purposes of collecting the data used in this study.

### 3.1.2 Building a common point of departure

As I have now established I make use of harvested data of digital traces left on Facebook through the Graph API to assemble the monads. In order to delimit this empirical test and make the comparison between the two approaches tangible for analysis I have selected a case representing a controversy.

Simultaneous with writing this thesis I have been employed as a student assistant at the Techno-Anthropological laboratory (TANTLab) at the Aalborg University campus in Copenhagen. The main focus of this employment has been to participate in a interdisciplinary research project *Public Health at a Crossroads* (PUSH) led by anthropologists at Aarhus University and co-funded by The Danish Cancer Society and Aarhus University Research Foundation respectively (Aarhus University 2018).

The aim of the project is to develop epidemiological and anthropological understanding of the role of social media in the Danish controversy revolving around the Human Papillomavirus (HPV) vaccine introduced in Danish vaccination program in 2009.

---

<sup>5</sup> HTTP can be compared to a mail service that brings text from one place to the other on the World Wide Web

As a part of this enterprise a subproject dealing with conversations on Facebook was executed. This led to a beneficial situation where the dataset I developed for the study presented here, could also function as a central input for the PUSH project.

The result is a HPV-dataset containing 3,059 entries relating to the HPV controversy. Developing the HPV-dataset has been a large part of the work undertaken in this study.

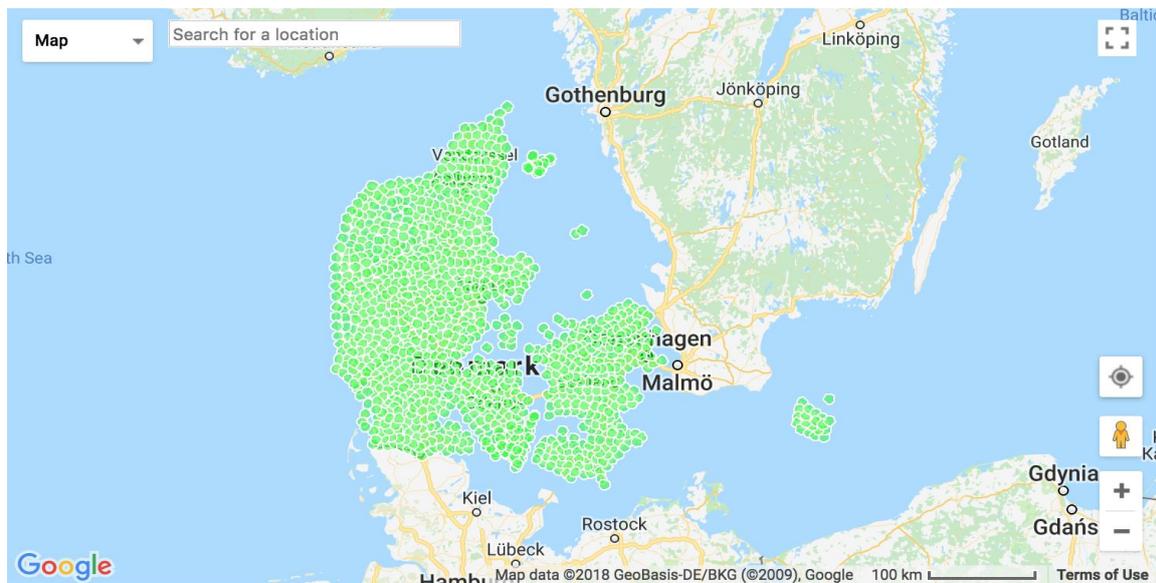
The HPV dataset is a combination of two different datasets I have developed for this study. The first contains manually curated Facebook pages and the other geographically curated Facebook pages.

I have constructed the manually curated dataset by doing research into what Facebook pages in one way or the other are associated with HPV. It includes pages ranging from activist communities declaring pro or con allegiance to the HPV-vaccine over official information fora, self-help fora to pages concerning indirect associated topics such as vaccines in general. I have not made the decisions on what pages to include in the manually curated dataset myself. Several of the members of the PUSH project has taken part in the curation. The manually curated dataset consists of the content from 102 pages.

The geographically curated dataset is based on pages that has a geographical association to Denmark. As such, it contains all sorts of different pages that has an explicit association to Denmark. This results in a dataset consisting of the content 72,000 Danish pages in the period between 01/01/2012 and 31/12/2017. This amounts to roughly 25 million posts, 600 million reactions, 125 million comments all in all generated by around 12 million unique users.

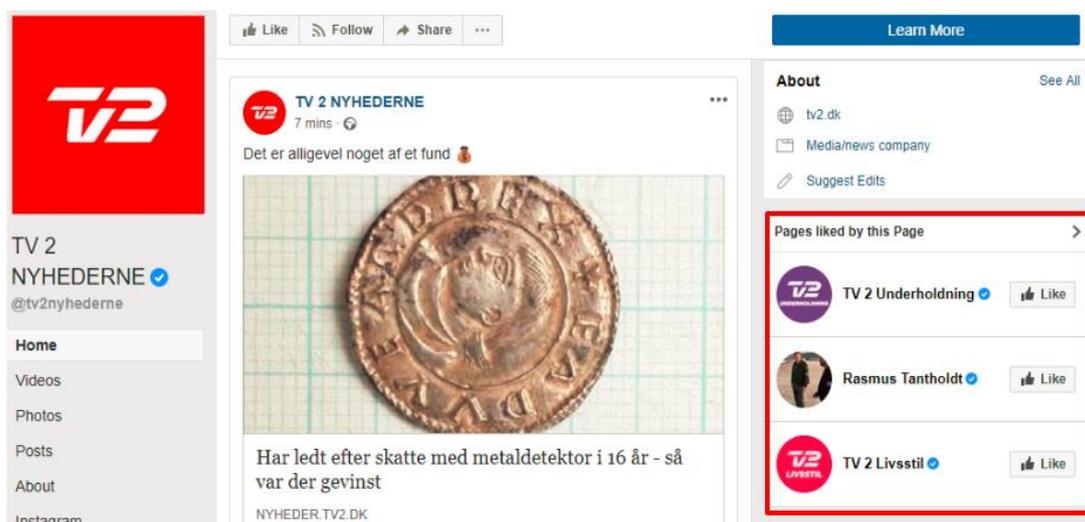
The starting point of geographical curation is a list of 1,086 coordinates of Denmark. Each set of coordinates represent a circle of 5 km within the Danish territory (Figure 3).

Figure 3 - This map is made using the “radius around a point” function on the website <https://www.freemaptools.com/>. Each node represents a circle with a radius of 5 kilometres. The geocoordinates of all the circles are outputted as a list



Through the search function of the API, it is possible to get a list of all the Facebook Locations<sup>6</sup> encompassed in such geographical circles and furthermore to get a list of pages affiliated with the Location. All the pages extracted from these Locations function as starting points of an expansion of the list of Danish pages called snowballing. Snowballing entails that each entry in the list is asked to point out new candidate members. Such candidates are identified as other pages liked by the initial pages (Figure 4).

Figure 4 - This figure shows the Facebook page of TV2 Nyhederne. The red frame indicates some of the other pages liked by TV2 Nyhederne (Facebook 2018D)



<sup>6</sup> The Location object on Facebook represents a geographical “offline” location. It is e.g. used by users to check-in at attractions and sights and by pages to tell where they reside in the physical world.

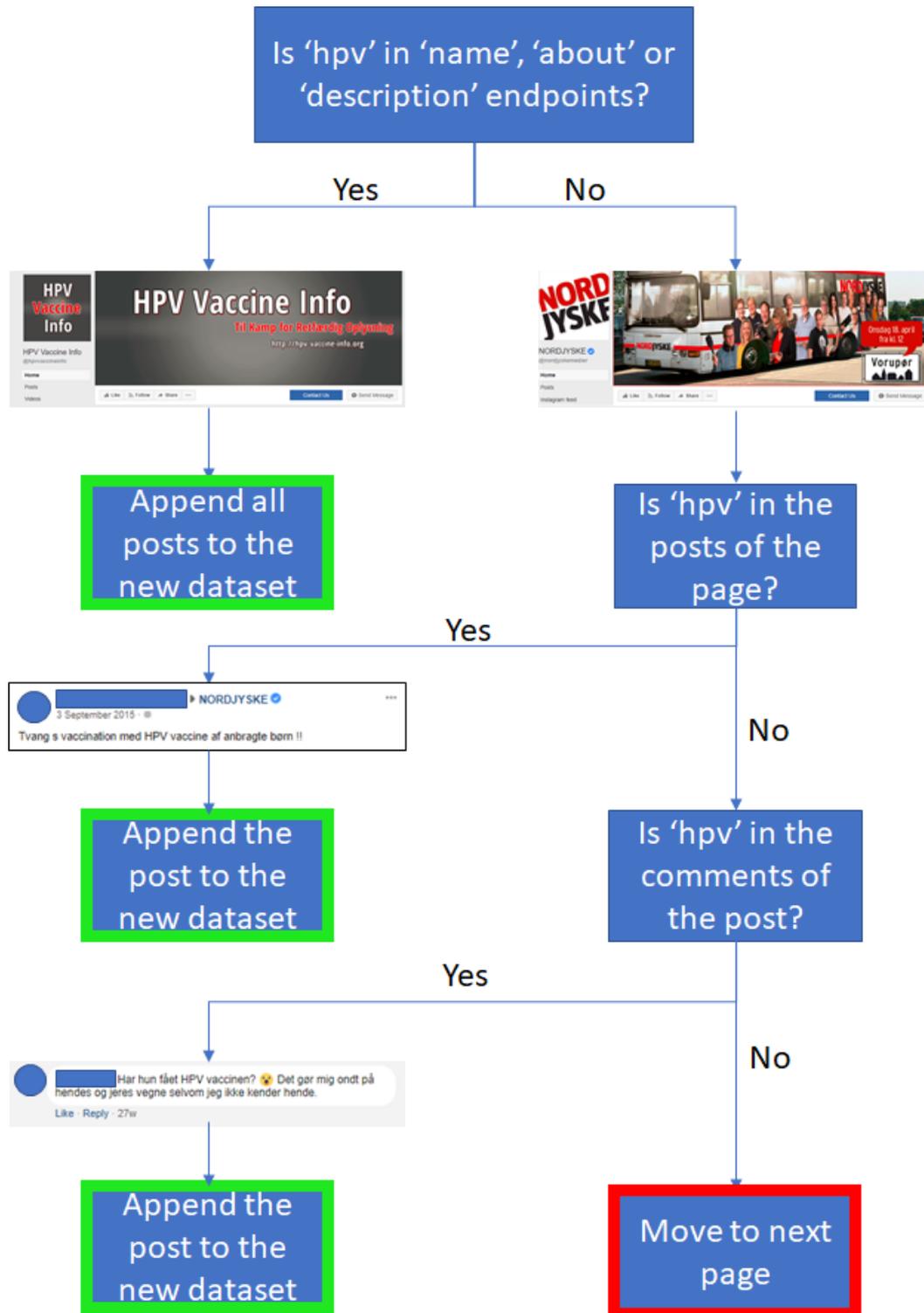
The candidates are then qualified by the questions: Are you grounded in an offline location? Is that offline location based within the borders of the Danish territory? If yes, the page is appended to the final harvest list and the snowball continues, starting from the page just appended. All new candidate members are asked the same questions. This process continues, until the snowball has reached the foot of the hill and run out of inertia. What has been picked up, and not discarded as foreign, constitutes the final list of pages to be harvested i.e. the 72.000 Danish Facebook pages.

### 3.1.2.1 Delimitation and combination of the datasets

Before combining the manually and the geographical curated datasets into a HPV dataset, delimitation is done based on the content's relation to HPV.

As mentioned earlier, this study chooses posts to be the monads in the network graphs. Therefore, posts are appended from the manually curated and the geographically curated dataset to the HPV dataset if they are directly associated with the controversy. This means that if 'HPV' has been mentioned in the message or comment thread of a post, the post is considered as directly associated. If the post has been published by a page where 'HPV' is present in the page name, about section or description it is considered as indirectly associated. The selection process is depicted in Figure 5.

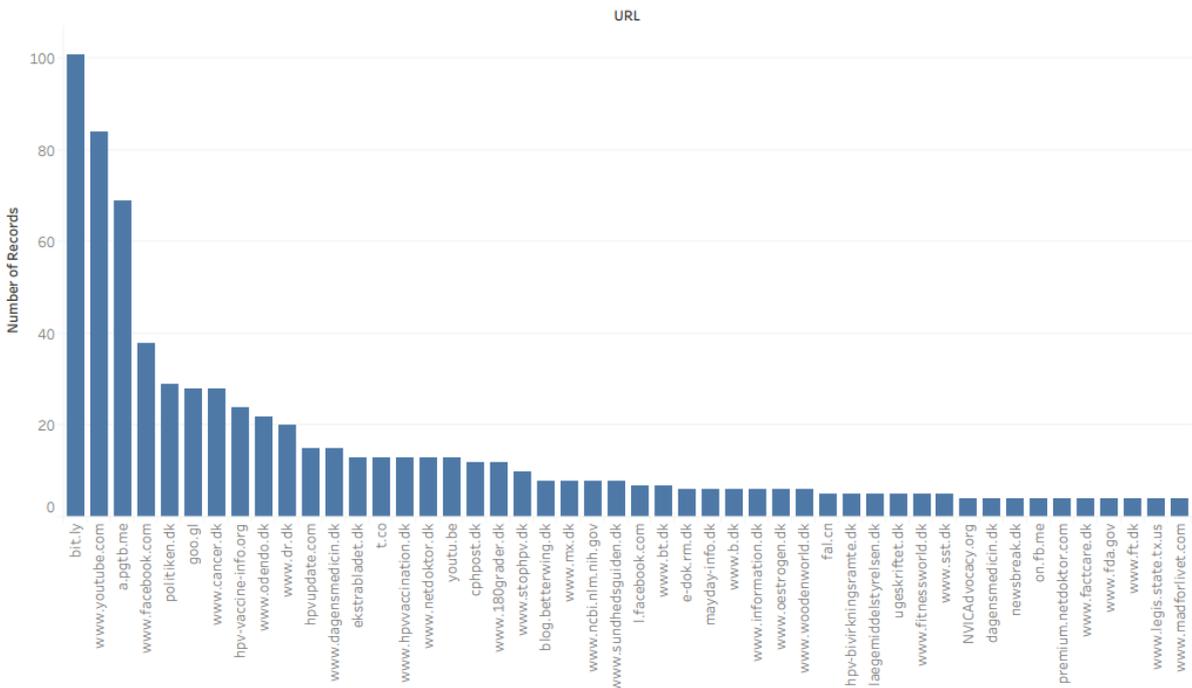
Figure 5 - This diagram depicts the selection process of the posts to include in the dataset. The green outline indicates that the post(s) are to be appended, and the red that the script should continue to next the page.



Having conducted the selection process on the geographically curated dataset narrows down the list of prospect posts from 25 million to 2,294. Removing duplicates based on the unique identifiers of the posts, reduces the number from 2,294 to 2,085.

Looking closer at the 2,085 prospects reveals that several posts are included because ‘HPV’ is present as part of a link. Most links are genuinely associated with the HPV controversy. A small subset of the links does not have any affiliation to the HPV controversy but are generically generated references as part of so called short links. As visualised in Figure 6 short links such as bit.ly, goo.gl and t.co take up a relatively large amount of the total amount of links.

Figure 6 - This graph shows the amount of the top 40 most commonly found sources of links in the geographically curated dataset.

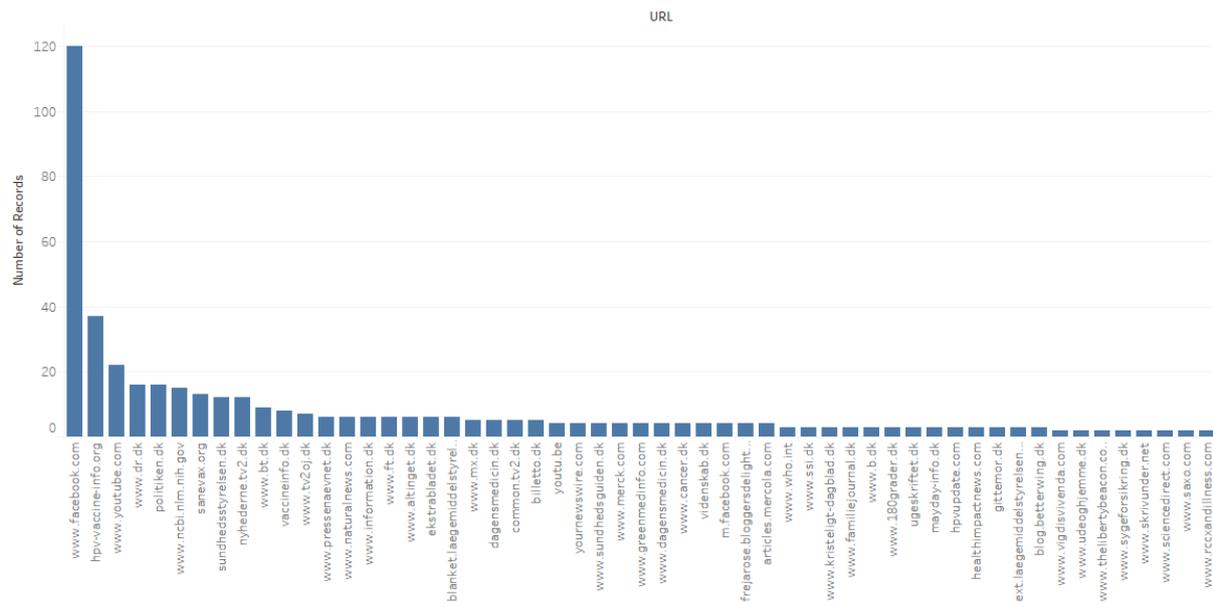


Removing posts that only refers to ‘HPV’ in a short link generates a final output from the geographically curated dataset amounting to 1,965 posts.

Conducting the same selection process on the manually curated dataset leaves us with 1,127 prospect posts. Removing duplicates narrows this to 1,111.

Performing a similar link analysis as above reveals, that ‘HPV’ only found in short links is not a problem for the manually curated dataset (Figure 7).

Figure 7 - This graph shows the top 50 most commonly found sources of links in the manually curated dataset.



The two datasets are now combined, resulting in a total of 3,059 posts from 771 different pages. The overlap between the two original datasets only represent two pages and 17 posts. This relatively low amount of overlaps is interesting, as it problematizes the two methods of generating datasets. This will be further explored in relation to digital methods in general in the section *Post-demographic and natively digital data*.

### 3.1.3 The operationalisation of monadic network graphs

*The following section describes how the network graphs representing Tarde's monads has been assembled by following respectively the issue centric and the medium centric approach. This is done through the HPV dataset introduced in the previous section.*

The monadic network graphs assembled for this study consists of *nodes* representing Facebook posts, and *edges* representing a link between the nodes. These links can be anything shared between the posts such as users' interactions or the co-occurrence of issue terms. Edges can be weighed. This means that the link between two nodes can be made stronger or weaker. This is e.g. done to represent how many users or issue terms link two posts together. By the use of the force-vector algorithm ForceAtlas2 the edges are made to function equivalent to springs. This means that they either repulse or retract what is connected to its ends (Jacomy et al 2014). In a network consisting of two nodes, if no edge is present, the nodes will be forced further and further away from one another. Visa versa if an edge is present, the stronger the weight, the closer it brings the two nodes to each other.

### 3.1.3.1 Operationalisation of the issue centric monadic network graph

The protocol for assembling the issue centric monadic network take inspiration from the concept of *issue mapping* as developed by Noortje Marres and Richard Rogers through what is labelled the *interface method* (Marres 2015; Marres and Rogers 2000; Marres and Gerlitz 2015). Furthermore, inspiration is taken from Participatory Data Design (PDD) as practiced through data sprints, in the way the issue specific vocabulary is established (Jensen et al 2017).

Data sprints are intensive workshops comprising the necessary competencies to pose and evaluate the relevance of research questions to a controversy and to make prototypes of the proposed outputs (Munk et al 2018). The necessary competencies encompass digital methods researchers and experts on the issues of the controversy (issue experts). As a part of the PUSH project, a data sprint was held ultimo May 2018 to assess the impact of social media on the HPV-controversy. Participants counted medical practitioners with expertise in HPV and side effects from the HPV- vaccine; anthropologists with intimate knowledge about those who suffer from HPV-vaccine side effects; and digital methods scholars including myself. As such, this data sprint served as an important opportunity to validate or disprove some of the assumptions I had about the issues in the controversy.

Issue mapping has roots in the Actor-Network Theory related method of *co-word analysis* (Marres and Gerlitz 2015, 28). In the 1980's Michel Callon and his colleagues, developed this method to detect "*pockets of innovation*" by highlighting emerging topics in scientific literature (Callon et al. 1983). As the name implies, co-word analysis detects words that occur in each other's vicinity in a corpus of text. The connection between two words is weighted (that is made stronger or weaker) based on the distance between the words in the text, as well as on the quantity of connections (Marres and Gerlitz 2015, 28).

Marres and Gerlitz interface method (2015) both draws inspiration and deviates from the digital methods approach outlined by Richard Rogers' prescription to follow the medium: "*while our approach seeks to repurpose 'natively digital' methods, we are interested in engaging critically and creatively with the remediation of social research methods in digital culture*" (Marres and Gerlitz 2015, 31). As such the method aligns with the ANT requirement of following the actors as "*it makes it possible to detect changes in topical associations over time without having to rely on previously defined categories, and their implied criteria of relevance*" (Marres and Gerlitz 2015, 29).

### 3.1.3.1.1 The issue dictionary as a way to operationalise topical networks

An *issue dictionary* is a list of keywords associated with an issue which can be used to generate networks. The entries of the issue dictionary is used to label entries in the dataset in a uniform way. An issue dictionary can be assembled by following various methods.

The first is by an automated extraction of words from a corpus of text from the dataset, as exemplified by Marres and Gerlitz in their issue mapping of *climate change* on Twitter (2015). The words are counted and assessed based on different parameters e.g. by introducing a list of stop-words that should not be included. A problem with this method is to assess whether the words are results of *digital bias* of the medium or actually associated with the issue (Marres 2015).

The second method is by the help of issue experts. Issue experts function as representatives of the issue publics represented in the data. They are assigned the authority to determine what keywords are of relevance and take part in the construction of the issue dictionary. A problem with this method is to make sure, that the issue experts not only represent one side of the controversy in order to avoid partial bias. In the case of the HPV data sprint mentioned earlier, it was therefore important to have both the perspectives of the practitioners as well as the perspectives of those affected by HPV-vaccine side effects. The latter was obtained by proxy of the anthropologists with intimate knowledge of those affected.

