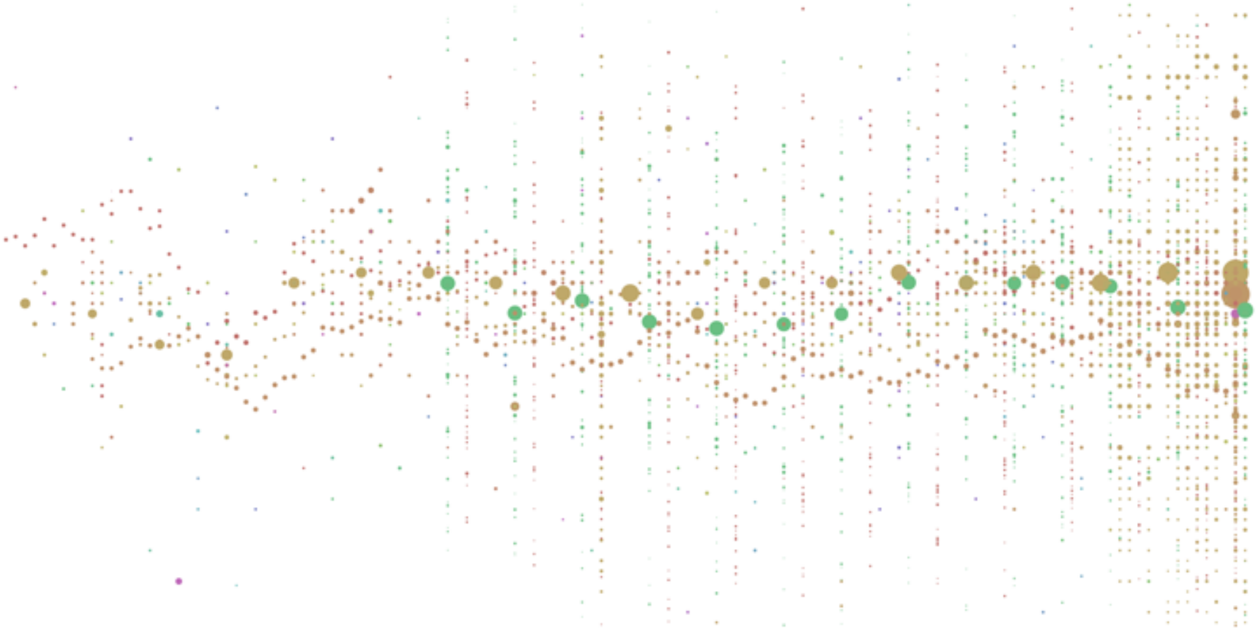
LINKÖPING UNIVERSITY



Interventions with Data: A Series of Workshops for Developing Alternatives to 'Big Data'. Project Report, April 2018

David Moats
Tema-T Technology and Social Change
Linköping University
david.moats@liu.se

Contents

Introduction 3
Workshop 1: Polling and Micro-Targeting 5
Workshop 2: Big Health Data 14
Workshop 3: Academic Rankings and Metrics 24
Findings 32
References 35
Participants 37

Introduction

Interventions with Data was a project which ran from June - December 2017, funded by a Research Initiation Grant from Riksbanken Jubileefund. The aim was to think through new ways of approaching so called 'big data', the kinds of digital traces like social media, feedback forms, online transactions, tracking data, GPS and open government data which are currently being gathered and analysed by governments, private companies and researchers. *The Economist* famously called data the new oil, a new natural resource which has spawned the creation of new industries and infrastructures.[1] Data flows between different companies as if through gushing pipelines and becomes stored in unstructured 'data lakes' and deposited in 'data warehouses'. While we should remain sceptical about claims to the new-ness of this data, and especially to discourses which refer to data as somehow 'naturally occurring', (data is by definition constructed and assumption laden: Gitelman, 2013), it is hard to deny that the hype around data is creating real transformations and anxieties in industry, governments and academia: redistributing roles and responsibilities between them.

But to continue *The Economist's* somewhat forced metaphor: if data is the new oil then there are important questions to be raised about how it is refined, turned into useful products and kept from leaking everywhere. Much of this data is processed with techniques like machine learning and artificial intelligence which attempt to automatically spot patterns in these massive datasets. These algorithms and the systems built around them are used to distribute resources, assign credit scores and deliver all sorts of content to our many devices. While these techniques have made great strides, they have also been accused of being reductionist or even dangerous. Legal Scholar Frank Pasquale (2015) argues that these systems are 'black boxed', their inner workings are unavailable for scrutiny, either to the people they effect or, sometimes even their creators. Cathy O'Neil (2016) has even gone so far as to describe the algorithmic systems used to process this data and as 'weapons of math destruction'.

We are in urgent need of alternative modes of analysis but these are not readily forthcoming. Many of the innovations in this area come from computer science or the natural sciences, when most of the data in question is at least nominally 'social' in character, involving interactions between humans or between humans and computers. Many of the sharpest critiques of these techniques come from philosophers, anthropologists and qualitative social scientists who have much to say about data and complex social phenomena, but who are also reluctant to get their hands dirty and experiment with computational techniques themselves. This is further hampered by old, somewhat outdated, splits between quantitative and qualitative methods, hermeneutic and positivist traditions in the social sciences, and tensions between adjacent disciplines (Barry et al., 2008). There are countless frameworks for diffusing these epistemological and disciplinary tensions from mixed methods to grounded theory (Fielding and Fielding, 2008; Glaser and Strauss, 1967) but these often presume the different camps and different methods as stable, singular (Hammersley, 1992) and separate entities to begin with, rather than examining the tensions and negotiations in practice (Neff et al., 2017).

[1] Available from: https://www.economist.com/news/leaders/21721656-data-economy-demands-new-approach-antitrust-rules-worlds-most-valuable-resource

Some researchers in the field of Science and Technology Studies (STS) have in recent years begun to experiment with the use of data visualisations to aide qualitative analyses and forms of public engagement, often involving social media or other publicly available data (Abildgaard et al., 2017; Rogers, 2013; Rogers and Marres, 2000; Venturini et al., 2014).[2] Visualisations in general may offer interesting alternatives because, while they necessarily involve algorithms and metrics, they foreground the role of (equipped) human interpretation in the process (Card et al., 1999). They also, it is claimed, open up the research process to a wider array of less technically-minded participants and topic experts, while at the same time, their seductive and flashy character creates new problems and blind spots to contend with (Coopmans, 2014; Kennedy et al., 2016). While there is a substantial literature about the implications of data visualisations for the social sciences, and implications for resolving 'quantitative' and 'qualitative' tensions (Venturini and Latour, 2010), there are few studies of how these tools might upset disciplinary identities and routines in situated encounters.

The project consisted of a series of three workshops dealing with different types of digital data. In these workshops we hoped to extend some of these experiments with more interpretivist visualisation techniques and apply them to different types of data (other than social media). The interventions in the title refers to the idea that we might learn something different about automated techniques of data analysis by trying (and sometimes, failing) to use them rather than just describing how they are currently used (Zuiderent-Jerak, 2015). Rather than detached observation, we actively engineered situations to trail alternative ways of analysing data. We chose three areas of social life which are being transformed by 'big data': political campaigns, health data, and academic metrics and rankings. The aim of each workshops was two-fold: to produce or mock up tools, approaches or visualisations and to reflect on the problems which emerge when different types of researchers use these techniques. While many researchers claim to be trans-disciplinary or beyond these tensions, we felt it was important to dwell on and explore these potential pitfalls and barriers to collaboration, while at the same time not presuming these splits as natural or given.