I propose a third method combining the two presented above. The third method entails automated extraction of words from the data, which are subsequently analysed by issue experts. By allowing the issue experts to help assess the list of extracted words from the text of corpora, the problems of digital and partial bias are arguably minimized. The method is not novel as such but has been used on several data sprints I have participated (Public Data Lab 2017; Bach et al forthcoming).

Automatic extraction of potential keywords for the issue dictionary is done through digital tools. Tools such as Cortext<sup>7</sup> and ANTA (Actor-Network Text Analysis) (Venturini and Guido 2012) are often used to extract n-grams<sup>8</sup> for the issue dictionary. These tools are often built on concepts from Computational Linguistics (CL) and Natural Language Processing (NLP) (Clark et. al. 2010). NLP is a concept where linguistics combined with some degree of

---

<sup>7</sup> Cortext.net

<sup>8</sup> An n-gram is a string of text constituted by n items.

machine learning capability enables automation of textual analysis. As such, a program built on NLP is able to “understand” a sentence in terms of nouns, verbs, adjectives etc.

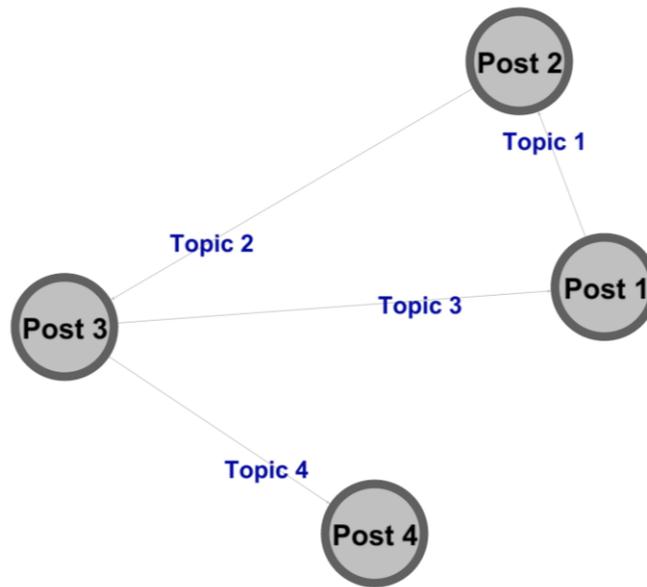
The corpus of text used in this study is in Danish language. Currently, the methods for conducting open source computer aided NLP on Danish language text are quite limited and the ones attuned to Danish language are empirically untested within the digital methods community. As a consequence, this thesis uses a simpler method for extracting keywords which is based on word count, summarising the number of times a word occurs in the corpus.

New problems are foregrounded using the method of word count. Generic words also known as stop words such as “is” or “are” overshadow the more interesting issue specific words such as “HPV-vaccine” or “aluminium”. As a countermeasure the stop words will be filtered out based on a combination of two lists of stop words (Torp 2018; Diaz 2016).

During the HPV data sprint, the list of most common words was validated by the issue experts. The issue experts assessed each keyword by its relevance to the controversy, and finally grouped keywords in issue specific topics. Each post in the dataset was then tagged by topics, if a keyword associated with the topic was found in the text of the post. This process can be compared to manually reading through each post and labelling it topically. The list of the topics and keywords can be found in appendix A.

The issue centric representation of the HPV controversy the monadic network graph is constructed by the occurrence of a topics e.g. “Science” in two different posts. If two posts in the dataset has both been tagged with the topic “Science” they will be associated by a link. As such, the network representation is not per se in line with the co-occurrence network introduced earlier. In the co-occurrence network, topics function as nodes. If two topics occur in the same text, a link will be generated. In this study the text (i.e. the post) function as nodes, whereas the topics function as links (Figure 8)

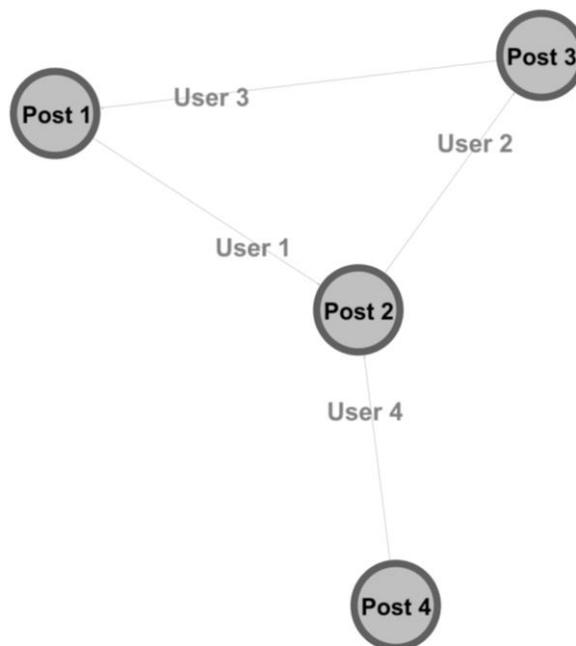
Figure 8 - This graph shows a simple post to post network linked by co-occurrence of topics.



### 3.1.3.2 Operationalisation of the medium centric monadic network

The medium centric approach for assembling the monadic network graph follows the ontology by which Facebook orders social life. This means that the posts in the dataset are linked through the Reactions and Comments of the users. This furthermore means that no additional information is to be appended to the dataset in order to assemble the network.

Figure 9 - This graph shows a simple post to post network. The posts are linked by co-occurring user interactions.



In the medium centric representation of the HPV-controversy the monadic network graph is constructed based on users' interactions with posts as depicted in Figure 9. Two posts will be linked if the same user has engaged with both posts. This means, that the list of users interacting on each post, is compared with the lists of all the other posts in the dataset severally. The weight of the link between two posts is determined by the number of overlapping user interactions between the two posts.

### 3.1.4 Metrics for comparing the two approaches

The metrics for comparing the issue centric and medium centric representations of the monadic network graphs in this study take inspiration from Latour's notions of *panorama* and *oligopticon* as introduced in the section *Panoramas and oligopticons*.

Network visualisations can be understood as the sorts of *panoramic* views of monadic structures. They portray a version of an overview, but not the distinct features of each of the monads. The entity list of the individual monad with all connections to other monads - as well as the attributes of the connections - are not for inspection. In order to render them visible, they have to be inspected severally through the narrow and short sighted '*oligoptic*' view of the monad.

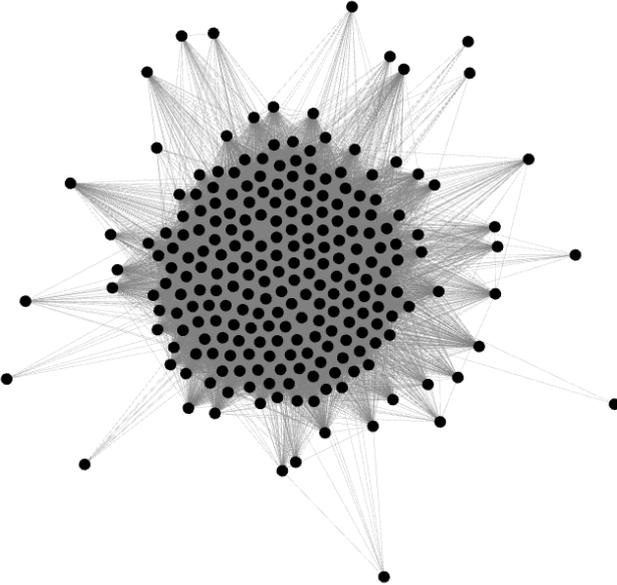
#### 3.1.4.1 Operationalisation of panoramic metrics

The panoramic metrics for comparing the issue centric and the medium centric representations of the monadic network graphs revolve around the distant view of the monadic composition and structure. One does not have to look far for an operationalisation of measures for composition and structure. In digital methods projects such operationalizations are common practice when evaluating networks eligibility for analysis. These evaluations often revolve around the same measures for composition and structure as is needed in this study. They build on a vocabulary composed by concepts from mathematics (e.g. clusters) and geography (e.g. centrality and bridging) (Venturini et al 2015, 1-2). This thesis will mainly draw on concept of *clustering* for evaluating structure and composition.

In order to make meaningful interpretations of a network it must show tendencies of *clustering*. Comparing to the analogy of the concert venue as described earlier, once the concert has started the audience attention is directed at the stage making it difficult to separate one group from the others. The same goes for networks. If some force is present that draws the attention of all other entities, it distorts the particularities of the smaller groups. And particularity of groups is exactly what is looked for, qua the allegiance with Actor-

Network Theory. This is exemplified in Figure 10. The figure shows a network of co-occurrences of words in texts. Because the words in the centre of the network are generic and therefore present in all textual entries, they are more likely to co-occur with all the other words in the network. This in turn generates what is known as a “hairballed” network. From such a network, it is difficult to pinpoint the particular groups of words.

Figure 10 - This graph shows a "hairballed" co-occurrence network based on words in texts.



Another network of Twitter data as shown in Figure 11 is much more diverse, allowing for the interpretation of clusters A, B and C.

Figure 11 - This shows a network where it is possible to identify three clusters (A, B and C) (Venturini et al 2015)



As a help to determine clusters, different calculations can be used. For example, the *modularity score* is an indicator of how clustered (or modular) the network is (Blondel 2008). A low modularity score indicates that few clusters are present in the network. This could mean that each node in the network is equally connected to all the other nodes. A high modularity score indicates that many clusters are present.

Although the modularity calculations by many is seen as a standard way of identifying clusters it is wise to be critical towards the results. The algorithm only takes structural parameters of the network into consideration, meaning that it cannot identify and exclude the generic nodes of the network nor does it take into account the topical content of the posts.

This means, that the network in some cases has to be filtered before clusters can be identified. This is typically done through degree measures where the connectedness of nodes is calculated and assessed. But due to the composition of the issue centric network, such filtration cannot be expected to exclude generic entries. This is because the generic qualities enter the network through the issue dictionary and are present in the network as links between the nodes. As such, clustering can best be obtained by manipulating the entries in the issue dictionary.

Assessing whether the *right* kind of clustering has been obtained is very much an up to the researcher to determine based on the tools available, the visual representation and experience. As a help, the researcher can ask the question: *Can meaningful interpretation of the network be made?*

In this study, the difference and similarities between the clusters of the issue centric and the medium centric network graphs will be used as a metric for comparison. As the networks are composed from the same dataset, such comparison includes an analysis of what entities are clustered together in the two approaches respectively. Furthermore, the structural comparison will be followed up by a topical comparison. As such, it will be analysed whether the clusters identified are of the same topical makeup across the two approaches.

#### 3.1.4.2 Operationalisation of oligoptic metrics

Imagine that you would be able to step into the network graph visualisation and take the place of one of the nodes. What would you see from this point of view? Probably only the links going from that specific monad to all the others in its network. Equipped with the magnifying glass and blinders these links can now be followed and studied severally.

The oligoptic metrics for comparing the issue centric and the medium centric representations of the monadic networks revolve around the proximal and narrow monadic point of view. This entails evaluating the item list of each monad severally.

The networks compared in this study is composed of up to as many as 3,059 monadic entities in the form of posts relating to other posts. A study analysing each of these entries severally would be an incomprehensible task for this study. Instead, a criterion is developed to select a few cases representing different types of monads.

This study proposes to use *degree* to inform such a selection. Degree covers the number of other monads a single monad relates to. In this study, degree potentially range from one for the least connected monad and 3,058 for the most connected monad.

The monads chosen for the comparison will be the monad with the highest and the lowest degrees in each network as well as a monad representing the average degree of each network. These cases are argued to represent both the best, worst and average monads in terms of their ability to assemble other monads in their item lists.

This setup allows for an analysis of what constitutes as connectedness and centrality in the two different approaches, which in turn helps inform how the addition of monadic entries differs between the two approaches.

### 3.1.5 Summary of the comparative study

In this chapter a recipe for the comparative study has been proposed and argued. The monads will be assembled based on a common dataset. The dataset has been composed by both manual and geographical curation and delimited by the selection of a case. The case selected is the HPV-controversy on Facebook, which is an offspring of a research project called PUSH.

The entities selected as monads are the Facebook posts from the HPV-dataset. These monads will function as the nodes in two distinct network graphs following the issue centric and the medium centric approaches respectively.

The issue centric and medium centric approaches define links in the monadic network graphs in two different ways. The former makes use of an issue dictionary co-composed by participants in a data sprint on the HPV controversy. This means that a link will be created

between two monads, if they share the same issue related topic. The latter make use of Facebook's own way of ordering social life. This means that a link will be created between two monads, if the same user has interacted on both posts.

Lastly, the metrics for comparison are based on two points of view. The panoramic comparison looks at the structural differences and similarities of the network graphs. This includes identifying structural and topical clusters and comparing across the networks. The oligoptic comparison take the monad's item list as its starting point. Identifying the most and least connected monad in each network as well as a monad representing average connectedness, informs how connectedness and centrality is distributed by following the two approaches respectively.

### 3.2 Methodological challenges in digital methods projects

*This section aims to shed a light on some of the problems encountered when repurposing digital data in this study.*

Repurposing digital data and tools to investigate social scientific questions is a central aspect of digital methods (Rogers 2009; Munk forthcoming; Marres and Gerlitz 2015; Latour et. al. 2012). Links on webpages can be seen as a way to ease visitors' navigation to content that is considered relevant but has in digital methods been repurposed to map authorities and hubs to aid the navigation in controversies (Munk 2014) through tools such as the Issue Crawler (Rogers 2018) and Hyphe (Jacomy et al 2016). Data from the Facebook API might be intended to help developers build applications for Facebook users but is repurposed in digital methods to map actors' engagement with issue topics (Public Data Lab 2017; Rieder 2013; Bach et al forthcoming) through tools such as Netvizz (Rieder 2013).

Harvesting digital data is not a neutral activity. It cannot be taken for granted that the data repurposed for digital methods analysis are unaltered copies of the order of the medium (Marres and Gerlitz 2015; Rogers 2009; Rogers 2012).

The first part of section problematizes the concepts of *natively digital* and *post-demographic* data in relation to the comparison between the issue and medium centric approaches. The second part addresses the *digital bias* (Marres 2015) exerted by tools, methods and socio-technical constructions such as an API. The third and last part introduce how the study comply with the newly introduced General Data Protection Regulation (GDPR).

### 3.2.1 Post-demographic and natively digital data

Richard Rogers argue that the *natively digital* is that which has been born in digital settings as opposed to what has been *digitized* (2009). Traditional demographics such as race, age, income, location and gender does not apply to the natively digital (Rogers 2009). Instead they have been replaced with what Rogers calls *post-demographics* employed 'info-politically' (Rogers 2004). This means that post-demographics such as tastes, interests, favourites, replies to invitations, attendance in events and likes of post are natively used to organize the distribution of information to the right people.

In this light, claiming that it is possible to create a dataset representing *the Danish public* as the geographically curated presented earlier is problematic. To introduce the demographic delimitation of something being *natively Danish* to Facebook data is on the one hand an arbitrary move. This is because Facebook data is inherently post-demographic and natively digital. But on the other hand, Facebook does offer a way to bridge the divide between the 'virtual' and the 'real'. The bridge is in the form of the quasi-demographic *Location*, that offers the means to 'offline' ground the online. The latter is also acknowledged by Rogers in his terms *online* and *offline grounding* (2009, 1).

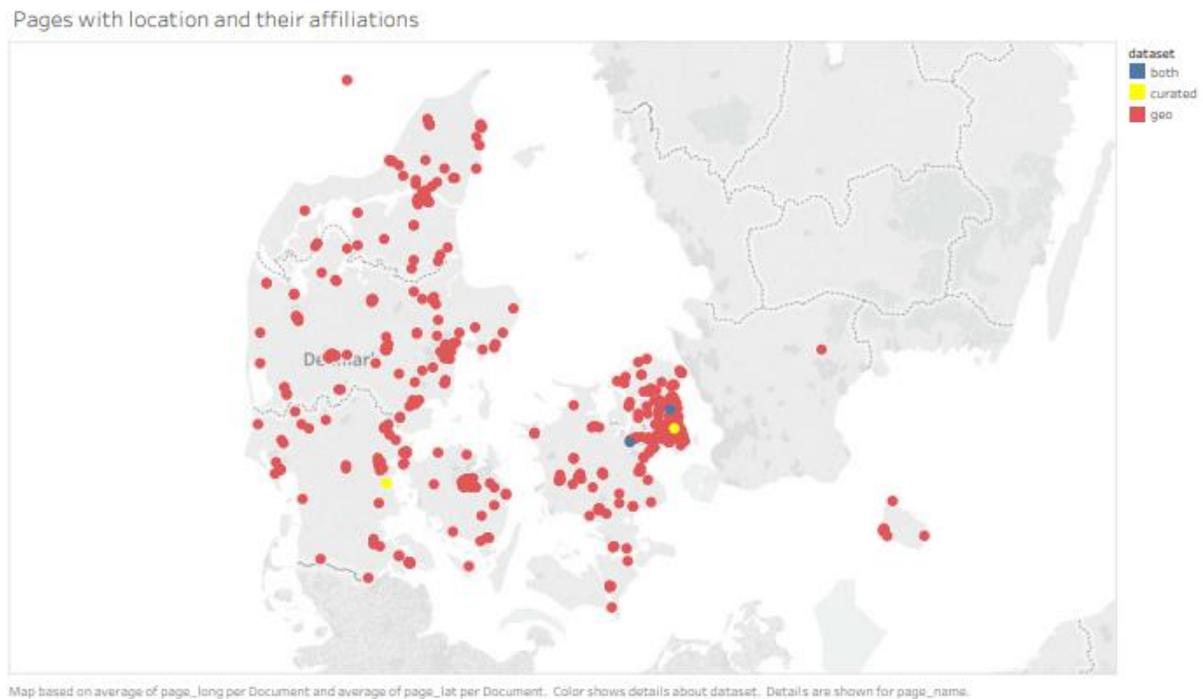
The problem then, is not the claim that the dataset represents a Danish public but to claim that it represents the Danish public. The latter would demand a way to be sure that the dataset is based on finite list of Danish pages, a list which does not publicly exist.

In order to grasp the extent of the problem, this a test was conducted to compare the two ways of curating the basis for the HPV dataset, geographically and manually respectively. The test was carried out before the datasets were combined.

Starting off by looking at the number of different pages included in the dataset shows that the manual curation represents 11 pages whereas the geographical curation represent 763. Only two pages overlap between the two datasets. This indicates, that the method used for generating the geographically curated dataset, might not have been able to capture all Danish Facebook pages.

Due to the approach for assembling the geographically curated dataset, all 763 pages contains location information. Looking at the 11 pages from the manual curation, only four of these contains information about the offline location (Figure 12).

Figure 12 - The figure shows a map of Denmark, where the pages with geographical information has been plotted. The red nodes represent pages from the geographically curated dataset, the yellow from the manually curated and the blue are overlaps that appear in both



The consequences can potentially be severe for the conclusions drawn from projects relying on natively Danish content. In the HPV example, interest was specifically on the Danish controversy. The problem in this case is that questions regarding representation can be difficult to answer, as there is no way of knowing how much of the controversy is covered with the given dataset.

Although the problem is of interest for digital methods in general, the methodological implications for the comparison carried out in this thesis are at worst considered minimal. As such, no further precautions are taken in terms of what public is encompassed in the dataset as part of this thesis, although the questions are still of high relevance in the digital methods communities in general.

### 3.2.2 Digital bias

The *God eye trick* where the analysis of a phenomenon is presented as pure and unaffected by the apparatus with which the phenomenon is observed has for long been critiqued within Science and Technology Studies. The argument that tools and methods impact observations and measurements of a study is well versed and established.

According to Marres and Gerlitz, the tools and methods used in the process of retrieving and analysing the repurposed digital data impact the resulting study.

*“Data tools make active contributions to the configuration of analytic practices through their sampling techniques, options for analysis and modes of visualization, which in turn help to concretize regimes of valuation”* (Marres and Gerlitz 2015, 31).

A significant problem faced by the transition from offline to online studies of the social, is *digital bias*. Digital bias encompasses the impact of the platforms themselves on the analysis (Marres 2015, Madsen 2012, Venturini and Guido 2012). The analysis of an object risks being undermined as researcher cannot be sure whether they study the object itself or the digital settings that render the object analysable.

According to Marres, digital bias can be approached in either a *precautionary* or an *affirmative* way (Marres 2015). The precautionary approach considers the influences of the digital settings as a source of noise which must be eliminated. Such elimination can be done by using multiple sources of data or by identifying the source of the bias and taking countermeasures.

What is seen as sources of noise by the precautionary approach, is by the affirmative approach seen as constitutive aspects of the controversies online. As such digital bias offers an empirical opportunity to study the representation of constitutive dynamics of the medium and the specificity with which social life is ordered through the medium.

The different ways users can react on posts serves as an example. In February 2016, Facebook introduced a new portfolio of ways to react. The familiar “Like” is now accompanied by five other types of reaction such as “Sad” and “Love” (Figure 13). Such changes quickly become the new normal for users but can have obvious implications for studies stretching over a longer period of time.

Figure 13 - This figure shows the six different ways of Reacting on Facebook introduced February 2016 (Facebook 2018E)



This means, that the dataset compiled for this study encompass data from four years where only the “Like” function is available and another two where the entire portfolio is in use. In this example it is of course quite easy to establish that the difference is a consequence of the medium introducing changes to the platform. But in other cases, it might not be as obvious.

This thesis argues, in line with the affirmative approach, that the contributions of the tools and methods used should not be targeted for immediate elimination. Instead, their configurative impact should be investigated and accounted for as a part of the study.

### 3.2.3 Compliance with GDPR

This study has in many ways been affected by the introduction of the General Data Protection Regulations (GDPR). GDPR was introduced by the European Union May 25th, 2018 and is enforced by the Danish Data Protection Agency. GDPR introduce a set of rules and regulations for handling personal data. These rules among other things demand that the data is to be stored secure locations and only people who are registered as data owners or processors can have access to the data. This only applies as long as the data contains personally identifiable data. Personal identifiers can be everything from a social security number to a unique identifier from Facebook.

Therefore, the datasets used in this study has been registered at the Danish Data Protection Agency by the supervisor of this thesis. The owners of the dataset include myself and my supervisor.