The workshops were all three days and were held at the Visualisation Centre in Linköping University's Norrköping Campus. They were modelled on the format of a hackathon, a gathering of programmers and topic experts over 2-3 days. However, there were a few key differences. Firstly, hackathons are often marked by imbalances between programmers and less tech-savvy participants (Munk et al., 2016; Ruppert et al., 2015). We thus made efforts to spend more time on defining problems and research questions and resisting the drive to swap these problems for technically solvable ones. Secondly, while we often started with pre-prepared data sets, the question of which data or what techniques we should use were purposely left open. Thirdly, the objective of the workshops was not to make something that definitively 'works' but to learn about the substantive topic and the research process through our sometimes-fumbling attempts to use digital tools. The workshops involved academics from Science and Technology Studies (STS), medical sociology, medicine, media studies, anthropology, information systems, computer science and library sciences. All the data we used was publicly available, but it still raised ethical questions, which we reflect on below.

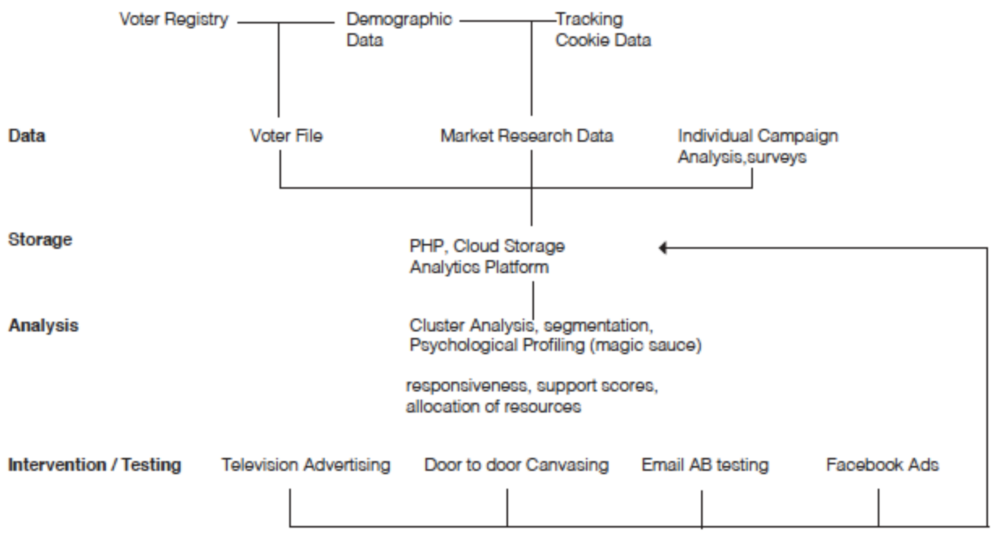
---

[2] See for example the Emaps Project: http://www.emapsproject.com/blog/

Workshop 1: Polling and Micro-Targeting

The first of the three workshops was prompted by recent, sensational claims that data analytics had played a central role in determining the result of the UK EU membership referendum, resulting in 'Brexit', and the election of Donal Trump in the United States.[3] The technique known as 'micro-targeting', which has been around since the late 90s, involves crafting specific political messages tailored to ever smaller sub-sections of people, algorithmically defined through seemingly inconsequential qualities like 'owning a pickup truck' or 'watching X-Factor'. In the 2016 election, companies like Cambridge Analytica claimed to have taken these techniques to the next level by using psychological profiling, trained on social media activity, to further tailor campaign messages the hidden hopes and fears of voters.

Yet as journalist Sasha Issenberg (2012) recently remarked, there is no conceivable way to attribute the success of political campaigns to technologies, particularly savvy experts or even tectonic shifts in the demographics of the electorate. So long as voting is a private act, then such post facto analyses will be conjecture at best. More recent revelations, relating to the abuses of these analytics companies have further brought into question their more hyperbolic claims.[4] The more interesting question to ask is – what do these hyperbolic claims do in terms of redistributing resources and responsibilities from traditional polling experts to data science and what are the implications of these shifts for politics.[5] In advance of the workshop, we mapped out some of the infrastructure required for large-scale voter micro-targeting (below)



The key ingredient here is that data about online behaviour, which is standard in market research, is then linked up to names in a voter registry and political messages are tested

[3] For example: https://motherboard.vice.com/en_us/article/mg9vvn/how-our-likes-helped-trump-win

[4] Cathy O'Neil 'Trump's 'Secret Sauce' is Just More Ketchup' https://www.bloomberg.com/view/articles/2017-02-01/trump-s-secret-sauce-is-just-more-ketchup

[5] Barocas (2012) talks about how micro-targeting techniques have the potential to destroy the sense of a common conversation: candidates are not required to stay on message or speak to a fictional 'center' if they can speak to voters individually. https://dl.acm.org/citation.cfm?id=2389671

iteratively in order to refine models of voter behaviour (Anstead, 2017; Kreiss, 2012). So many of the more interesting questions about the role of data analytics in politics are contingent on first identifying the different elements of this infrastructure – which companies are involved in collecting, analysing and selling data for political ends? This proved difficult to determine because these infrastructures are by design, hidden, due to negative perceptions of these tactics.

As we will return to later, these secretive aspects of the topic and our analytic distance from the field (few of us were experts on the topic) may have inclined us to ask certain sorts of questions as opposed to others.

Election Spending

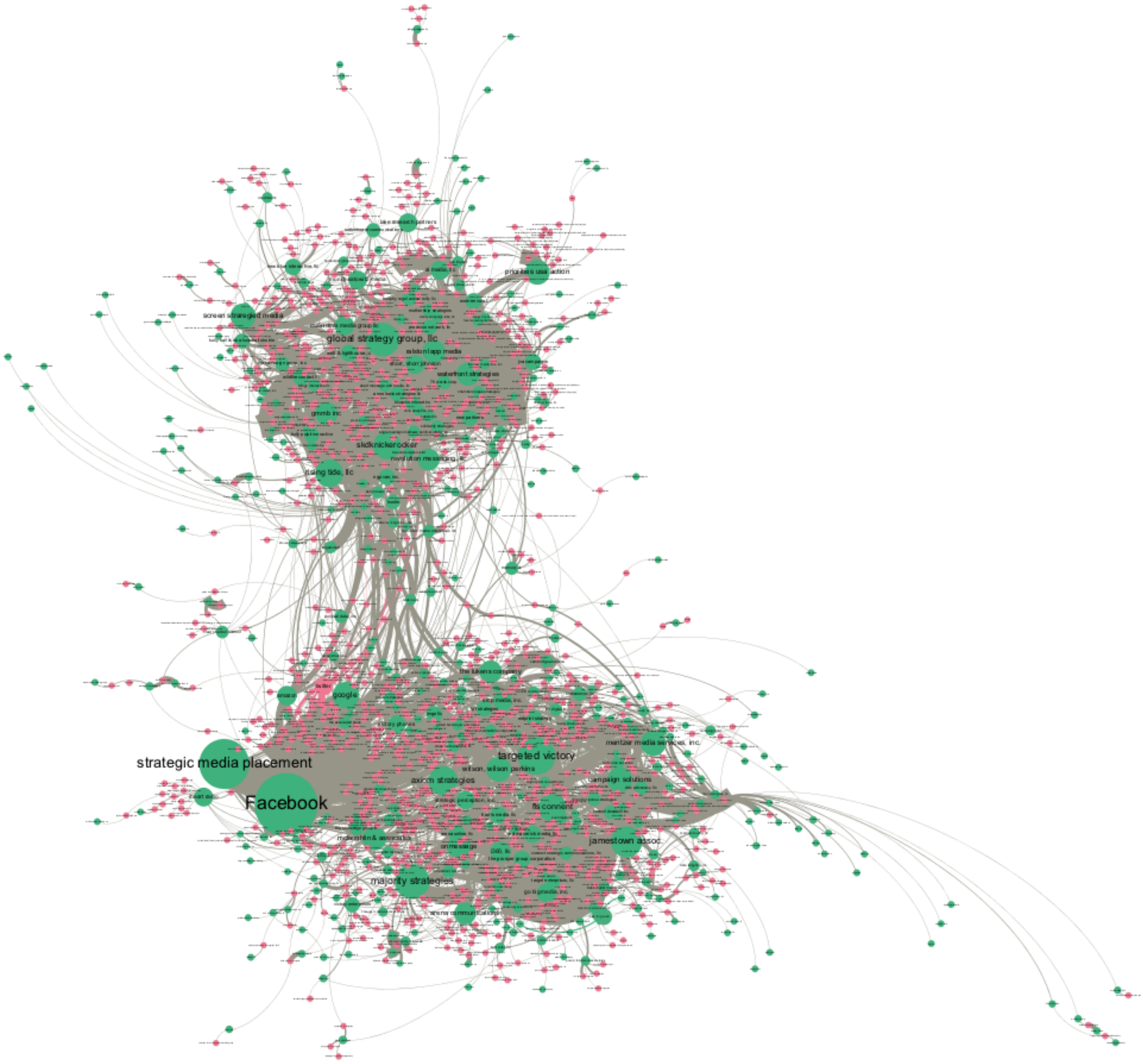
One group started with a particularly compelling data source: the electoral registers for the United States and the UK. These are public databases which list expenditures by political campaigns and their proxies (I.e. SuperPacs) in a given election. These sorts of open government data are made available to all in the name of transparency but as with many such initiatives, they also limit access in subtle ways (Tkacz, 2014). For example, they restrict the amount of records one can download at a time, and because the bandwidth is slow, users are forced to use targeted searches or small date ranges.