GDPR of course has a limiting impact on the realm of possibilities for digital methods in general. For this study in particular the it mainly impacted the data sprint held as part of the PUSH project. During the data sprint, only the data owners were allowed to look at the entire dataset, and from that extract aggregates of data without personal identifiers.

## 3.3 Facebook as an ethnographic field

In this section I ask the question how harvesting data from Facebook through conversations with the Graph API can be perceived ethnographic fieldwork. What might we learn from changing the perception of the API from a purely technical instrument to a gatekeeper and an informant in the form of a socio-technical assembly?

*“When I sat down in front of my computer that evening in the beginning of January I had only one question on my mind: How were we going to harvest five years of information from 72.000 pages in only 26 days? The harvest alone would probably take all the time we had left, and we didn’t even have a script. No, it all had to be built from scratch. Anders had a script that he had written for another project, but it was ineffective and demanded a lot of supervision. What we needed was a robust and dependable script that would be able to supervise itself. I only had very little experience writing syntax for the Graph API, so I immediately started looking for documentation on how it should be structured and tested it through the Explorer. This is not much different that the work I was used to as a mechanical engineer, I thought to myself. No, the problems are purely technical, so the solution must be as well. I just had to conform with the fact, that the more time I spend on the script, the less I would have to do proper techno-anthropological research. My interviews and observations would have to be squeezed in between reading Dewey and Latour.” – Field note*

Labelling the process of harvesting data from the Facebook API as ethnographic fieldwork is probably not self-evident for the reader of this study. And as should be clear from the field note above it wasn’t for me either.

It took about a week to build the first functioning version of the script. During that week, and the following three it took to finalise the harvest the perception of the API radically changed for me. In the end it felt like I had spent 26 days in a virtual World surrounded by a constant blue glare from the computer screen, only communicating through commands of text. And in my efforts to understand and learn how to communicate with the API, I felt like I was gradually being let behind a set of curtains of the Feeds and Walls, the Pages and Posts. Behind those curtains was the “real” Facebook in its flows of data, its social graphs and lists of users, interactions, links and rankings. This was the representation of Facebook I wanted to gain proximity to. Not only to harvest all the flows of data, but also to gain an understanding of how this mediator impacted the representations of social life.

This section can be seen as contemplation on what sorts of knowledge I gained through this unintended and unexpected transformation of the API from a purely technical entity to an informant of this study.

But first I wish to establish a foundation of how ethnography in general handles the digital.

### 3.3.1 Ethnography of the digital

*“After all the data had been harvested through the API, a great task of organizing, counting and triangulation lay ahead of us. All the posts, pages, users, likes, comments and locations had to be placed in the right folders and the right files. After a week or so this work was done, and the next task was to delimit the dataset by the search term “HPV”. It took an hour or so to search through all 25 million posts, outputting in total 3059 posts relating to the controversy. This was when the surprises started pouring down. We knew in beforehand, that we were likely to harvest an amount of data that would make some kind of user profiling possible. But to which degree, we had no idea. A part of the study was to investigate whether the controversy would function as a gateway to new spheres of interest. Our intuition told us that people participating in the HPV controversy would be likely to also be interested in e.g. alternative medicine. We would test this by locating the most active users in the controversy (in terms of numbers likes, comments and posts) and then search the entire geographically curated dataset for all other interactions these users had made. The latter would then be representing “alternative spheres of interest”. As I had done so many times before to validate the method used for counting entries in the dataset, I typed in “facebook.com/” in my browser address bar followed by the unique identifier of the most active user. This would lead me to the user’s profile. What I had not realised, was that this was the first time that I had done this with a user. Until now, I had only been counting posts and pages. All this time the API had functioned as the prime informant and more or less the only object of study and the users as static and anonymous data points. But suddenly and to my great surprise this balance was completely shifted. What had before only been present as an arbitrary identifier of a user, was now transformed into a profile of a living and breathing human being. And combined with the information, that this user was the most active in the HPV controversy sparked a creepy sensation in me. It really felt like doing ethnography.”* Field note ultimo march

Latour argues that researchers undertaking the tiresome and obsessive job it is to assemble the many traces of actors’ social associations should look towards ethnography and anthropology for inspiration (Latour 2005).

In their introduction to *Laboratory Life*, Latour and Woolgar describe how observers of scientific culture must strike a balance between their own partial familiarity with aspects of scientific activity, and the scientists’ versions of laboratory life:

*“In practice, observers steer a middle path between the two extreme roles of total newcomer (an unattainable ideal) and that of complete participant (who in going native is unable usefully to communicate to his community of fellow observers)”.* (Latour and Woolgar 1979, 44).

*Going native* refers to what happens when the observer is emerged too deep in the culture which is observed. One perspective of ethnography in this view is about gaining proximity to capture the *emic* versions of social life and at the same time keeping a distance in order for the describe for the community of fellow observers the *etic* version. (Malinowski 1999; Goodenough 1970).

Gaining proximity traditionally means *being out there*. Doing fieldwork like participant observation, interviews and transcripts in direct engagement with the “natives” are considered hallmarks of ethnographic methods that traditionally entails *being out there* (Boellstorff et al 2012; Spradley 1980).

But how do you do fieldwork out there, when out there is a virtual place?

Ethnographic fieldwork in virtual worlds has been carried out in large numbers in the past couple of decades. For more than two years, anthropologist Tom Boellstorff conducted ethnographic fieldwork in the virtual world of *Second Life*, living among and observing its residents (Boellstorff 2015). Bonnie Nardi did something similar in the virtual *World of Warcraft* (Nardi 2010).

The relationship between offline and online culture has also been investigated as an ethnography. In his book *Social Media in an English Village* Daniel Miller describes how English culture is manifested in the way the residents of the village use social media (Miller 2016).

Common to these two ways of doing ethnography is that they engage with either avatars of offline humans or study the relationship between the offline and online culture.

This leaves digital methods in particular and STS in general in a soft spot when it comes to ethnography, as there are seemingly no humans to observe. The socio-materiel constructions we engage with does not fall under the categories for which methodological precedence has been created in ethnography of digital fields.

In the article *An ethnography by any other name*, ethnographer Michael Agar concludes that the definition of what counts as 'real' ethnography is in fact a more complicated question now than ever (2006, 3). Some argue that real ethnography cannot be carried out outside an academic anthropology department, in less than a year, and with an already existing problem in mind (2006, 4). Others argue that anthropology is about applying what has been taught to a problem of interest opening the definition of 'the field' up for interpretation (Agar 2006, 4).

Agar propose that there are many different types of ethnography, and whether or not it is acceptable comes down to the ethnographer's ability to follow the hallmark ethnographic principles (some of which are listed above). Boellstorff and colleagues argue that the ethnographic methods can easily be attuned to a virtual world. As an example, they describe the domain of "textual listening" as a form of interviewing (Boellstorff et al 2012).

Instead of trying to find the 'real' ethnography, Agar urges to pursue *abduction* (Agar 2006, 11). As opposed to deduction, where new conclusions are drawn from old concepts, and induction, where new materials are fitted into available concepts, abduction is where new concepts enters ethnography. It does so by allowing the ethnographer to be "*lead away*" from the old concepts by turning habits and certainties of the observed into *rich points* for surprising observations.

In order for abduction to fit the dynamic nature of ethnography, it has to be applied *iteratively*. As the research goes on, early abductions may be altered by information gained in new abductions. In the case of this study, such iterative alterations occurred every time the script was changed as a consequence of lessons learned through the previous script.

Lastly, abductions are *recursive* from the Latin for "*run again*" (Agar 2006, 13). This does not mean that ethnography is all about trial and error. It means that whenever a surprise in a rich point is pursued another surprise comes up which now needs to be pursued as well. As such recursion helps the ethnographer understand when a particular rich point has been exhausted and why some abductions are richer than others (Agar 2006, 13).

With basis in Agar, Boellsdorff and Munk this thesis argue that time is ripe to encompass the natively digital mediators as spaces for ethnographic enquiry. While the digital ethnographer has exchanged the recorder, notepad and pen with computers, scripts and external hard drives; the jotting notes with ad hoc learned programming skills; and the street corner with the desktop view of the Facebook platform, the observation enterprise is still based on recursive iterative abductions made from the rich points of the API.

The natively digital does not only include the online profiles of offline people, but also the repurposed tools and methods of the medium. This includes the API with which many conversations have been carried out throughout this study.

At first sight, the API might look like a mere *intermediary* where the output is inferable from the input. But once proximity to the API is obtained it becomes visible that it is a socio-technical construct, a hybrid between a database, a gatekeeper, an interviewee, rules and regulations, and a *mediator* in the Latourian sense as it is continuously assembled by all its constituent parts.

### 3.3.2 Gaining and sustaining rapport

*“Something weird happened last night. The harvester had stopped, no warning signs whatsoever, simply stopped. This is so strange! According to Facebook, we have at least a month to get the data, before shutting down access. And we were actually quite close, with only around 20% of the Events left to be downloaded. The error message informed me that “Access to this data is temporarily disabled for non-active apps or apps that have not recently accessed this data due to changes we are making to the Facebook Platform”. What the hell does that mean? We had become so familiar with how to converse with our informant- the API- our requests had been fine tuned to maximise the yield of each call, and the informant had been kind enough to answer our questions each time. Besides of course for that one time, where we asked to many questions at the same time, leading to our informant ignoring us for 24 hours (which it was kind enough to tell us). But this time it did not tell us anything. We were nowhere near the maximum amount of calls per minute, and the profile, that performed the proxy between us and the informant was still online. As a mean to narrow down possible sources to the problem, I tried to update the token. No result. The token is a key phrase, required to open the door the data. If the key phrase whispered through the small hatch in the door is correct, the informant will place the requested data on the other side of the door in the form of a small json package. But if the key phrase does not correspond with the access level of the required data, the json package will only contain a message from the informant explaining that you have tried to acquire data, which is not yours to acquire. But the message received - access to this data is temporarily disabled” was utterly ambiguous. No matter what token I tried, the message was the same. It gave the impression that the informant we had been speaking with all along, had suddenly become someone else, a hybrid being.” – Field notes*

The field notes above describes some of the final contemplations I made on my conversations with the informant - the Application Programming Interface of Facebook.

*Gaining rapport* is not a concept reserved for exploring and accounting for exotic and inaccessible cultures such as the Balinese or the Argonauts of the Western Pacific (Malinowski 1999; Geertz 1972) or the heroin addicts in Kentucky (Agar 2006). When in contact with a digital informant such as the API, I found that gaining rapport is perhaps even more critical. This is both in respect to gain access to the data safeguarded by the API as well as to study the API's contributions to the configuration of the study of said data.

Authentication is the prime source of rapport. It was easy for me to obtain through the developers' interface, but as made evident in the field note above, it had to be sustained.

What I learned was, that it is crucial to know the rules of how to converse with the API. Just like any other conversation, if you ask too quickly, the other part of the conversation will interrupt you at some point. For example, rules to abide to sustain rapport that I stumbled upon encompass the maximum number of sub questions a question can contain as well as the frequency with which the questions can be asked. The authentication process occurs anew each time a question is asked. If the rules are broken, the punishment range from a written warning (Figure 14) to silent treatment for an unknown duration of time (Figure 15).

*Figure 14 - This snapshot shows an error message from Facebook when trying to access an object that is not available.*

```
{
  "error": {
    "message": "(#100) Tried accessing nonexistent field (gf) on node type (User)",
    "type": "OAuthException",
    "code": 100,
    "fbtrace_id": "AjMEQHVVAAQO"
  }
}
```

Figure 15 - This is a snapshot of the message from Facebook to the inbox of the app used to harvest data from the API. It informs that the hourly rate limit has been reached and that access can be obtained again later

## Rate Limit Reached

attention

**TANTLAB** has reached **100%** of the hourly rate limit. All API calls to your app will fail until your app falls below the throttling limit.

Here's a breakdown of the methods causing your app to hit the rate limit:

### By Call Count:

GET {page-post-id}/comments 33005%

GET {page-post-id}/reactions 6443%

GET {comment-id}/reactions 5018%

### By Total Time:

GET {page-post-id}/comments 10430%

GET {page-post-id}/reactions 8221%

GET {comment-id}/reactions 4895%

### By Total CPU Time:

GET {page-post-id}/comments 38550%

GET {page-post-id}/reactions 7948%

GET {comment-id}/reactions 3724%

The situations where the conversation breaks down and rapport is lost are the *rich points* where surprising abductions can be made. A lot of information about the socio-technical construct of the informant can be gathered by observing the behaviour of the API. The 'socio' in socio-technical is foregrounded so to speak. When I broke the embedded and static rules of the API - intentionally or not - those rules became explicitly enforced. I could then choose to adapt this new knowledge iteratively in further enquiry, or not. In the example presented in Figure 15, a simple pause not longer than the blink of an eye was enough to remedy the problem and return to a state of mutual acceptance.

Rapport can also be lost because the rules changes as described in the field note above. A few days before the abrupt loss of rapport, news had come out that the analytical company Cambridge Analytica had gained access to private Facebook information on up to as many as 87 million users. The data had been retrieved through the Facebook API by the company Global Science Research back in 2014 (Braga 2018). Their Facebook-app called *thisisyourdigitallife*, had supposedly made psychological profiles of the user based on all sorts of interactions such as likes, shares and comments on posts, as well as information on religion, sexual orientation and location. The problem was seemingly that the problem was that the data had been shared with Cambridge Analytica, who would later use the data for

profiling and micro-targeting potential voters in the 2016 U.S. presidential election in favour of Donald Trump (Brage 2018).

At that point my communication with the API had become trivial and a practical matter of harvesting data. My computer would automatically ask questions to and receive answers from the informant. All obstacles that could break rapport had been removed in my efforts to learn how to converse with the informant. But suddenly the conversation breaks down. The informant returns to an explicit hybrid state, and the actions of an executive decision in Facebook manifests itself through the functions of the API. The gate was shut.

This episode shows some of the challenges with complete participation. When deeply emerged in the practicalities of programming, I lost the ability to become surprised and intrigued by the routine tasks because they become tacit knowledge. But when rapport was lost, I was cast 'back' to a state of a novice and distant participant. As such, both gaining and losing rapport might be a great irritation for the practical task at hand, but also constitute some of the most valuable rich points themselves because also help navigate the distance to the object of study that enables iterative and recursive abduction.

### 3.3.3 Tools as “armchair harvesting”?

*This section examines the benefits of constructing new tools for conversing with API's as opposed to using existing.*

Tools for extracting data through the Facebook API already exist. By using such tools as Netvizz (Rieder 2013), the access to the API is mediated by another program allowing the observer to feed off of the rapport already established between the program and the API.

To feed off of the rapport of other programs, bears resemblance to what has been called *armchair anthropology*. Armchair anthropology refers to the practice of gathering information from a field by proxy of either skilled observers, such as ethnographers, or unskilled observers, such as merchants, seamen and other travellers. The practice was especially popular amongst British anthropologists in the 19th century as they otherwise would have to travel for long durations in order to find and reach their natives (Malinowski 1999). Fast forwarding to present time and the study of digital natives through the API, physical distance is no longer an issue. Instead it has been replaced by another sort of distance, in the sense that the ethnographer either has to climb the steep learning curves of programming in order to reach to the digital natives or apply tools such as Netvizz climb them in his or her stead.

Armchair practices might at first seem convenient, as the proxy might be in a better position to negotiate terms and conditions. A critique raised by Bronislaw Malinowski against the armchair practices is about the nature of the accounts brought back from the field. By leaving the fieldwork to others, the armchair researcher risk to be presented with *etic* representations as though they were *emic* representations (Malinowski 1999).

*“It would be easy to quote works of high repute, and with a scientific hall-mark on them, in which wholesale generalisations are laid down before us, and we are not informed at all by what actual experiences the writers have reached their conclusion.”* (Malinowski 1999, 13)

Fast forwarding to present time again, the critique raised by Malinowski also applies to what could be called *armchair harvesting*. Choices diverging from the ontological order of the medium from which the data is extracted, has already been taken in many folds by the proxy. And those choices are in a tool such as Netvizz obscured by the hybrid that is the API in combination with yet another socio-technical black box in form of the tool. As such, the empirical study would not be of the social World as ordered by the medium, but as ordered by a black boxed hybrid constantly being reconfigured by its constituent parts.

In this study, the observer chose to climb the learning curves of programming and cover the distance to the natively digital himself.

The critique raised above could obviously also be directed at the program written by the observer to automatically question the informant. But the major difference is that by directly engaging with the informant, by learning its language, studying its every move and by gaining rapport, the observer keeps his proximity and so to speak puts up his tent in the middle of the village. It might seem like splitting hairs or as being a moot point but is in fact quite the opposite. By not allowing the study to be based on the rapport of others, but instead go the long way around by learning to program, all choices are left open for the observer and the study can be attuned to the ontological orders of the medium - instead of a mediated version of that.

Furthermore, the observer misses out on a potentially great opportunity to learn about and from the API. Both in relation to learn how to gain and sustain rapport and what the breakdowns of rapport brings along of opportunities as argued in the previous chapter. But also, in relation to understand how the social is actually represented and order by the

medium. Having used Netvizz it would for instance not have been possible to learn how the public/private divide changes from the browser representation to that of the API through tags.

The two small illustrations below exemplify the critique of armchair harvesting vividly. Both illustrations are networks of a Facebook post connected with users who engage with the post. The link between the post and the user represent either a comment or a reaction. The first network (Figure 16) is based on data collected through the Netvizz tool, and the second based on the tool of the observer (Figure 17).

*Figure 16 - This figure shows a network based on data collected from Netvizz. Note that the links between the users and the post does not have any information*

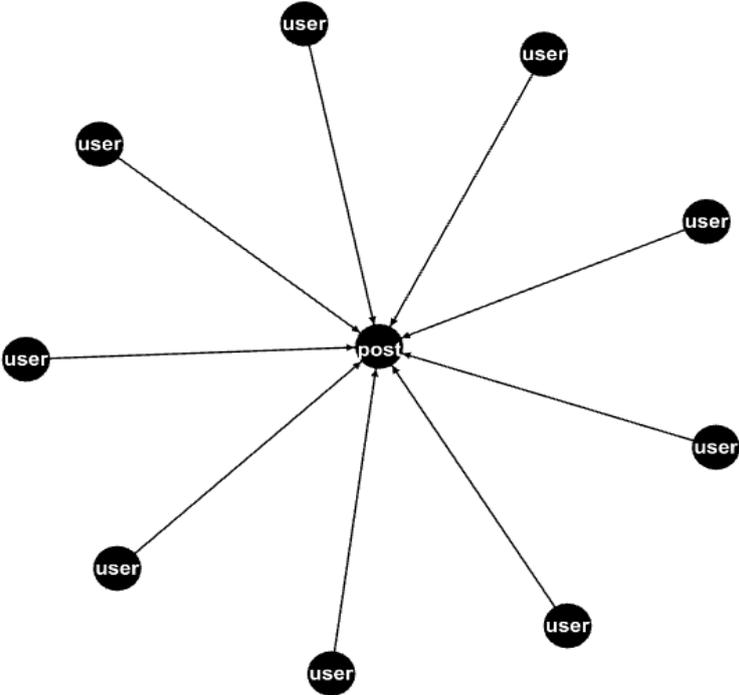
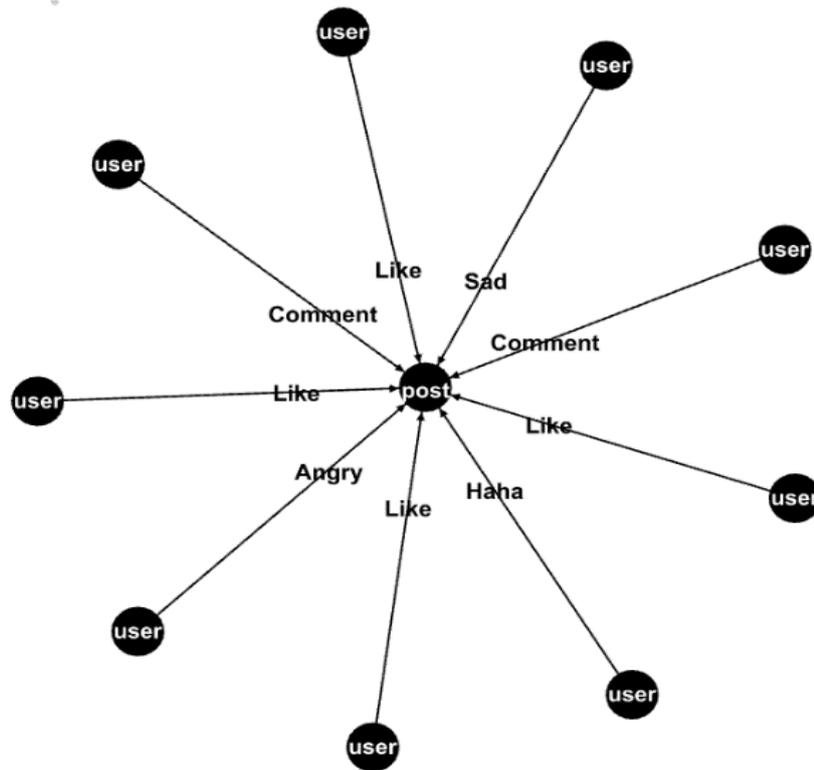


Figure 17 - This figure shows a network based on data collected through the script written for this study. Note that the links between the users and the post contains information about the type of link



What should be evident is that the networks are completely identical in form. What separates them are the labels on the edges between the users and the post. I.e. the comments and reactions of the users on each post. When using Netvizz, it is not possible to qualify what type of relation the users have to the post, as a consequence of a fixed order embedded in the tool, even though the information is available through the API. This qualification is only possible via the program written by the observer.

## 4 Analysis of the empirical comparison

The two networks assembled when following the issue centric and the medium centric approaches, respectively, can be seen as two different representations of the monads linked to the HPV controversy. In the previous chapter, the methodological differences of the two approaches the consequences for how monads are assembled were discussed.

This chapter concerns the empirical assessment and comparison of the two approaches in order to answer the problem statement of this study, namely: *How does the addition of*

*monadic entries differ between issue centric and medium centric approaches to digital methods? And what are the implications for social theory?*

The chapter is divided in two sections, each contributing different angles to shed light on these questions. The first takes a panoramic point of view of the networks in order to scrutinize and compare structures and groupings. As described earlier, the panoramic point of view shows the entire monadic structure as a network, but it does not allow inspection of the monad's item list. The second show how the monads are interconnected and represented differently by the two networks from an oligoptic point of view. In the last part of each of the sections, the empirical implications using each of the two approaches are explicitly analysed, compared and summarised.

The second questions concerning the implications of the two approaches for social theory will mainly be handled in the discussion and conclusion.

## 4.1 The panoramic comparison

As established earlier, the input for both networks were 3,059 Facebook posts concerning the HPV controversy. Both networks are monopartite, meaning that they consist of only one type of nodes, in this case posts from Facebook. In short, both are post-to-post networks.

The real issue is how we establish connections (edges) between the nodes. This is where the two approaches differ. Following the issue centric approach, posts are connected based on the co-occurrence of issue related topics. Following the medium centric approach, posts are connected based on the interactions of the users that engage with them.

### 4.1.1 The medium centric network

The result of the medium centric approach is a network consisting of 2,174 posts linked by 181,738 shared issue terms<sup>9</sup>. The network is shown in Figure 18 below. Each monad is a post from the HPV dataset, linked to other monads by users' interactions.

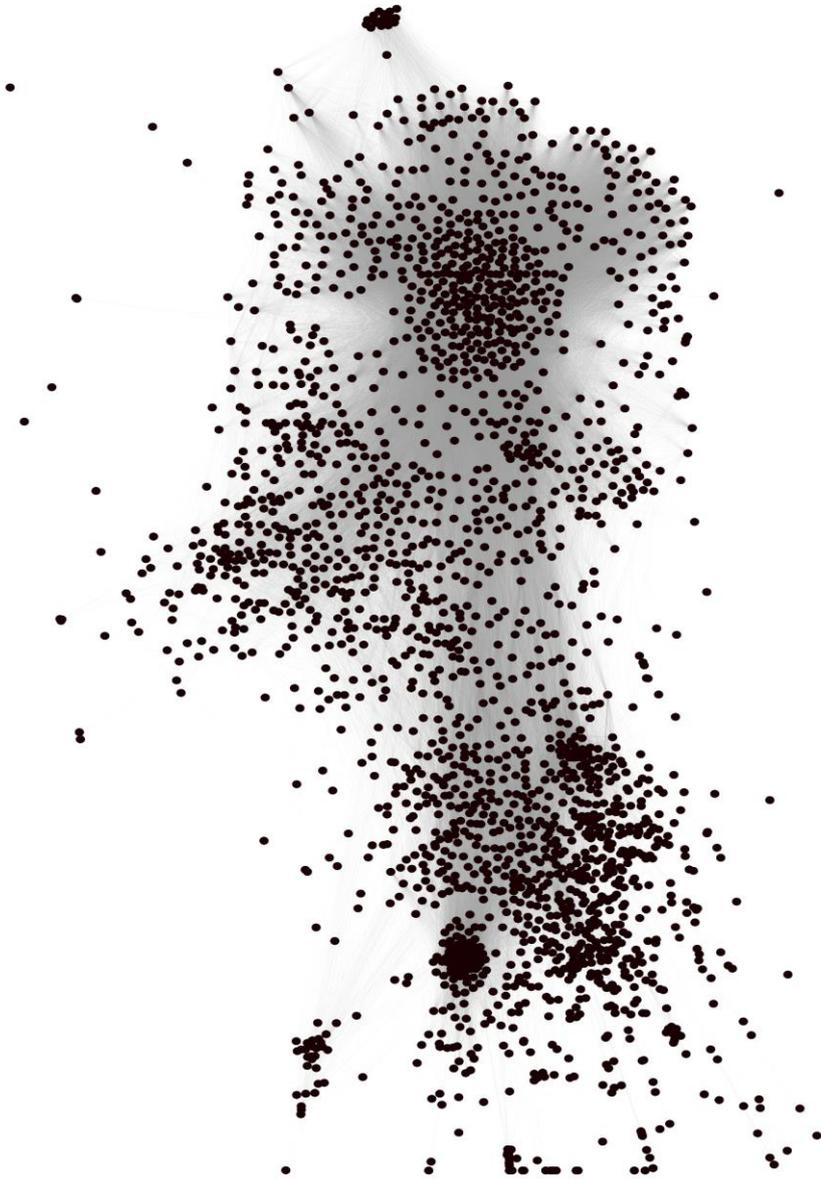
The entity list of the average monad consists of approximately 84 other monads. This is a somewhat arbitrary and bland measure by itself, as there are no rules of thumb to help assess whether it is many or few. It all comes down to how the network is structured in clusters and

---

<sup>9</sup> Although the HPV dataset consists of 3,059 posts, only those that relate to other posts through user interactions are part of the network.

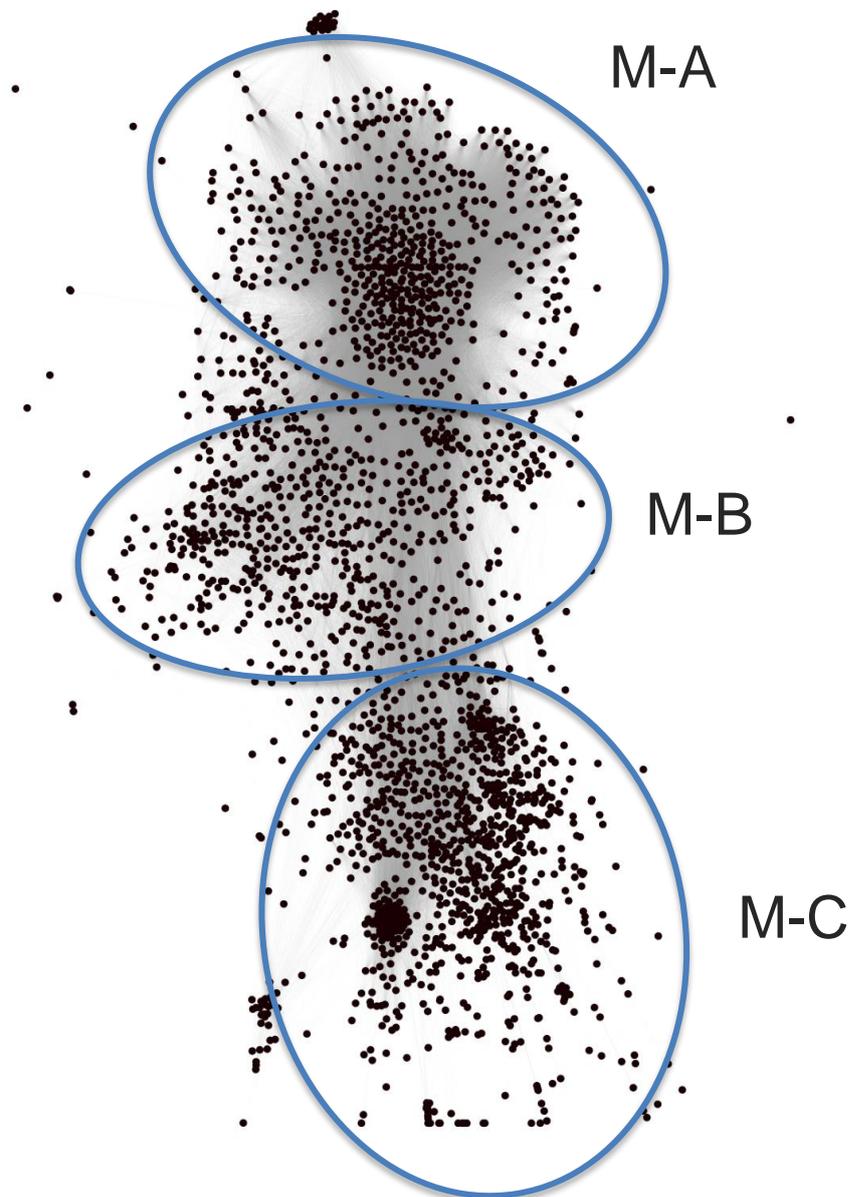
groups. But comparing the measure with the equivalent from the issue centric network, gives an idea of the relative structural differences, as the two networks originate from the same dataset.

Figure 18 - This graph illustrates the medium centric network.



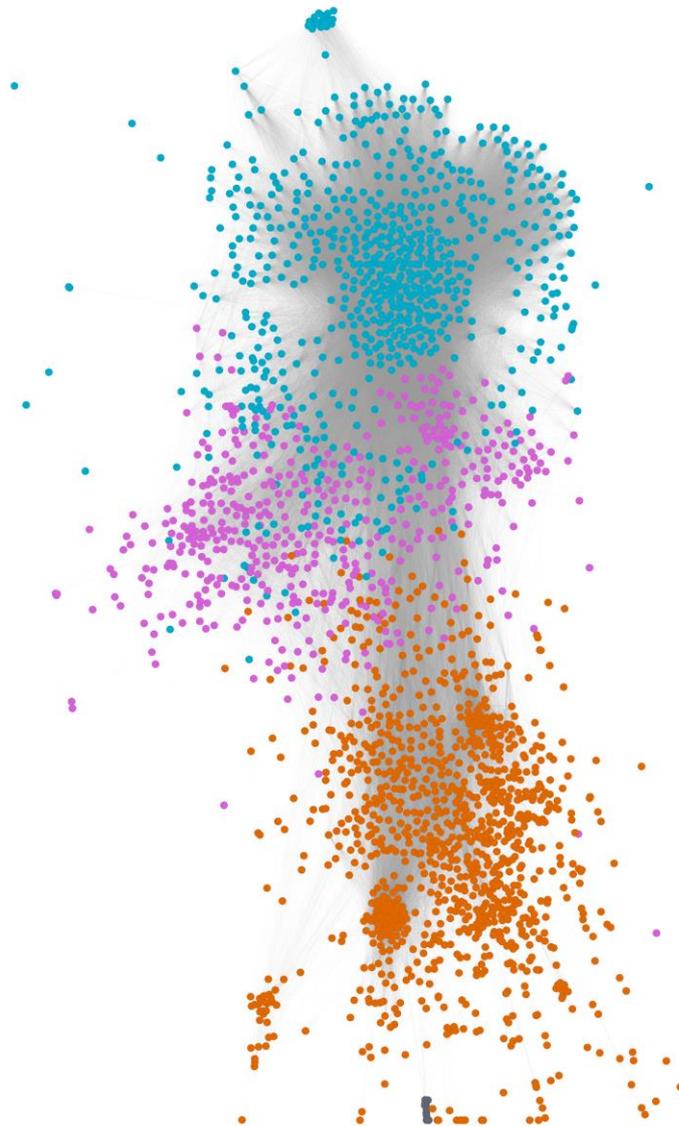
Visual inspection of the network identifies three main clusters (Figure 19). For future ease, they will be called M-A, M-B and M-C (labelled 'M' for their association to the medium centric network). The clusters are more or less the same size. Within each cluster several smaller clusters can be identified. From visual inspection, cluster M-C overall seems denser than M-A and M-B. This could mean that there is a relatively larger degree of user overlap in this cluster.

Figure 19 - This is the same illustration as figure 17. The circles represent the three clusters identified.



The modularity algorithm that outputs cluster classes based on calculations of each node's connection to all the other nodes in the network, identifies the same division (Figure 20). It also adds additional information to the analysis, as it questions whether M-B is a cluster in itself, or if it is a mix between M-A and M-C. This is visible as all the three node colours are present in cluster M-B (Figure 20).

Figure 20 - This graph illustrates the medium centric network where the nodes are coloured by the modularity score to identify clusters

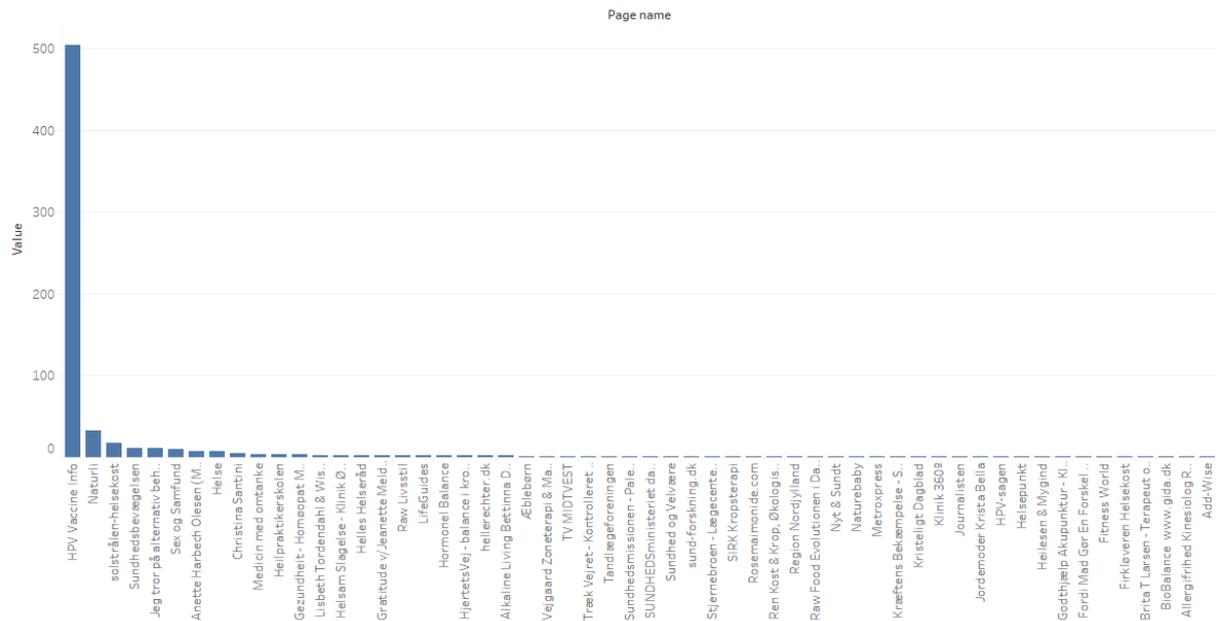


Based on the preliminary structural analysis, the medium centric approach points to a conclusion that the monads assembled by Facebook interactions with the HPV controversy consist of three main clusters. The following sections will dive into the content of each of the clusters to investigate whether the structural division extends to a thematic division.

#### 4.1.1.1 Interpreting cluster M-A

Cluster M-A consist of 678 posts authored by 56 pages. The most active page is called *HPV Vaccine Info* followed by a range of pages significantly less active (Figure 21).

Figure 21 - This diagram shows the number of posts from the 50 most active pages in cluster M-A



Looking at the content of the posts in cluster M-A, it can best be characterised as the conception of a Deweyan public. For the members of this public, the HPV controversy revolves around the consequences of introducing the vaccine. They claim that the vaccine cause side effects and that other alternatives are better suited. They constitute a public, because these consequences are not taken care of by the public institutions responsible for introducing the vaccine. They express a large degree of distrust in the public institutions such as the National Board of Health, experts such as doctors, professors and public health researchers, and NGOs that are pro HPV-vaccine such as The Danish Cancer Society. Their stories are generally driven by personal stories and evident by persons affected by side effects (example of posts is presented in Figure 22).

Figure 22 - This figure show two posts representing the public against the HPV-Vaccine. The post on the left tells a story about how the Danish Cancer Society is funded by the medical industry (HPV Vaccine Info 2018). The post on the right tells a story about a medical doctor warning against the HPV-Vaccine (Naturli 2018).

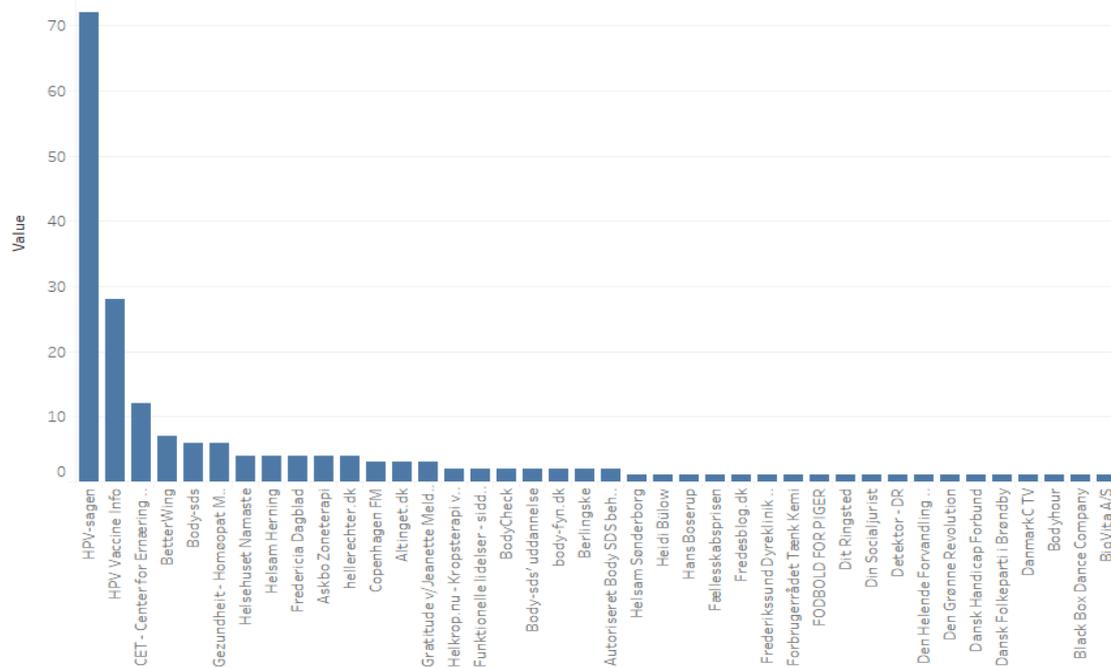


The cluster will at large be interpreted as a public movement sceptic towards the HPV vaccine and active in circulating alternatives both to the vaccine and to cure the side effects.

#### 4.1.1.2 Interpreting cluster M-B

Cluster M-B consist of 463 authored by 91 pages. The pages that contribute the most are listed in Figure 23 and is as cluster M-A mainly constituted by what is characterised as *Public Movement* pages. It also includes a range of other sources including media outlets such as Radio24syv, Politiken and TV2 News, and 'alternative' practitioners of Body SDS and zone therapy.

Figure 23 - This diagram shows the number of posts from the 50 most active pages in cluster M-B



The content is mainly driven by a efforts to define what is right and wrong in the controversy not only by the use of personal accounts as in Cluster A, but also by scientific research. The expert community is being portrayed as divided, as practitioners critical towards the vaccine are amongst the most cited in this cluster.

As such, this cluster is interpreted to be representing a public in allegiance with both parts of the expert community and the established press.

#### 4.1.1.3 Interpreting cluster M-C

Cluster M-C consist of 1,023 authored by 400 pages. It is mainly constituted by NGO's such as Sex and Samfund, public health institutions and the media (Figure 24). For the two former focus is mainly on pro HPV-Vaccine content, addressing the severity of the consequences of HPV related cancers and the drop in participation in the vaccine program (Figure 25). For the latter it is possible to trace a sort of introspective process of the media and their role in the controversy (Figure 26).

Figure 24 - This diagram shows the number of posts from the 50 most active pages in cluster M-C

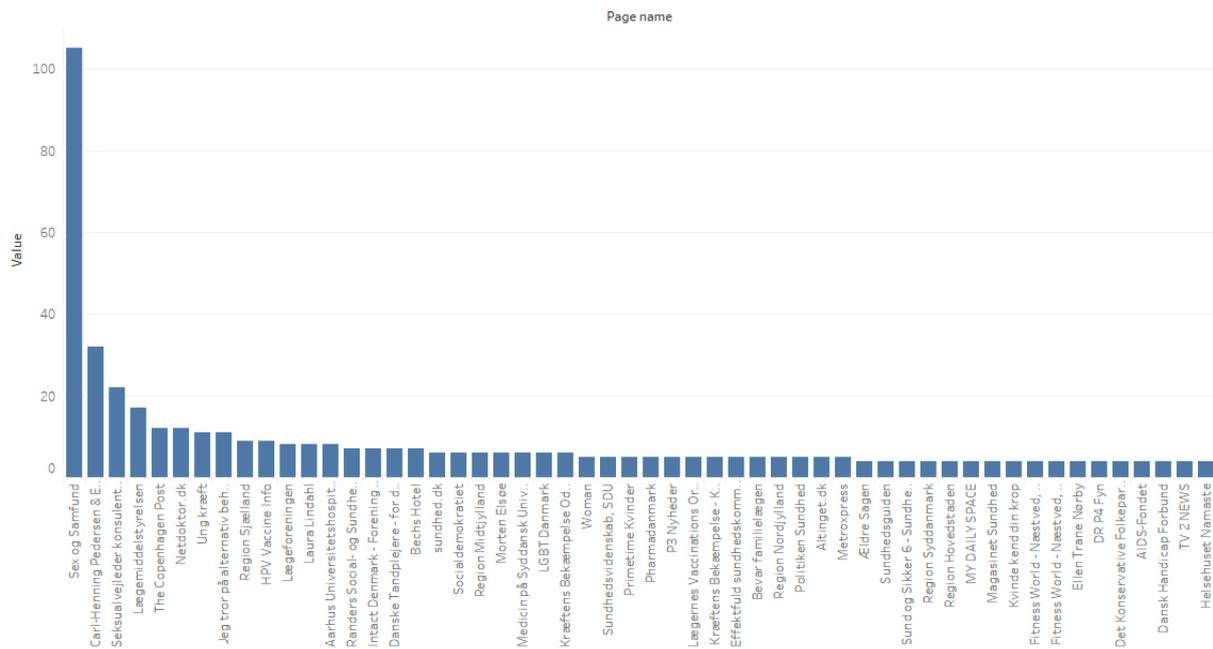


Figure 25 - This figure show two pro HPV-vaccine posts from cluster M-C. The one to the left tell the story about a woman who has suffered from cervical cancer (Lægemiddelstyrelsen 2018) and the one to the right tells the story about how the vaccine percentage has risen from 28 to 67 in only four months (Kræftens Bekæmpelse 2018)

 **Lægemiddelstyrelsen** shared a video. 24 July 2017 · 🌐

Det kræver mod, råstyrke og fightergen at gennemgå et kræftsygdomsforløb. Det ændrer ens liv - for altid. Det kræver endnu mere overskud at stå frem og dele ens historie, men det er enormt vigtigt, da det kan hjælpe andre. Derfor vil vi gerne sige tak til Louise.

HPV-vaccinen redder liv. Du kan finde mere information på [www.stophpv.dk](http://www.stophpv.dk)



 **Kræftens Bekæmpelse - Middelfart** 23 April 2013 · Middelfart · 🌐

Stor ros til de unge kvinder i Middelfart Kommune - ved årsskiftet var 67 procent af de unge kvinder i Middelfart Kommune vaccineret mod livmoderhalskræft. Det er fantastisk flot, at andelen af vaccinerede er steget fra 28 til 67 procent på kun fire måneder. De unge kvinder fra årgang 85-92 har altså i vid udstrækning fulgt opfordringerne om at få den gratis HPV-vaccine.