The participants started with a few arbitrary choices, limiting our search to individual expenditures over $1000, and disbursements over $10,000 for the US and a similar level for the UK. We also limited the records to the years 2013-2016 so we could focus on the 2016 election and EU referendum. This raised interesting questions like: how long does a campaign work in advance of an election or what size expenditures are most interesting? Bigger expenditures would seem to indicate greater involvement in campaign activities, but lots of smaller expenditures might be used to conceal shady purchases. It was also pointed out that any truly illegal activities would be kept off the books entirely.[6]

If these two lists were combined, we could represent them as a bi-partite network diagram (a network with two types of nodes) connecting 'payers' (political campaigns or more likely shell companies) and their 'payees' (various suppliers, consultants and services including data analysis and targeted advertising). Combining the UK and US lists was relatively easy because they could just line up the columns (ignoring for the time the difference between USD and GBP) but there was another problem: though this data was extremely well formatted, it was 'messy', being collated from a host of different organisations and sometimes converted from paper submissions. Through keyword searchers, the group realised that the company 'Facebook' was spelled 23 different ways (Facebook, Face Book, Facebook, inc. etc.). Making a network diagram inclined them to resolve these alternative spellings into singular entities. But what counts as an entity and where does the entity stop? Is it sufficient to combine alternate spellings of the company? What about including subsidiary companies as well?

[6] In the UK, it has been claimed that campaigns have hidden spending either in less regulated local elections or in overseas companies: http://www.cbc.ca/news/canada/aggregateiq-aiq-brexit-vote-leave-beleave-whistleblower-1.4592056.

This required some manual work to decide which entities should be combined and this took the participants the better part of the second day. The resulting list was then converted into a network and, using a free program called Gephi, the network was visualised with a spatialisation algorithm which arranges the network in two dimensional space so that entities with more mutual connections are brought closer together.