Men vi mangler stadig at få de sidste kvinder med, for at vaccinationsprocenten kommer helt i top, siger Irma Greve fra Kræftens Bekæmpelses Lokalfdeling i Middelfart. Jeg vil derfor opfordre til, at alle kvinder årgang 85-92 kommer af sted og får de tre stik, som vaccinen består af. Det tager et halvt år at blive vaccineret, og derfor skal I sørge for at starte inden sommerferien, så I kan nå at få den gratis vaccination i 2013, for tilbuddet udløber ved årsskiftet.

Er du allerede er vaccineret, så spørg dine veninder, om de også er kommet i gang. For husk, at kombinationen af vaccination og regelmæssig deltagelse i screening betyder, at vi i fremtiden kan undgå livmoderhalskræft", slutter Irma Greve.

Figure 26 - This figure show two posts about the news media's role in the HPV controversy. On the left is a post from the news media Zetland, reporting that misinformation cause mistrust in expert institutions (Zetland 2018). On the right is a post about how the critical media coverage has frightened the public (Journalisterne 2018).



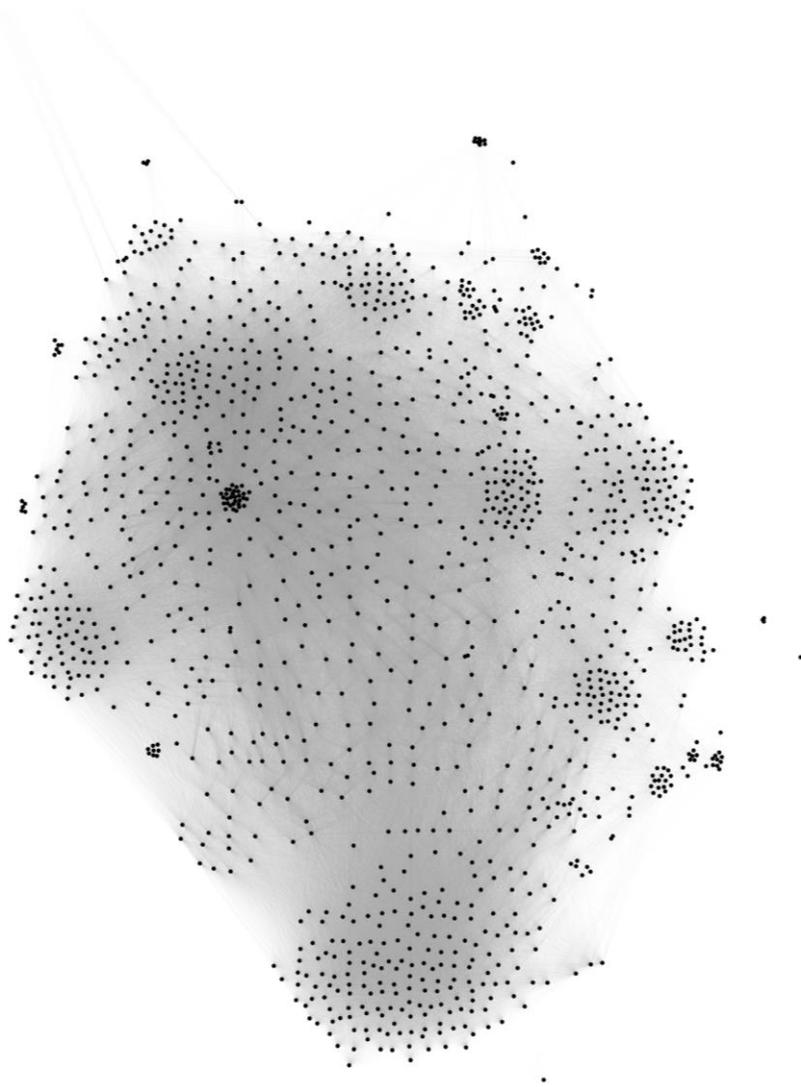
Cluster M-C is interpreted as representing the public institutions. They are pro HPV and trying to impact the debate with facts and turn the problematization from side effects and instead focus on the consequences of HPV in general.

Obviously, the interpretations done on the three clusters are not exhaustive. A lot more could be put in focus and explored e.g. the sub-clusters within the main clusters. But such analysis does not serve the purpose of this study as the main aim is to compare how the posts are distributed in the issue centric and the medium centric approaches respectively. By turning to the issue centric approach below, the comparison of the two approaches from a panoramic view will become more visible.

#### 4.1.2 The issue centric network

As mentioned earlier, the issue centric network is assembled co-occurring keywords in posts. If a keyword is present in two posts, they will be connected in the network. As such, the network is composed of 1,358 posts and 204,412 co-occurrences of keywords, which generates an average of 150 connections per monad. This indicates that the issue centric network is much better internally connected compared to the medium centric (with an average of 84 in the medium centric). As is visible in Figure 27, the division of the network in clusters is not as self-evident as the medium centric network.

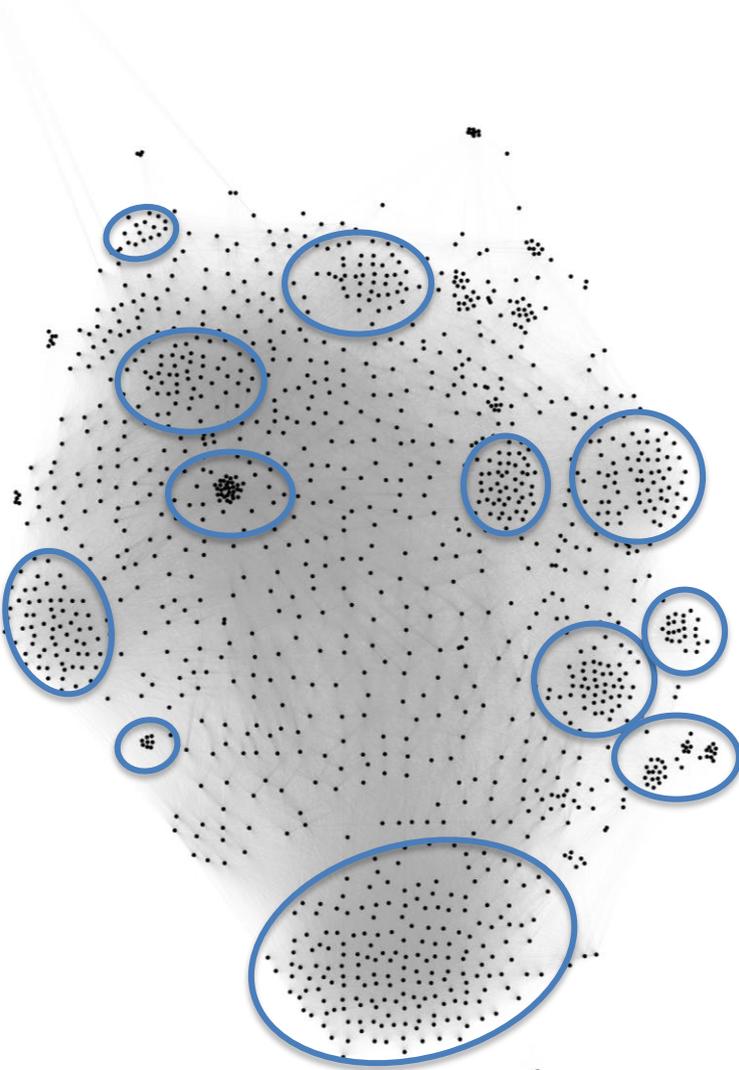
Figure 27 - This graph shows the issue centric network.



Although the issue dictionary for this study has been iterated upon several times by simple algorithms and issue experts, it becomes evident that the method is still not entirely attuned to Danish language. This is evident because the network is very compact and very interconnected. This means that the keywords used in the issue dictionary are too generic and therefore present in too large a part of the posts. They are as such not sufficiently issue specific. This study as such makes explicit a need to further develop the methods and perhaps tools for composing issue dictionaries. For now, this study will rely on the clusters calculated by the modularity algorithm, and the problem with the issue dictionary will be discussed in detail in the chapter *Revisiting the issue dictionary*.

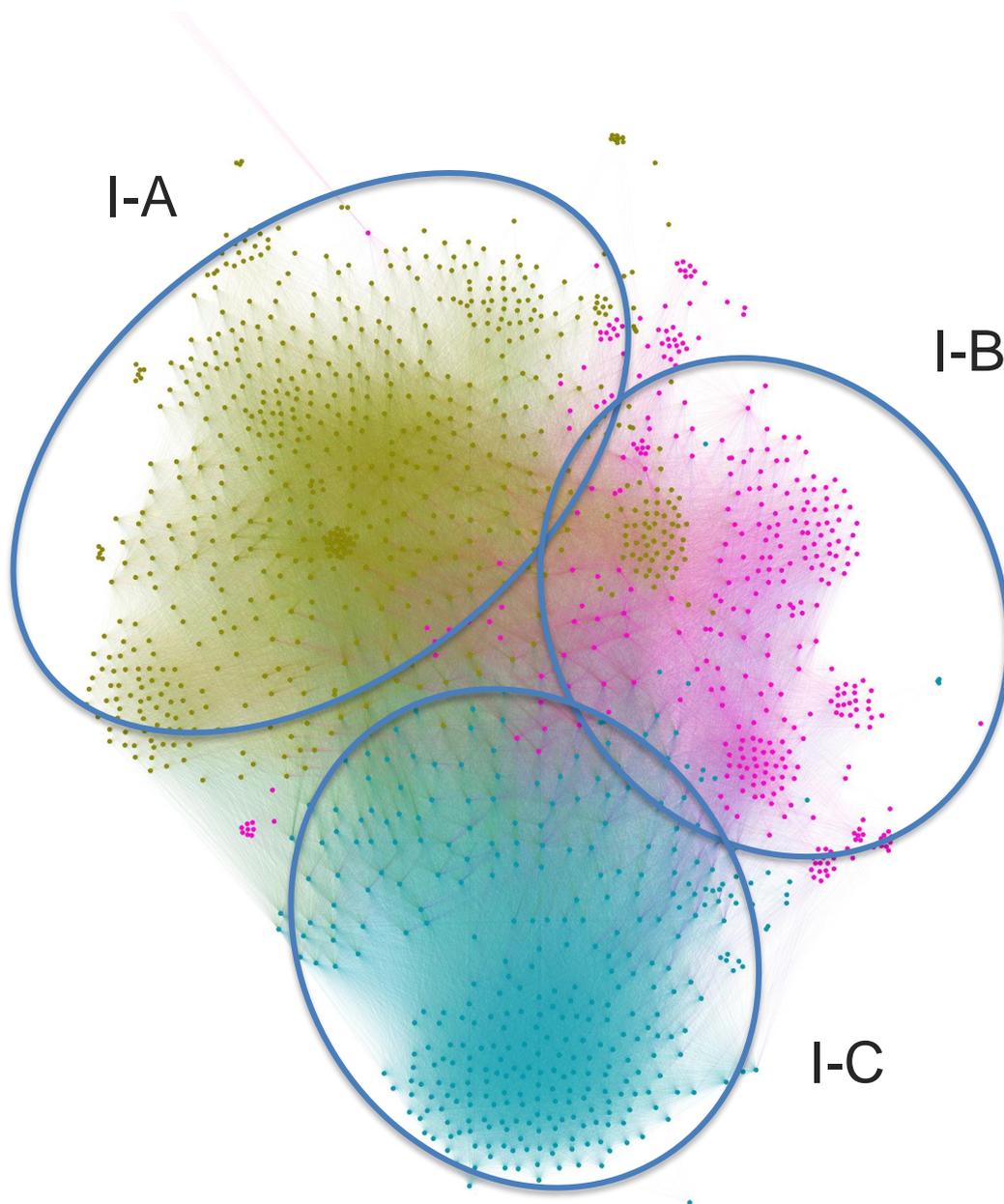
Although difficult to identify clear clusters in the network, smaller ones are identifiable (Figure 28).

Figure 28 - This graph shows the small clusters in the issue centric network



Applying the modularity algorithm on this approach also, these smaller clusters can somewhat be seen as the centres of the three larger clusters I-A, I-B and I-C, which is identified by the modularity algorithm (Figure 29).

Figure 29 - This graph shows how the modularity calculations divides the network into three clusters identifiable by the different colours of the nodes

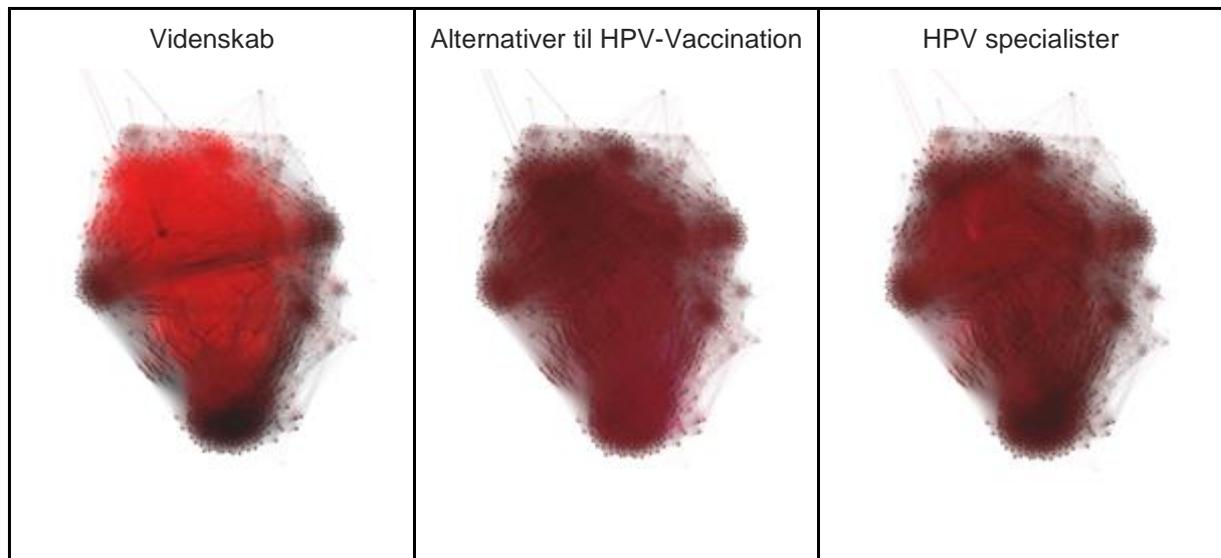


As a consequence of the way the issue centric network has been constructed, each node constitutes a Facebook post and each post has been tagged with topics according to the mentions of specific keywords (*The issue dictionary as a way to operationalise topical networks*).

Having this information, enables a content analysis of the clusters based on a coding of topics. Such information is not natively available for the medium centric approach on Facebook.

Figure 30 shows three different versions of the network that has been coloured based on the concentration of posts tagged with specific topics. Red nodes indicate posts that has been tagged, and black posts that has not.

Figure 30 - The visualisations in the figure shows three different themes that spread out over the network.



General for these three topics “Videnskab”, “Alternativer til HPV-Vaccination” and “HPV specialister” is, that they do not clearly relate to a single cluster but spread over most of the network. Furthermore, they all revolve around a vocabulary related to a scientific realm as they encompass keywords such as “Science”, “Research”, “Screening” as well as a list of names of experts, doctors and researchers.

#### 4.1.2.1 Interpreting cluster I-A

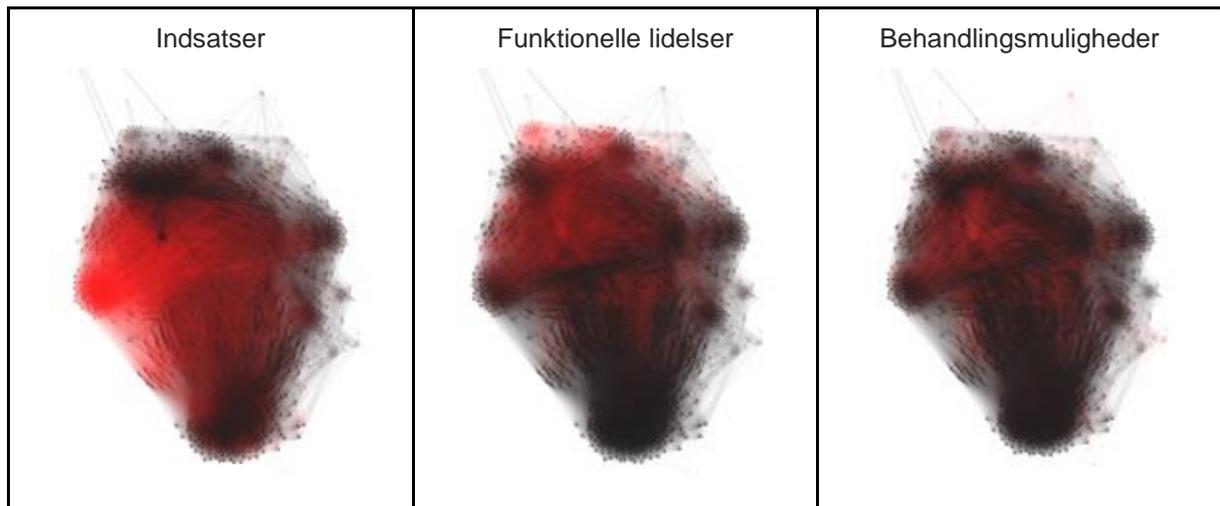
Other topics has a more distinct connection to a specific cluster. Cluster I-A has the highest concentration of the topics “Indsatser”<sup>10</sup>, “Funktionelle lidelser”<sup>11</sup> and “Behandlingsmuligheder”<sup>12</sup>, which can be said to cover ways to treat HPV related side-effects (Figure 31). General to this cluster is scepticism towards public institutions and can best be compared to the *Public movement* cluster (M-A) of the medium network.

<sup>10</sup> Efforts

<sup>11</sup> Translated here as Bodily Distress Syndrome

<sup>12</sup> Treatment possibilities

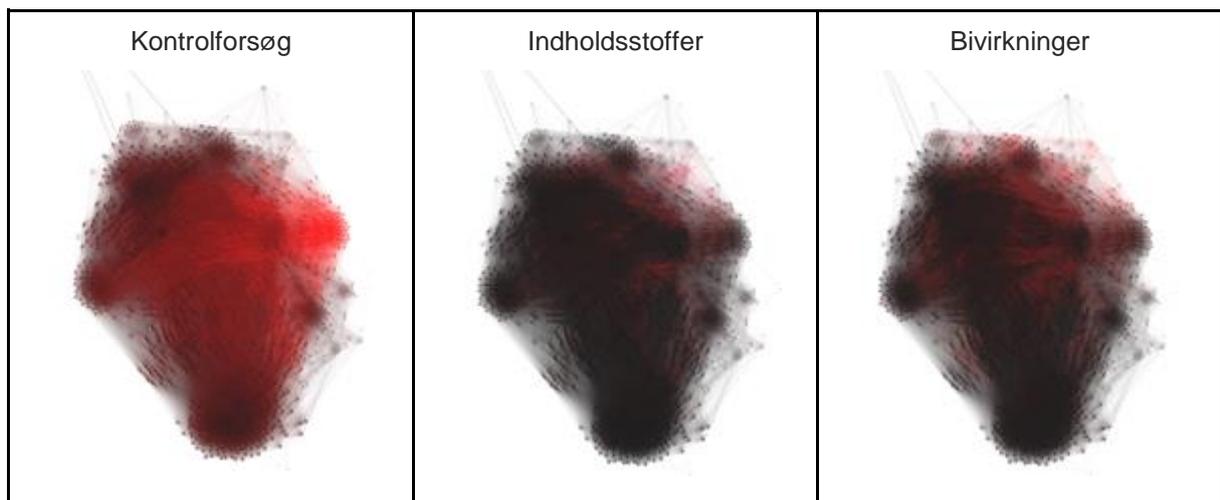
Figure 31 - The visualisations show three different topics that mainly attain to cluster I-A



#### 4.1.2.2 Interpreting cluster I-B

Cluster I-B has the highest concentration of the topics “Kontrollforsøg”<sup>13</sup>, “Indholdsstoffer”<sup>14</sup> and “Bivirkninger”<sup>15</sup>, which can be said to cover how the composition of vaccine lead to side effects (Figure 32). General for this cluster is a negative and sceptic attitude towards the commercial interests in the HPV-controversy and to public institutions in general.

Figure 32 - The visualisations show three different topics that mainly attain to cluster I-B



<sup>13</sup> Experiments controlling for side effects of the vaccine

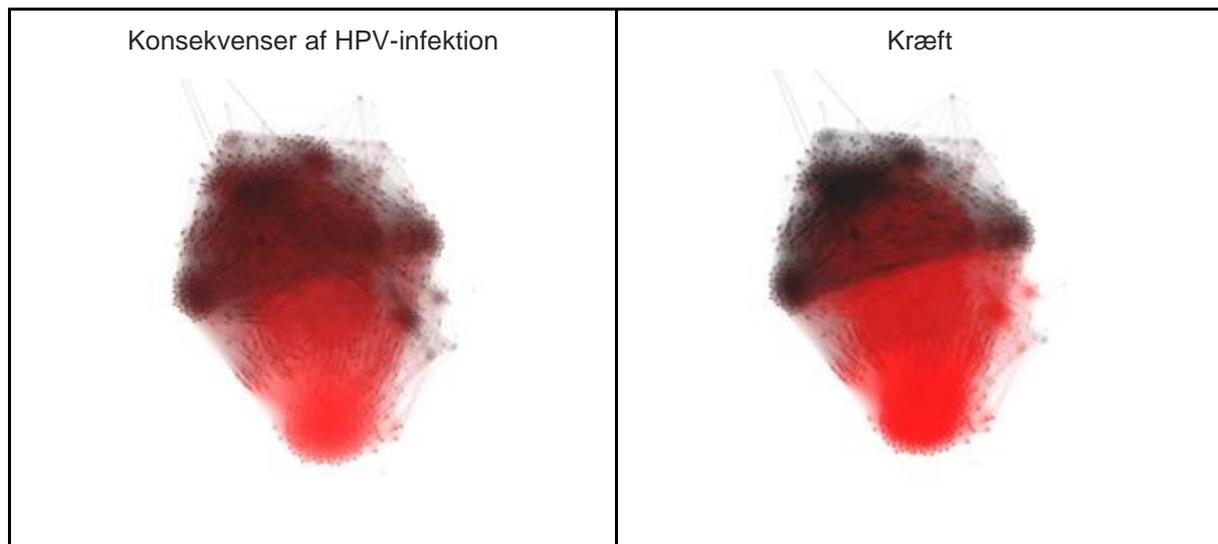
<sup>14</sup> Ingredients (in the vaccine)

<sup>15</sup> Side effects from the vaccine

#### 4.1.2.3 Interpreting cluster I-C

Cluster I-C show the clearest connection between the topics. It has the highest concentration of the topics “Konsekvenser af HPV-infektion”<sup>16</sup> and “Kræft”<sup>17</sup> and covers stories about the consequences of a HPV-infection such as cancer (Figure 33). The cluster does not focus much on side effects and is in general positive towards the vaccination program. This cluster can best be compared to the *Public Institutions* cluster of the medium centric network (M-C).