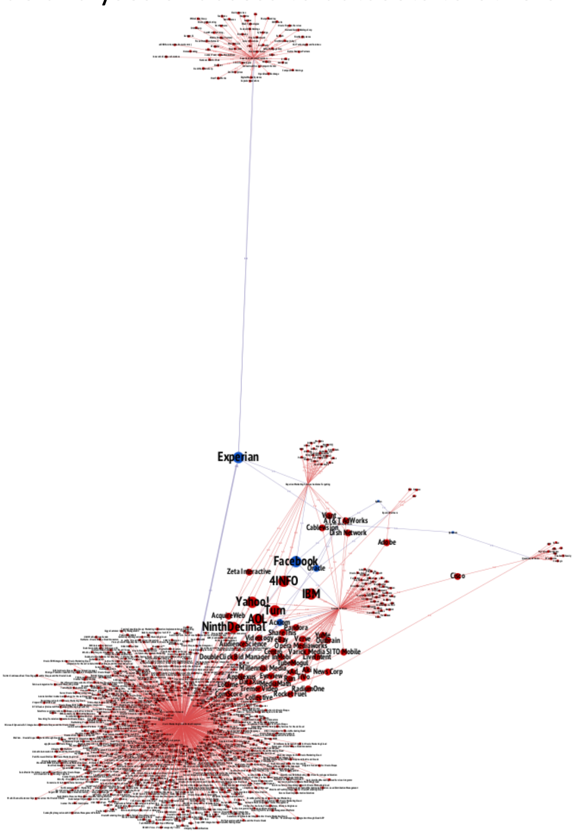


Payers (in red) and Payees (in green) in the UK and US election cycle 2014-6. Payees are sized by number of connections in this version of the graph.

To the average reader, the galaxy of coloured dots above might look imposing or confusing, but to the participants in the room, labouring with the data and patiently waiting for results, the result was exciting. Researchers who frequently use these graphs often hope that the graph will show distinct clusters, which might reveal possible tensions or more-or-less coherent groups which one could interrogate. Compared to some of the blobs we were looking at during the workshop, this network had a definite shape and structure to it. Through checking the contents of the clusters in the database, the group found that the two major clusters did not correspond to US and UK as one might expect; the top cluster seemed to consist of mostly US Democratic party candidates and organisations and their payees, while the bottom cluster seemed to contain the US republican party and several of the tech giants (Facebook, Google etc.) and most major UK payers. The tech companies are the main bridge between the two because they work across the political spectrum and internationally. However as Anders Koed Madsen, one of the participants noted: 'these are the clusters according to *this* algorithm…', at which point he clicked through several settings and windows, displaying different configurations of the network, complicating the seemingly clear 'finding'. It is helpful when considering these moments to get away from an easy language of 'discovery', 'findings' or 'revealing' which downplay the *work* that goes into making something self-evidently present, 'just there' (Coopmans, 2014; Goodwin, 1994; Lynch, 1988). Sometimes our intuitions, the data, and a particular graph with its multitude of settings, harmonize in such a way that they appear to 'tell a story' (or allow us to tell a story to each other).[7] Essential to this story is how some 'findings' are positioned against what one might 'expect' to find – even though these expectations might not have been consciously apparent to anyone before the fact. What is interesting is that once this relationship is established, it becomes hard to see other possible stories.

Facebook's Partners

Another group came to the workshop with an already well-formed approach to the topic. Anne Helmond and Fernando van der Vlist collected a list of Facebook, Twitter and other social networking sites' 'approved partners' to create a database of companies which could, in theory, have access to vast stores of social media data, and thus act as 'data brokers' who could provide analyses or access to datasets to others.