Figure 33 - The visualisations show three different topics that mainly attain to cluster I-C



#### 4.1.3 Analysing the panoramic similarities and differences

*In the previous sections the networks were introduced and analysed severally, leading to a division of each of the networks into three clusters based on the structure of the network as well as the topics of the posts. This section will focus on a panoramic comparison between the two networks, using the clusters as points of reference.*

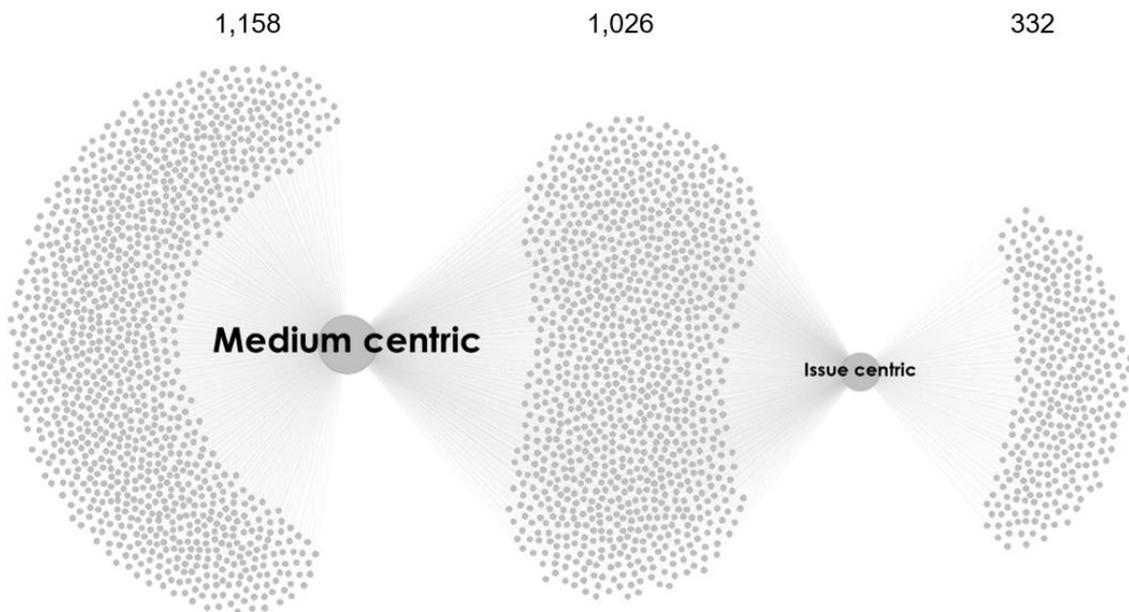
This section starts out with analysing what subsets of the dataset the issue centric and medium centric approaches view as entitled to be part of a Deweyan HPV public.

This is done by combining the two networks in order to generate a new network. The new network consisting of 2,516 unique posts when doublets has been removed (Figure 34). 332 posts are exclusively included in the issue centric network, whereas 1,158 are exclusively included in the medium centric. This leaves 1,026 posts that are shared between the two networks.

<sup>16</sup>Consequences of HPV-infection

<sup>17</sup>Cancer

Figure 34 - This graph illustrates the overlap between the two networks representing the issue centric and medium centric approaches respectively. The numbers refer to the amount of monads in each group.



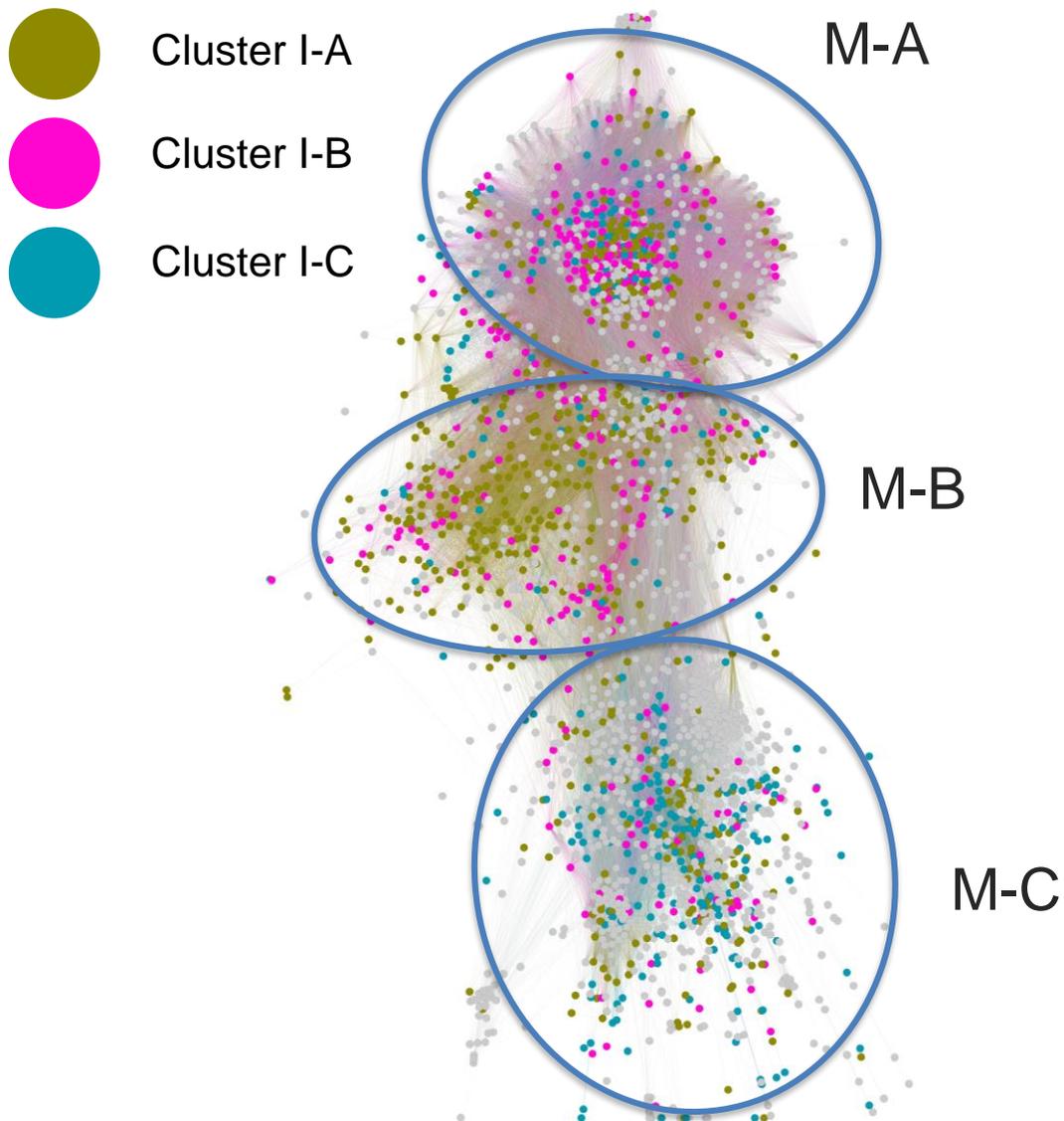
The fact that the two networks are assembled by significantly different subsets of the dataset is interesting. But what subsets are included and excluded?

Having looked at the broad comparison between the two networks, the following will look into how the clusters of the two networks compare to one another. As shown in the previous section, some degree of thematic overlap between the different clusters of the two networks is hinted. For example, the medium centric cluster labeled Public Institutions (M-C) overlaps with thematics and topics found in the issue centric cluster about the consequences of the HPV-infection (I-C). But does this thematic overlap shine through in a structural comparison? If so, the two similar clusters would be located in the same area, if information is carried from one network to the other.

#### 4.1.3.1 Overlaying the issue centric clusters on the medium centric network

As is visible from Figure 35, the issue centric clusters do not clearly resemble those of the medium centric. That being said, the highest concentration of nodes from issue cluster I-A can be found in medium cluster M-B. The same goes for issue cluster I-B which is most often found in medium cluster M-A and lastly for issue cluster I-C which is most often found in medium cluster M-C. The visual assessment might be complicated because the number of shared nodes between the two networks is only half of the total size of the medium centric network.

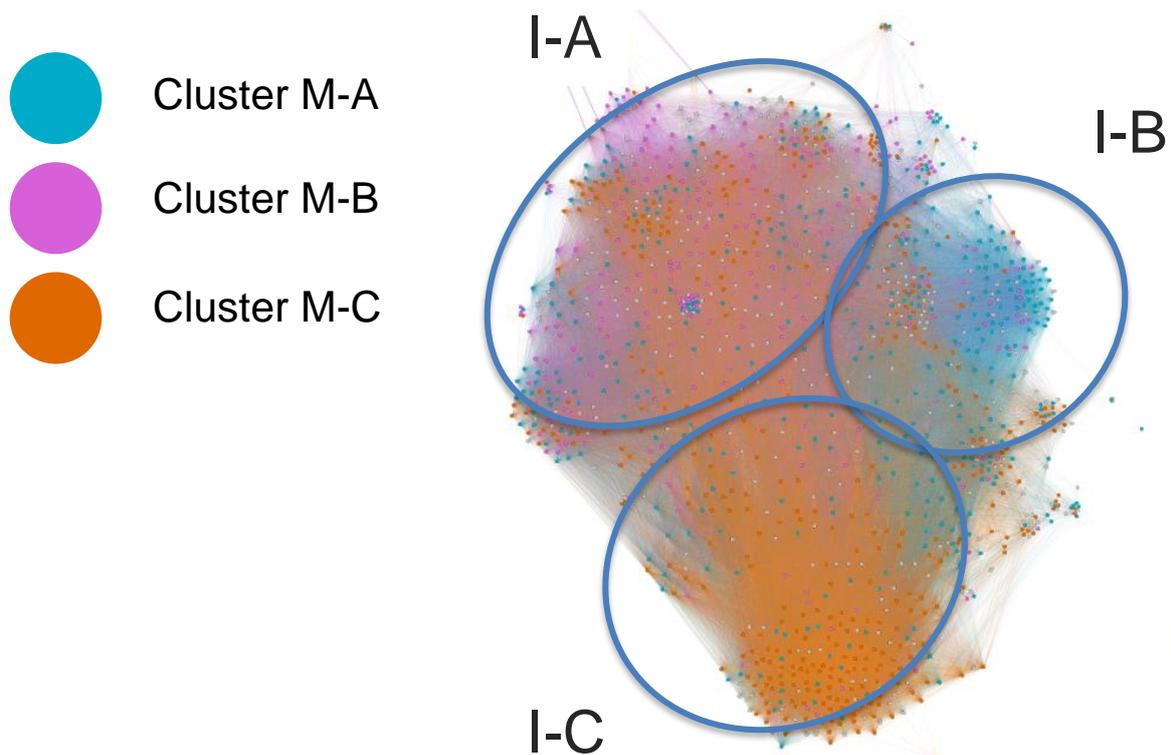
Figure 35 - This graph illustrates the medium centric network, where nodes has been coloured in accordance with their association to the three clusters in the issue centric network. Green represent cluster I-A, pink cluster I-B and blue cluster I-C



#### 4.1.3.2 Overlaying the medium centric clusters on the issue centric network

Examining the overlay the other way around gives a slightly clearer picture, although the clusters are still intertwined. As visible in Figure 36, the distribution is more or less identical compared to the overlay made above. The medium cluster M-A is mostly situated in issue cluster I-B, M-B mostly in I-A and lastly M-C mostly in I-C.

Figure 36 - This graph illustrates the issue centric network where the nodes has been coloured in accordance with their association with the medium centric clusters. Blue represents cluster M-A, pink cluster M-B and orange cluster M-C.



These comparisons indicate that the clusters found in the issue centric network to some extent resembled the ones found in the medium centric. This can be interpreted as a sign that the issue centric network has been successful in encapsulating some of the topics and associated keywords of the controversy, but not all of them (qua the large amount of posts not included in the issue centric network).

#### 4.1.4 Summary of the panoramic comparison

The two networks are structurally very different. Where the medium centric has easily identifiable clusters (Figure 19), the issue centric is much better connected internally leading to a much denser network (Figure 27).

Furthermore, the networks have managed to assemble the issue publics around the HPV-controversy in very different subset sizes. Most notably the overlap of posts between the two networks only counts for  $\frac{1}{3}$  of the entire dataset and the amount of posts relating exclusively to the medium centric networks counts approximately the same (Figure 34).

Three clusters were found in each of the networks, and these clusters to some degree resemble each other across the networks both topically and structurally.

These empirical findings offer a few tentative substantial conclusions, going forward to the oligoptic comparison. Due to the relative low amount of overlapping entities between the two representations, it is suspected that the monadic assemblies will also differ notably in the amount of overlapping entities. Furthermore, due to the structural differences in the network it is suspected that the internal compound keeping the networks together manifested in monads' entity lists will also differ in amount.

What needs to be analysed now, is the connections between the posts on an oligoptic level to empirically validate or disapprove the tentative conclusions.

## 4.2 The oligoptic comparison

This section analyses the differences and similarities between the assemblies of the monads offered by the issue centric and medium centric networks, respectively. As explained earlier, this is done by selecting three cases based on highest, lowest and average degree.

### 4.2.1 Case 1: Highest degree

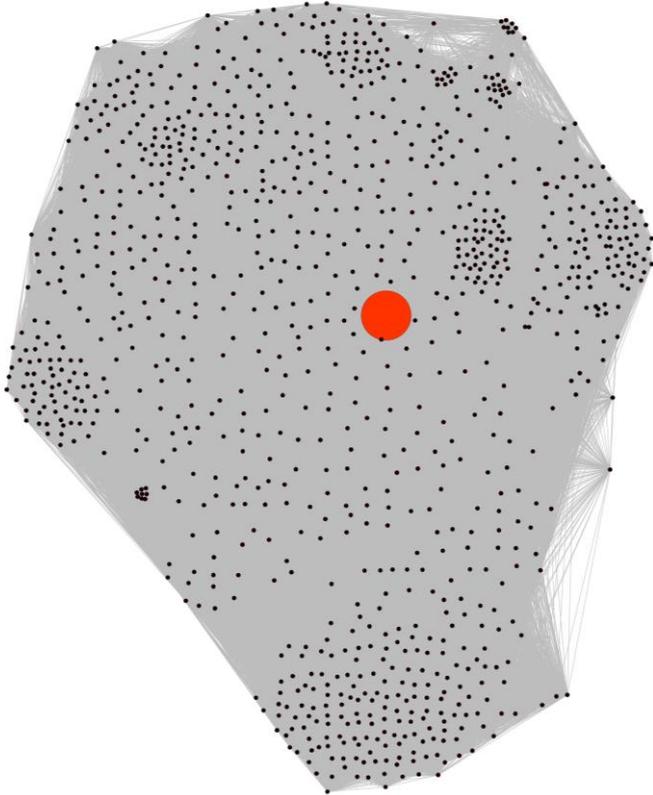
In the issue centric network, the monad with the highest degree, is the one that use most of the keywords from the issue dictionary. In the medium centric network, the monad with the highest degree is the one that shares the most user interactions with other posts.

#### 4.2.1.1 Highest degree from the issue centric network

The monad with the highest degree in the issue centric network, is a post by a page called *Medicin med omtanke*. The content is a letter from the medical doctor Claus Werner-Jensen, explaining the timeline of the use of the HPV vaccine brand called Gardasil.

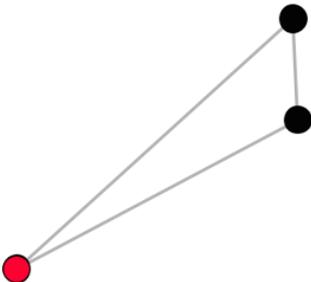
The post mentions all of the keywords of the issue dictionary, creating a total 1,045 links to other posts. As is visible in Figure 37, the network constituted by the entity list of the monad resembles the original issue centric network.

Figure 37 - This graph depicts the network around the most connected monad in the issue centric approach. The red node indicates the location of the monad being analysed.



Surprisingly, the medium centric representation of this monad reveals an entirely different and quite remarkable story than the issue centric. In the medium centric assembly, the monad is only linked to two other monads, making it one of the least connected in the network (Figure 38).

Figure 38 - This graph depicts the same monad as figure 36, but here assembled by the medium centric approach. The red node indicates the location of the monad being analysed.



This striking difference between the two representations might in fact in the most illustrative way encapsulate the relevance of this study. In the former representation, the monad is represented as the most central stage for the issues of the controversy, and in the latter represented as the least central. This will unquestionably lead to substantially different conclusions in an analysis.

#### 4.2.1.2 Highest degree from the medium centric network

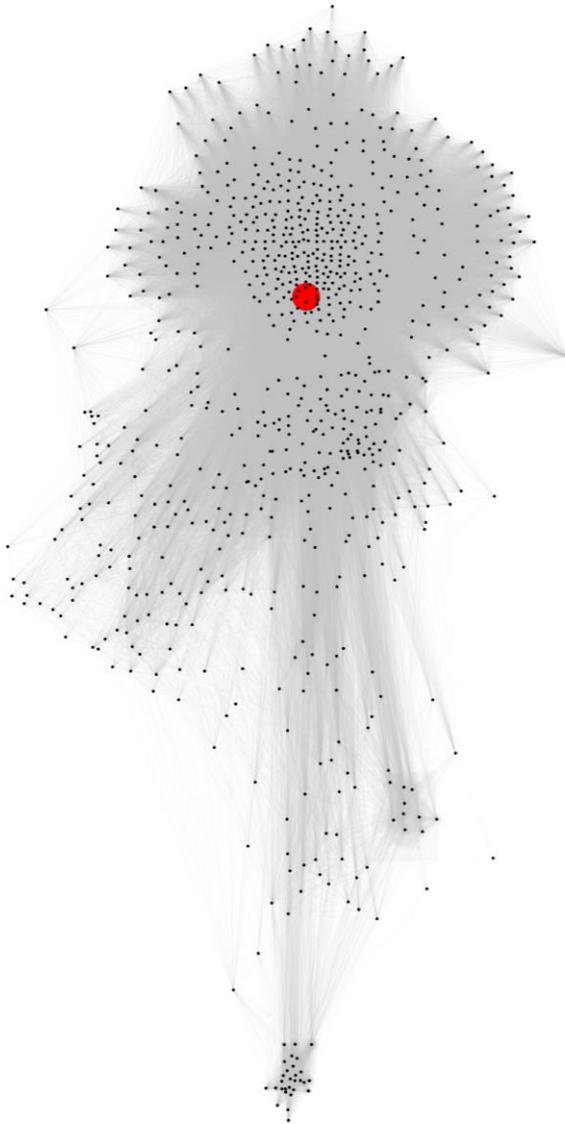
Turning towards the medium centric network, the most well-connected monad in terms of degree is a post from the page *HPV Vaccine Info*. The post claims that the Danish Health Authorities cover up information allegedly showing, that the HPV-vaccine increases the chances of cervical cancer (Figure 39).

Figure 39 - The post of the monad with the highest degree in the medium centric approach (*HPV Vaccine Info 2018B*)



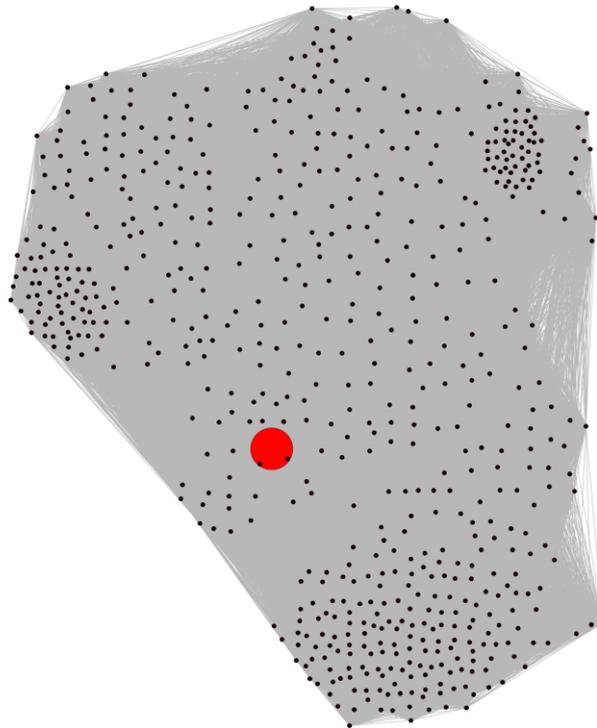
The post is connected to 815 other posts and also resembles original network (Figure 40).

Figure 40 - This graph depicts the network around the most connected monad in the medium centric approach. The red node indicates the location of the monad being analysed.



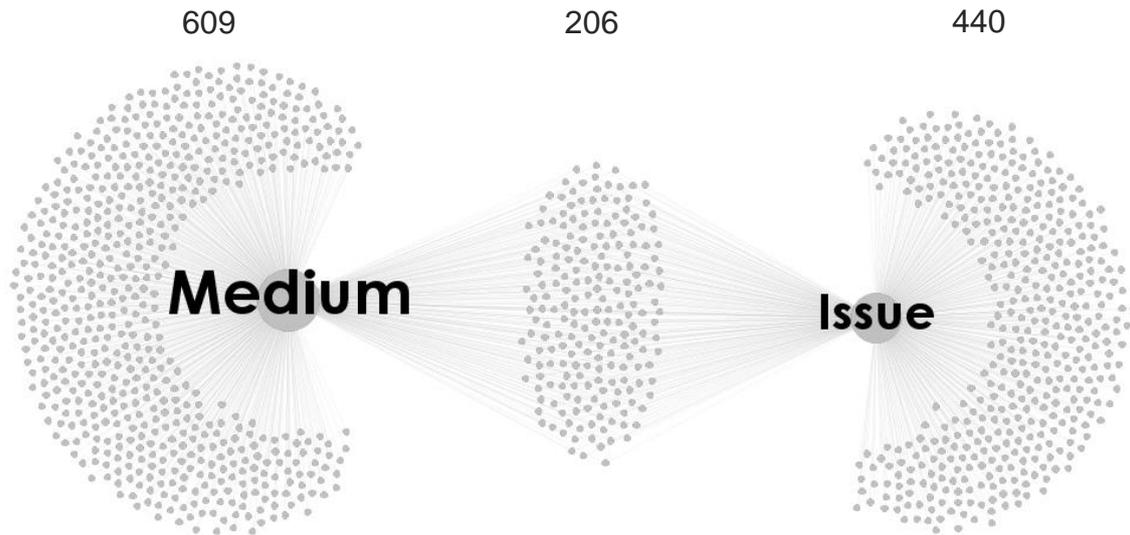
In the issue centric network this monad is connected to 646, again in a network that resembles the original (Figure 41).

Figure 41 - This graph depicts the same monad as Figure 40, but here assembled by the issue centric approach. The red node indicates the location of the monad being analysed.



The overlap between the two representations of the monad is shown in Figure 42. The overlapping elements between the two ways of assembling the monad is no more than 206, following the tendency showed earlier when analysing the structural assembly of the networks. This points towards a conclusion - for the highly connected monads at least - that even though close numbers, the two approaches compose the entity lists in completely different ways.

Figure 42 - This graph illustrates the overlap between the two representations of the monad with the highest degree in the medium centric approach. The numbers refer to the amount of monads in each group.



#### 4.2.2 Case 2: Average degree

The average degree is calculated based on the total number connections in the network divided by the total number of monads. For the issue centric network the average degree is 150 and for the medium centric 84.

##### 4.2.2.1 Average degree from the issue centric network

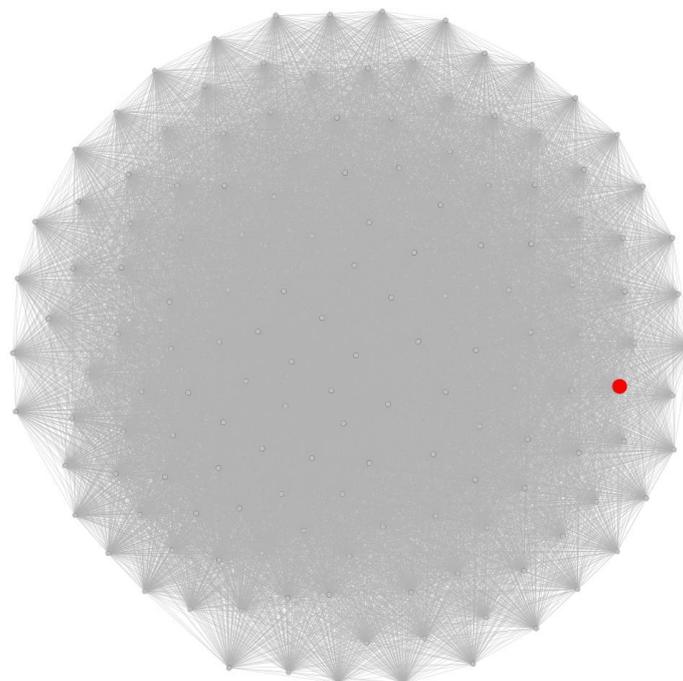
The case selected to represent the average monad in the issue centric network is in the form of a post from the page *Anette Harbech Olesen (Madforlivet.com)* (Figure 43). The post itself does not address HPV, but it is mentioned in one of the comments.

Figure 43 - The post of the monad with the average degree in the issue centric approach (Olesen 2018).



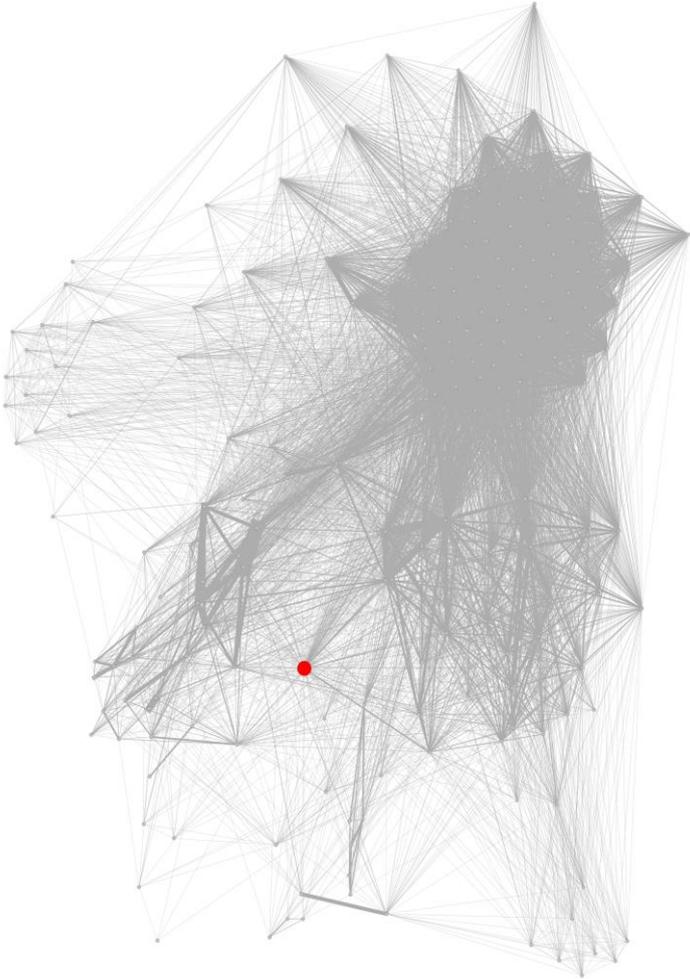
The circle-like shape of the ego network of the monad, informs that most of the nodes are connected, and as such does not show any tendencies of clustering (Figure 44).

Figure 44 - This graph illustrates the network around the average monad from the issue centric approach. The red node indicates the monad being analysed.



Performing the same filtration in the medium centric show a different story. Here, the network is more diverse, and the average monad followed is in the outskirts of a larger cluster (Figure 45). The number of other monads in the network is 193.

*Figure 45 - This graph depicts the same monad as Figure 44, but here assembled by the medium centric approach. The red node indicates the location of the monad being analysed.*

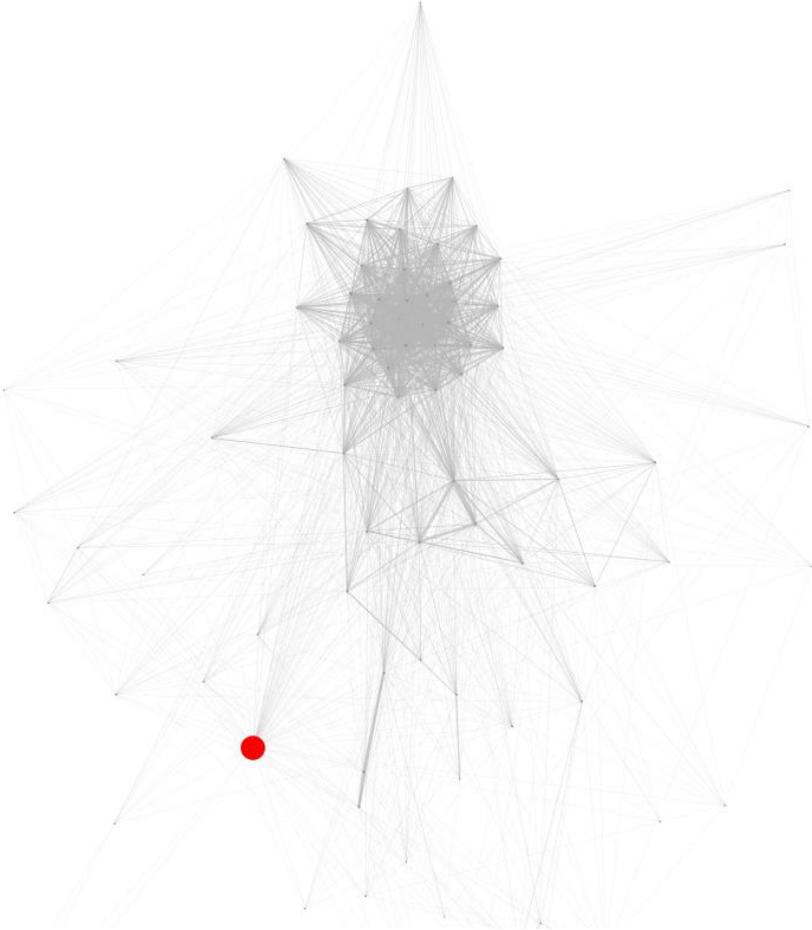


The overlap between the two ways of assembling the monad follows previous: only a few of the total amount are present in both representations (Figure 46).



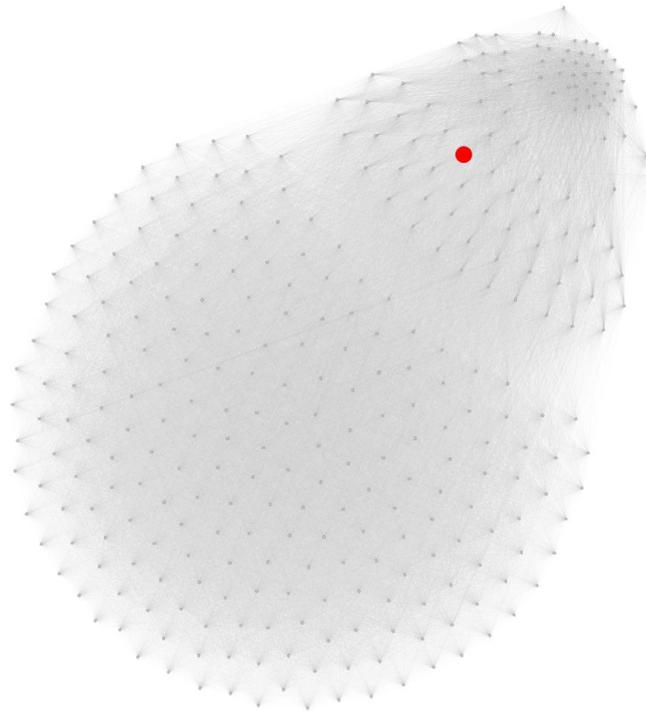
As with the previous medium derived networks, tendencies to clustering are present, and the monad in question is once again in the perimeter of the main grouping of nodes (Figure 48).

*Figure 48 - This graph illustrates the network around the average monad from the medium centric approach. The red node indicates the monad being analysed.*



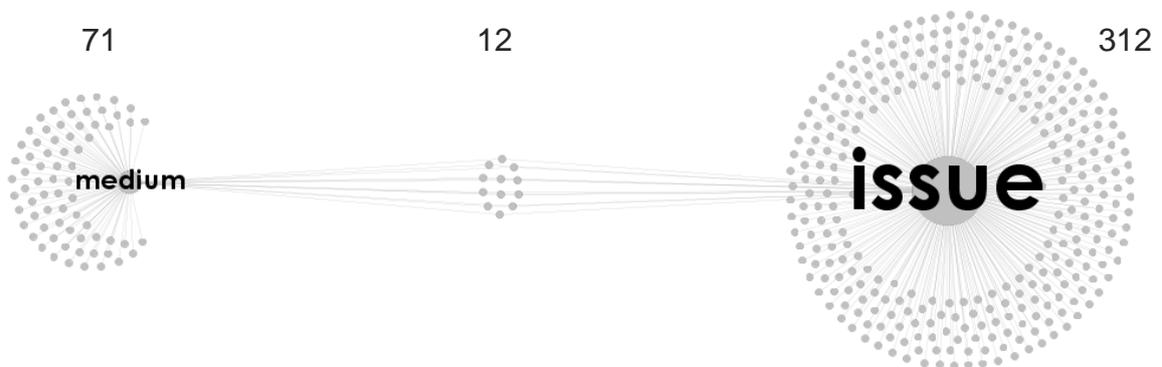
In the issue centric network, this monad is connected to 325 other monads, which is almost four times as many as in the medium network (Figure 49).

Figure 49- This graph depicts the same monad as Figure 48, but here assembled by the issue centric approach. The red node indicates the location of the monad being analysed.



Calculating the overlap show once again that the representations are very different from one another (Figure 50).

Figure 50 - This graph illustrates the overlap between the two representations of the monad with the average degree in the medium centric approach. The numbers refer to the amount of monads in each group.



#### 4.2.3 Case 3: Low degree

The low degree case does not really make sense to compare visually. This is because at least one of the representations in the comparison is only connected to one other monad.

#### 4.2.3.1 Low degree from the issue centric network

The post representing the issue centric approach for assembling the low degree monads only has one other monad linked to it. In the medium centric network, the monad is connected to 116 other monads.

#### 4.2.3.2 Low degree from the medium centric network

Finally, the post representing the medium centric assembly of the low degree monads also only has one other monad linked to it. In the issue centric network, the monad is connected to 28 other monads.

#### 4.2.4 Oligoptic summary

Having empirically analysed and compared networks through the oligoptic view one finding outshines all others: the issue centric and medium centric approaches assemble the monadic structures very differently.

The most connected monad in the issue centric network turns out to be one of the least connected in the medium centric. The least connected monads with a degree of 1, turns out to be relatively well connected with 28 and 116 links to other monads for the issue centric and the medium centric respectively. The representations that resemble one another the most in respect to the amount of connections are in fact the ones with the average degree.

When looking at the overlaps between the two representations, the difference become even more striking, as they in all cases can best be defined as minimal.

The conclusion of the oligoptic comparison based on the empirical evidence is that the general trend points towards an almost complete misalignment between the two representations of the monads. This may potentially lead to substantially different conclusions drawn from digital methods projects. This point will be developed further in the next chapter, where the implications will be discussed in relations to social theory in general.

## 5 Discussion

In this study I have taken the long way through the comparison of the issue centric and the medium centric approaches in order to empirically account for their implications on the study of the social. In the previous chapter I analysed the panoramic and oligoptic views of the monads and how the monads are represented differently by the two approaches.

It is now time to critically reflect on the outcomes of the comparison. Both in relation to the circumstances under which the network graphs of the monads were produced. But also in relation to how each of the two approaches impact the study of the social.

In the first section of this chapter I discuss what I call *Robust methods*. This entails discussing how the tools and methods used in this study impact the results, as well as how the lessons learned might help improve the quality of said tools and methods.

In the second section I compile the empirical data, signs and pointers accumulated in the analysis into a discussion about how the each of the two approaches informs and impact the study of the social.

### 5.1 Robust methods

This section can be seen as an input to the general discussion in digital methods on the impact of methods, tools and protocols on studies of the social. In the first part I will revisit the issue dictionary and discuss how the composition may impact the results of the study. In the second part I focus on the impact of the way the oligoptic comparison has been carried out.

#### 5.1.1 Revisiting the issue dictionary

As has been hinted several times during this study, the composition of the issue dictionary determines the composition of an issue centric network. If the keywords used are too generic, the resulting network will become highly connected, dense and difficult to analyse. This is because the generic terms “pulls” the network together and overshadows the particular terms representing the issues of the controversy. I have been aware of this risk all along. I have tried to introduce different measures to filter out the most generic terms. These measures include lists of stop-words combined with the issue specific knowledge of the participants in the PUSH-project. But in the iterative process of creating networks and adjusting the issue dictionary according to the composition of the networks, it became clear

that the counter measures employed were simply not sufficient to generate satisfying results (i.e. clearly clustered networks and foregrounding the particular). At first, I saw this as a major setback for the study. How could I possibly say something meaningful about the networks, if they refused to be analysed? But I soon realised, that rather than seeing it as a threat to the study, it should be seen as a classic case of stumbling on a new problem when trying to solve the original. What I was trying to achieve had not been operationalised into a protocol in digital methods literature. Speaking from the experiences gathered in this study, such an operationalisation could very well be its own thesis or paper worthy - it is not a job easily done.

To demonstrate the importance of a well composed issue dictionary, take the example of a topic that the issue experts proposed for the issue dictionary. The topic 'Time markers'<sup>18</sup> encompass Danish equivalents for the keywords 'begin', 'week', 'first' etc. At least one of these keywords were present in each of the posts in the issue centric network, leading to what is known as a "hairball" network where all nodes are highly connected. As the name implies, such a network is completely intertwined and compact. This in turn makes grouping, clustering and categorisations extremely difficult. Because of the risk of generic keywords, researcher has to navigate the complex field of tension between wanting to capture the issues of the controversy as laid out by the experts, and the ability to make meaningful analysis on the result.

The consequences of such navigation has definitely impacted the representation of the monads in the issue centric approach and consequently the comparison with the medium centric. Without a thorough empirical investigation, it is difficult to assess exactly how. That being said, I argue that the network composed for this study follow the general traits of the issue centric approach. Albeit the keywords to assemble the monads might not have been entirely attuned to the controversy, the network in itself and the principles of assembly remain valid.

The development of a protocol for composing issue dictionaries is not only of interest for this particular study. It is also deemed highly relevant for the digital methods community in general, as the medium centric approach might not be practically possible in future research projects. This can be seen as a consequence of both GDPR and the Cambridge Analytica controversy. Data repurposed by digital methods researchers to assemble the medium centric networks has either become locked up in the walled gardens of the social media

---

<sup>18</sup> Tidsmarkører

platforms or become subject to the protective rules of GDPR. This means that all information that makes users uniquely identifiable in a network are as of now off limits.

When the information that makes each monadic entry uniquely identifiable is no longer available, questions on actions and reactions, ties and affiliation, groupings and anti-groupings becomes difficult to dissect and answer. At least when following the medium centric approach. It becomes impossible to identify who engages in the different conversations. The public has become faceless and unquantifiable through the API. The observer is left with strings of text pointing in only one direction. Every single utterance stands by itself.

This could, and maybe also should to some extent, arguably be seen as a victory to privacy rights. Third party developers no longer have access to endpoints which could be exploited for commercial and political gain through e.g. micro targeting of users. In the name of privacy rights, public pressure has allowed Facebook to introduce walls around the data, not only protecting it from misuse by companies like Cambridge Analytica, but also from academic inquiry into the social life being organised on Facebook.

In this light the need for functioning and robust protocols for issue dictionaries attuned to Danish language seems more important than ever.

### 5.1.2 Oligoptic comparison based on cases

As part of the oligoptic comparison, I selected three cases from each network graph based on a measure for connectedness and centrality. The three cases selected were the monads with the highest and lowest degrees, as well as a representative for average degree.

The argument I used was primarily that a study based on each of the 3,059 monads treated severally, would be too comprehensive for this study. To limit this number, I therefore chose to select a suitable amount of cases. The choice reflects an assumption, that the distribution of degree can be used as a measure to select representative and generalisable cases.

The problem is, that such assumptions can only be rooted in a gut feeling as a mixture of experience and fingers crossed. And if there is one thing this study has shown it is that when it comes to digital methods assumptions should always be empirically supported. The alternative could very well be to embrace the predefined categories of Durkheim.

I do not wish to claim that the validity of the study has been jeopardised in general. But I wish to acknowledge, that it introduces limitations in regard to generalisability. The cases selected may represent the high, low and average monads in the HPV-controversy. And the assumption that this selection represent relevant categories may also apply to the study of the HPV-controversy. But in order to make the conclusions of this study truly generalisable, I propose that the selection criteria for what counts as representativity is thoroughly investigated.

Such an investigation could take its starting point in this study. By analysing the distribution of degree across the monads, it might be possible to in debt account for, what is represented by high, low and average degree.

As the empiricist I consider myself to be, I do not wish to jump to any conclusions and introduce predefined categories. Instead categories must be extracted from the data. It must be made clear, what is actually represented. Else, the entire endeavour will be to no avail.

## 5.2 The implications for social theory

This study's empirical contribution clearly shows, that Tarde's monads are represented in very different ways by the issue centric and medium centric approach respectively. Both in regard to the structural compositions of the networks, where the subsets of monads attaining only the issue centric and the medium centric respectively are more or less identical in numbers of entities (Figure 34). But also in regard to the monads taken severally, where the most central issue centric monad turns out to be one of the least central medium centric monads.

In the following, the implications of the empirical results for the study of the social will be discussed. The first section lay out the implications of the two ways of attributing centrality to actors and stages, whereas the second discuss the implications of how the monads are assembled.

### 5.2.1 Assessing central actors and stages

Most digital methods projects in one way or the other use centrality measures to determine what actors and stages function as authorities in a controversy. This in turn means, that the way centrality is operationalised in each of the two approaches determines how authority is attributed to the monads.

What is deemed as a centrality marker for the issue centric monad is that it uses the *issue specific vocabulary* on *issue specific stages*. The issue specific stages are in this study appointed in beforehand by the composition of the dataset. They are therefore shared by both the issue centric and medium centric representations. But the issue specific vocabulary is left for the researcher and the issue experts to derive. This leads to the conclusion that a monad's ability to muster allies correlates with its ability to match the list of keywords composed for the controversy.

In the case of the medium centric approach, centrality is operationalised as user overlaps with other monads. This leads to the conclusion that centrality of a monad correlates with its ability to attract user interactions.

Through the oligoptic comparison we see that these two ways of operationalizing centrality is in complete misalignment with each other. The most central issue centric monad turned out to be one of the least central medium centric monads.

As I argued in the introduction the issue centric mapping exercise will be in favour of the ones who manage to assemble the most durable vocabulary. Regardless of the exposure to other actors, the centrality measure of the issue centric approach dictates that using the controversy vocabulary distributes centrality in the debate. Is this the "right" representation of centrality in a controversy? As ANT describes actors (and monads) are human and non-human ensembles that acts or makes others act. How does the most central issue centric monad make others act, if nobody (besides the researchers studying the controversy) notice its presence?

In the medium centric network graph, it is not possible to determine how many different actors are responsible for the connections between monads. In the case of the most connected medium centric monad, all connections to other monads could potentially have been made by a single user on Facebook. As such, the approach does not discriminate based on the amount of unique actors. As long as a user is detected between two monads, a link will be created. If several users are detected, the link becomes stronger.

The medium centric distribution of centrality further complicates the definition of the "right" representation of the term. If a single user can be responsible for linking all the monads to each other in the medium centric network, how is that a "better" representation of centrality than using the right vocabulary?

The consequences of the two approaches could not have been derived, had it not been for an empirical comparison between the assembly of the monads. And the questions derived from the consequences are of utmost importance for researchers doing digital methods projects. It is now clearly evident that the representation of centrality varies between the two approaches. It will probably also vary from medium to medium and from controversy to controversy. When someone based on a network graph claim that this or that monad is the most central, powerful or important, the first question we should ask is how centrality has been operationalised and what empirical evidence brings such a conclusion to be.

### 5.2.2 Assembling the monads as representations of issue publics

Following the previous section's friendly scolding of digital methods researchers, I will now discuss the particularities of the two ways of assembling the monads. This will be related to the consequences of how the issue public is represented.

When I have discussed the topic of my thesis with others, for some reason the conversation often turns toward the role of social media in public disputes. Some say, that social media has accelerated the spread of misinformation in so-called filter bubbles. Others say that social media is a victory for democracy, because everybody now gets to have a say in public matters.

Personally, I consider the question mute. I argue, that social media does not per se change how controversies play out. Misinformation is not new (Public Data Lab 2017), media has always accelerated (Latour 1986) and if Dewey's definition of the public extends to present time the public is still constituted by consequences of transactions that turn into controversies. What has changed is the realm of possibilities to reassemble and analyse the different issue publics, because their associational traces are available online.

This means, that the two different ways to assemble the monads of the HPV-controversy may also be seen as two different ways of assembling issue publics of the controversy.

On the last day of the PUSH data sprint, one of the anthropologists said that she in the dataset had found what she believed was the origin of the public mobilisation in the HPV-dispute. It happened, she told, as the organisation Sex and Samfund refused to discuss the alleged side-effects from the HPV-vaccine with worried parents on Facebook. Instead Sex og Samfund referred to the guidelines of the National Board of Health. As a result a Deweyan

public composed by the worried parents and others impacted by the consequences of the HPV-vaccine was sparked into being.

The medium centric assembly of the monads shows a clear divide between two problematizations that reside in the HPV controversy. One problematisation attains to the consequences of the HPV-vaccine and the other attains to the consequences of the HPV-infection. Furthermore, this problematisation divide in the networked representation is supported by a divide between public authorities on the one side and issue public on the other. As such, the medium centric approach follows the tell-tales of Dewey.

The issue centric assembly on the other hand, show clear divides between the topics of the dispute. But neither the problematisation divide nor the authority/public divide is possible to derive clearly from the network.

I therefore argue, that where the medium centric approach may be used to show *who* problematises, the issue centric approach may be used to expand on *what* is being problematised.

## 6 Conclusion

In this thesis I have been dedicated to disentangle two distinct dogmas that has been collapsed in digital methods: *follow the medium* and *follow the actors in controversies* respectively.

I have identified these two dogmas in digital methods literature as two distinct ways of assembling the social in a way that is consistent with both Tarde and Actor-Network Theory.

The main problem I have addressed in this thesis is, that no actual test of the implications of the different ways to represent the social and to build the monad has been carried out. This is a problem, as it from an empirical point of view obscures what the one has to offer compared with the other.

The first contribution of this thesis is thus methodological. By separating and operationalising the two dogmas into two approaches – the *issue centric* and the *medium centric* respectively – it is for the first time possible to conduct a transparent comparison and evaluation of the consequences for the representation of the social. The output of these novel

operationalisations is in both approaches a network graph representing monads' social associations in the HPV-controversy.

The second contribution of this thesis is empirical. From the novel operationalisations of the two approaches an empirical analysis has been carried out. Results show that the two approaches represent the monads and their social associations of the HPV-controversy in very different ways. A clear example of this difference is found in the low number of overlapping monads in the network graphs (Figure 34). This conclusion is extended when looking at the clusters of monads found in each network. Comparing the clusters found in the issue centric and the medium centric network graphs respectively show no clear signs of similarity. Neither on a structural level nor on a topical level. Finally, results show that the two approaches identify central and well-connected monads in very distinct ways. The issue centric approach attribute centrality and authority to the monads that use the *right* vocabulary. The medium centric approach attribute centrality and authority to the monads that has the most interactions from other actors of the controversy. As shown, these two definitions of centrality do not resemble one another.

The empirical findings need to be understood in the context of the methods and tools used in the study. As discussed, the method for constructing the issue dictionary is in need of a thorough analysis in order to assess how it empirically impacts the assembly of the network. I therefore recommend that further research into this subject is carried out. The same applies to the way the oligoptic cases has been selected.

The third contribution of this thesis is theoretical. Taking the empirical conclusions into consideration, the implications of non-reflected use of interwoven methods are potentially severe. We cannot any longer assume that the two approaches result in the same presentation of a controversy. By choosing the one approach or the other, we at the same time make a theoretical choice of how social associations are perceived.

Finally, I propose, that the results presented here should first and foremost be taken as a palpable reminder of the importance of understanding in dept the machinery set in motion when choosing an item from the digital methods toolbox. Each time a tool is used, a method is applied, and a theory followed it entails a subscription to build in assumptions and definitions. As I have shown these assumptions and definitions are often implicit and intertwined. But through inspections of their heritage and the empirical particularities they can be separated, and their implications can be made explicit. This is, in my opinion, what digital methods researchers should strive for.

## 7 References

Aarhus University. 2018. *PUSH - Public Health at a crossroads*. Aarhus: AU.

[http://pure.au.dk/portal/da/projects/push--public-health-at-a-crossroads\(f5b63775-e94c-40cd-a2e6-a4773a88477f\).html](http://pure.au.dk/portal/da/projects/push--public-health-at-a-crossroads(f5b63775-e94c-40cd-a2e6-a4773a88477f).html)

Agar, M. (2006). *An ethnography by any other name*. Forum : Qualitative Social Research

Barry, A., and Thrift, N. 2007. *Gabriel Tarde: Imitation, Invention and Economy*. *Economy and Society*, 36:4, 509 - 525

Bloor, D. and Wright, J. D. 2015. *Sociology of the Strong Program*. In *International Encyclopedia of the Social & Behavioral Sciences* (pp. 592-597). Burlington: Elsevier.

Boellstorff, T., Nardi, B., Pearce, C., and Taylor, T.L. 2012. *Ethnography and virtual worlds: A handbook of method*. Princeton, N.J., Princeton University Press

Boellstorff, T. 2015. *Coming of Age in Second Life: An Anthropologist Explores the Virtually Human*. Princeton, N.J.:Princeton University Press

Braga, M. (2018) *How one researcher harvested data from 50 million people - and Facebook was designed to help*. <http://www.cbc.ca/news/technology/facebook-cambridge-analytica-friends-api-by-design-1.4583337>

Bruun, C. J., Lauritsen, P. and Olesen, F. 2007. *Introduktion til STS: science, technology, society*. Kbh.:Hans Reitzel

Callon, M. and Latour, B. 1981 *Unscrewing the big Leviathan; or how actors macrostructure reality and how sociologists help them to do so*. In K. Knorr et A. Cicourel (editors) *Advances in Social Theory and Methodology*, Routledge and Kegan Paul, Londres, pp. 277-303.

Callon, M., Courtial, J.-P., Turner, W. A. and Bauin, S. 1983, *From translations to problematic networks: An introduction to co-word analysis*, *Social Science Information*, 22 (2): 191–235.

Callon, M. 1987. *Society in the Making: The Study of Technology as a Tool For Sociological Analysis*. In Bijker, W., Hughes, T., & Pinch, T. (eds.) *The Social Construction of Technological Systems*. Cambridge: MIT Press.

Callon, M. (1986) "Some Elements of a Sociology of Translation: The Domestication of the Scallops and the Fishermen of St.Brieuc Bay." In Law, J. (ed.) *Power, Action & Belief: A New Sociology of Knowledge?* London: Routledge & Kegan Paul.

Clark, A., Fox C., and Lappin S. 2010 *The Handbook of Computational Linguistics and Natural Language Processing*. Blackwell. ISBN:9781405155816

Collins, H. M. 1983. *An empirical relativist programme in the sociology of scientific knowledge*. In: Knorr-Cetina, K. and Mulkay, M. eds. *Science Observed: Perspectives on the Social Study of Science*, London: Sage, pp. 85-114.

Cressman, D. 2009. *A brief overview of ANT*

<http://faculty.georgetown.edu/irvinem/theory/Cressman-ABriefOverviewofANT.pdf>

Diaz, G. 2016. *Danish stopwords collection*. <https://github.com/stopwords-iso/stopwords-da>

Durkheim, E. 1951. *Suicide: A Study in Sociology*. Glencoe: The Free Press, transl. John A. Spaulding and George Simpson

Durkheim, E. 2000. *Den sociologiske metode*. København: Hans Reitzels forlag

Epstein, S. 1996. *Impure science: AIDS, activism, and the politics of knowledge* (Vol. 7). Univ of California Press.

Epstein, S. 1995. *The Construction of Lay Expertise: AIDS Activism and the Forging of Credibility in the Reform of Clinical Trials*. *Science, Technology, & Human Values*, Vol. 20, No. 4, Special Issue: Constructivist Perspectives on Medical Work: Medical Practices and Science and Technology Studies pp. 408-437

Facebook. 2015. *Changelog Version 2.3*. [https://developers.facebook.com/docs/graph-api/changelog/archive#v2\\_3\\_new\\_features](https://developers.facebook.com/docs/graph-api/changelog/archive#v2_3_new_features)

Facebook. 2017. *Changelog Version 2.11*. <https://developers.facebook.com/docs/graph-api/changelog/version2.11>

Facebook. 2018A. Graph API Overview.  
<https://developers.facebook.com/docs/graph-api/overview>

Facebook. 2018B. *Changelog Version 2.12*. <https://developers.facebook.com/docs/graph-api/changelog/version2.12>

Facebook. 2018C. *Facebook About*.  
[https://www.facebook.com/pg/facebook/about/?ref=page\\_internal](https://www.facebook.com/pg/facebook/about/?ref=page_internal)

Facebook. 2018D. *TV 2 Nyhederne*. <https://www.facebook.com/tv2nyhederne/>

Facebook. 2018E. *Reactions now available globally*.  
<https://newsroom.fb.com/news/2016/02/reactions-now-available-globally/>

Geertz, C. 1972. Deep Play: Notes on the Balinese Cockfight. In *Daedalus*, 101(1), 1-37.

Goodenough, Ward (1970) "Describing a Culture" in *Description and Comparison in Cultural Anthropology* Cambridge: Cambridge University Press pp 104-119

HPV Vaccine Info. 2018. *HPV Vaccine Info*.  
<https://www.facebook.com/hpvvaccineinfo/posts/1104552622996321>

HPV Vaccine Info. 2018B. *HPV Vaccine Info*.  
<https://www.facebook.com/hpvvaccineinfo/posts/1317301755054739>

Heilpraktikerskolen. 2018. *Heilpraktikerskolen*.  
<https://www.facebook.com/heilpraktikerskolen/posts/1025238950831712>

Jacomy, M., Venturini, T., Heymann, S., Bastian, M. 2014. *ForceAtlas2, a Continuous Graph Layout Algorithm for Handy Network Visualization Designed for the Gephi Software*. *PLoS ONE* 9(6): e98679. <https://doi.org/10.1371/journal.pone.0098679>

Jacomy, M., Girard, P., Ooghe-Tabanou, B. and Venturini, T. 2016. *Hyphe, a Curation-Oriented Approach to Web Crawling for the Social Sciences*. Paper presented at INTERNATIONAL AAAI CONFERENCE ON WEB AND SOCIAL MEDIA, Köln, 2016.

Jensen, T., E. 2003. Aktør-Netværksteori – en sociologi om kendsgerninger, karakterer og kammuslinger. Papers in Organization, No. 48

Jensen, T.E., Madsen, A.K., Misfeldt, M., Munk, A.K, Tamborg, A. 2017. *Participatorisk Data Design: En ressource for capacity building*. (H.K. Krogstrup ed.) Samskabelse og Capacity Building, KBH:Hans Reitzel.

Journalisterne. 2018. *Journalisterne*.

<https://www.facebook.com/journalistendk/posts/1316841081704951>

Kuhn, T.S. 1962 *The Structure of Scientific Revolutions*. Chicago: University of Chicago Press.

Kræftens Bekæmpelse. 2018. *Facebook Post*.

<https://www.facebook.com/CancerMiddelfart/posts/491908560864485>

Latour, B. 2005B. *From Realpolitik to Dingpolitik – or How to Make Things Public*. In Latour, B. (ed) Making Things Public. MIT Press:Cambridge MA

Latour, B. 1999. *On Recalling Ant*. The Sociological Review, 47, 15-25. doi: 10.1111/j.1467-954X.1999.tb03480.x

Latour, B., 2018. *The Debate*. <http://www.bruno-latour.fr/sites/default/files/downloads/TARDE-DURKHEIM-GB.pdf>

Latour et. al. 2012. The whole is always smaller than its parts – a digital test of Gabriel Tarde's monads. The British Journal of Sociology 2012 Volume 63 Issue 4

Latour, B., 2005. *Reassembling the social*. Oxford : Oxford University Press

Latour, B., 1998 *Thought Experiments in Social Science: from the Social Contract to Virtual Society*. 1st Virtual Society? Annual Public Lecture. London: Brunel University  
<http://www.artefaktum.hu/it/Latour.htm>

Latour, B. 1986. *Visualisation and Cognition: Drawing Things Together*. In H. Kuklick (editor) *Knowledge and Society Studies in the Sociology of Culture Past and Present* Jai Press vol. 6, pp. 1-40.

Latour, B. 1996. On actor-network theory. A few clarifications plus more than a few complications. *Soziale Welt*, vol. 47, pp. 369-381.

Latour, B., and Woolgar, S. 1979. *Laboratory life: the construction of scientific facts*. Princeton: Princeton University Press, ISBN: 0691094187

Latour, B. 1993. *The pasteurization of France*. Harvard University Press.

Lippmann, W., 1993. *The Phantom Public* Transaction Publishers, New Brunswick, New Jersey ISBN: 1-56000-677-3

Lægemiddelstyrelsen. 2018. *Lægemiddelstyrelsen*.

<https://www.facebook.com/laegemiddel/posts/273401956477181>

Madsen, A. Koed. 2012. 'Web-visions as Controversy-lenses. *Interdisciplinary Science Review* 1 (37): 51-68.

Marres, N., 2012, The redistribution of methods: on intervention in digital social research, broadly conceived, *The Sociological Review*, 60: 139–165.

Marres, N. 2015. *Why map issues? On controversy analysis as a digital method*. *Science, Technology and Human Values*, online first, doi: 10.1177/0162243915574602.

Marres, N., and Rogers, R. 2005. *Recipe for tracing the fate of issues and their publics on the Web*. in B. Latour and P. Weibel (eds), *Making Things Public: Atmospheres of Democracy*, 922–935, Cambridge, MA: MIT Press.

Marres, N. and Gerlitz, C. 2016. *Interface methods: renegotiating relations between digital social research, STS and sociology*. *The Sociological Review*, Vol. 64, 21–46 (2016) DOI: 10.1111/1467-954X.12314

Miller, D. 2016. *Social Media in an English Village: Or How to Keep People at Just the Right Distance*. London: UCL Press.

Munk, A. K. and Abrahamsson, S. 2012. *Empiricist Interventions: Strategy and Tactics on the Ontological Battlefield*. *Science Studies*, Vol. 25 (2012) No. 1, 52-70

Munk, A. K. 2014. *Mapping Wind Energy Controversies Online: Introduction to Methods and Datasets*.

[https://papers.ssrn.com/sol3/Delivery.cfm/SSRN\\_ID2595287\\_code2345181.pdf?abstractid=2595287&mirid=1](https://papers.ssrn.com/sol3/Delivery.cfm/SSRN_ID2595287_code2345181.pdf?abstractid=2595287&mirid=1)

Munk, A. K. 2013. *Techno-Anthropology and the Digital Natives*, in Børsen (ed.) *What is Techno-Anthropology?*

Munk, A. K., Meunier, A., & Venturini, T. (2018). *Data sprints: A collaborative format in digital controversy mapping*. (J. Vertesi et al eds.) *DigitalSTS: A Handbook and Fieldguide*, Princeton University Press

Nardi, B. A. 2010. *My life as a night elf priest: an anthropological account of World of Warcraft*. Ann Arbor:University of Michigan Press.

Naturli. 2018. *Naturli*. <https://www.facebook.com/naturli.dk/posts/286212284871147>

Olesen, A. H. 2018. *Anette Harbech Olesen (Madforlivet.com)*.

<https://www.facebook.com/harbecholesen/posts/628276960527419>

Pinch, T. & Bijker, W.E. 1987. *The Social Construction of Facts and Artefacts*. I: Pinch, T. & Hughes, T.H. & Bijker, W.E., *The Social Construction of Technological Systems*. London: MIT Press. ISBN: 0-262-52137-7

Public Data Lab (2017). *A field guide to fake news and other information disorders*.

<https://fakenews.publicdatalab.org/>

Rieder, B. (2015) *The end of Netvizz?* <http://thepoliticsofsystems.net/2015/01/the-end-of-netvizz/>

Rieder, B. (2013). *Studying Facebook via data extraction*. Proceedings of the 5th Annual ACM Web Science Conference on - WebSci '13 Proceedings of the 5th Annual ACM Web Science Conference: 346-355.

Rogers, R. 2009. *The end of the virtual*. Inaugural speech.

[http://www.govcom.org/rogers\\_paris\\_medialab.pdf](http://www.govcom.org/rogers_paris_medialab.pdf)

Rogers, R. (2012). *Mapping and the Politics of Web Space*. Theory, Culture & Society 29(4/5) 193–219

Rogers, R. (2004). *Information Politics on the Web*. MIT Press, Cambridge, MA.

Rogers, R. 2018. *Short Description of the Issue Crawler and Allied Tools*.

[http://www.govcom.org/publications/full\\_list/rogers\\_issuecrawler\\_context.pdf](http://www.govcom.org/publications/full_list/rogers_issuecrawler_context.pdf)

Strathern, M. 1996. *Cutting the Network*. Journal of the Royal Anthropological Institute 2.3 : 517-535.

Spencer, H. 1940. *The man versus the state*. Caldwell

Tarde, G. 1903. *The Laws of Imitation*. New York: Henry Holt

Tarde, G. 1999 [1898]. *Les Lois Sociales*. Paris: Les Empêcheurs de Penser en Rond

Tarde, G. 2003 [1899]. *Les Transformations du Pouvoir*. Paris: Les Empêcheurs de Penser en Rond

Torp, B. 2018. *Dansk stopords liste*.

<https://gist.github.com/berteltrorp/0cf8a0c7afea7f25ed754f24cfc2467b>

Blondel V. D., Guillaume, J.L., Lambiotte, R. and Lefebvre, E. 2008. *Fast unfolding of communities in large networks*. in Journal of Statistical Mechanics: Theory and Experiment 2008 (10), P1000

Venturini, T., and Guido, D. 2012. *Once Upon a Text: An ANT Tale in Text Analysis*. Sociologica 62 (3). doi:10.2383/72700.

[https://medialab.sciencespo.fr/publications/Venturini\\_Guido-Once\\_Upon\\_A\\_Text.pdf](https://medialab.sciencespo.fr/publications/Venturini_Guido-Once_Upon_A_Text.pdf).

Venturini, T. 2012. Building on Faults: How to Represent Controversies with Digital Methods. *Public Understanding of Science* 7 (21): 796-812.

Venturini, T., Jacomy, J. and Pereira, D. 2015. *Visual network analysis*.  
[http://www.tommasoventurini.it/wp/wp-content/uploads/2014/08/Venturini-Jacomy\\_Visual-Network-Analysis\\_WorkingPaper.pdf](http://www.tommasoventurini.it/wp/wp-content/uploads/2014/08/Venturini-Jacomy_Visual-Network-Analysis_WorkingPaper.pdf)

Spradley, J. P. 1980. *Participant Observation*. Fort Worth:Harcourt Brace Jovanovich College Publishers

Vikkelsø, S. 2007. Description as intervention: Engagement and resistance in Actor-Network Analyses. *Science as Culture* 16(3):297-309

Woolgar, S. (ed.). 2002. *Virtual Society? Technology, Cyberbole, Reality*. Oxford: Oxford University Press

Zetland. 2018. *Zetland*. <https://www.facebook.com/zetlanddk/posts/1528422667185376>

## 8 Appendix

### 8.1 Appendix A – Table of topics and keywords

Topics	Keywords
Bivirkninger	Svimmel; høj puls; lammelse; smerter i underlivet; nældefeber; nervesystemet; kronisk hovedpine; CFS/ME; POTS; CRPS; autoimmun
Kontrollforsøg	Merck; MSD; GlaxoSmithKline; gardasil; saltvandsplacebo; gardasil; aluminiumindhold; gær; inaktiv virus; cervarix; FUTURE2
Behandlingsmuligheder	Homøopat; kostomlægning; vitaminer; udrenningskur; Vega test; Stig Gerdes; C-vitamin; Body SDS; zoneterapi; hormonterapi; rødkløver; akupunktur; kinesiologi; mineraler magnesium; c-vitamin; Lifewave akuplastre; kosttilskud; intravenøs modgift; naturmedicin; medicinsk cannabis; valle-protein; L-Lysin; Synkope Centeret; Frederiksberg Hospital; Jesper Mehlsen
Funktionelle lidelser	Funktionelle Lidelser; funktionel lidelse; Fibromyalgi; Somatoform; Whiplash; Infralyd; overfølsomhed; marianne rosendal; DSAM; Stig Gerdes; Trygfonden; Per Fink; BDS; IBS; træthedssyndrom; psykiatrisk
HPV-skade specialister	Stig Gerdes; Inger Lis Riis; Ole Kåre Føli; Helle Zeymer; Maiken Andersen; Leonard Coldwell; Deirdre Little; Jesper Mehlsen; Lægemiddelstyrelsen
Alternativer til HPV vaccination	Screening; celleskrab
Konsekvenser af HPV-infektion	Celleforandring; keglesnit; livmoderhalskræft; kønsvorter; papillomavirus
Videnskab	Forskning; evidens; facts; postfaktu; fake news; falske nyheder; information; viden; videnskab; rationel
Indholdsstoffer	aluminium; adjuvans; gær; cerevisiae
Kræft	Livmoderhalskræft; Kræftens Bekæmpelse; Vidunderlivet; HPV infektion; celleforandringer; cancer; HPV-virus
Indsatser	gratis vaccination; børnevaccinationsprogram; catch-up program; Sundhedsstyrelsen