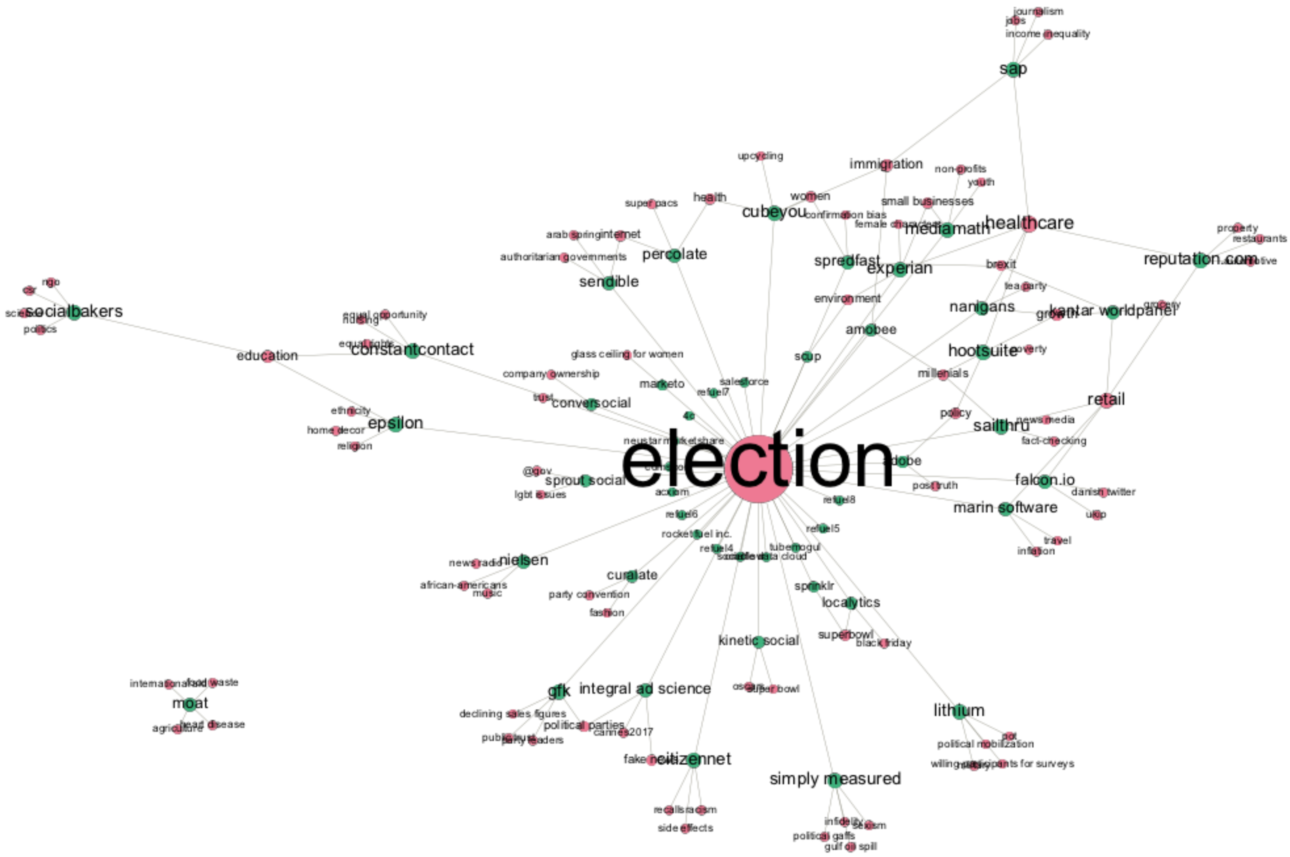
Partner networks for Facebook's marketing data partners.

[7] Thinking of these graphs as helping us tell stories draws attention to the fact that these graphs are being analysed qualitatively and interpretively and resists the stereotype that qualitative methods are about stories while quantitative methods are about numbers.

Above we see another imposing network diagram, this time of Facebook's partners and partners of the partners. A halo of these companies encircles Facebook in the lower part of the graph while to the upper right of the largest cluster are partners with several partners of their own. Experian in the middle, connects through a subsidiary to a swarm of other partners at the top. But which of these companies might use this data or insights from it to sell to political campaigns? The team started by searching their database for partner companies (mostly market researchers) which advertised themselves as specialising in political topics. They also manually cross-referenced their list with the payees in the campaign spending database. The group found that many Facebook partners were selling services to political campaigns but also a tremendous amount of money was also going to Facebook directly. $34 million was also spent on i360, a company claiming to have a database of +18 consumers and voters.

Another sub-group started investigating the way that these politically inclined marketers construed 'politics'. For example, some marketers and data brokers professed to have specialisms in 'gun control' or 'abortion' or 'education policy'. Starting with the list of Facebook partners, they queried the text of the partners websites for the terms 'politics' and 'political' to obtain a list of companies claiming to specialise in political topics. They, then manually pruned the results to focus on those pages with the most obvious political ties and then for each of them gathered 5 political issues they claimed to specialise in.

These issues, which were phrased in a variety of ways, where then manually grouped under common headings like 'education' and represented as a network of partners and their issues.



Network of companies (green) and issues (red)

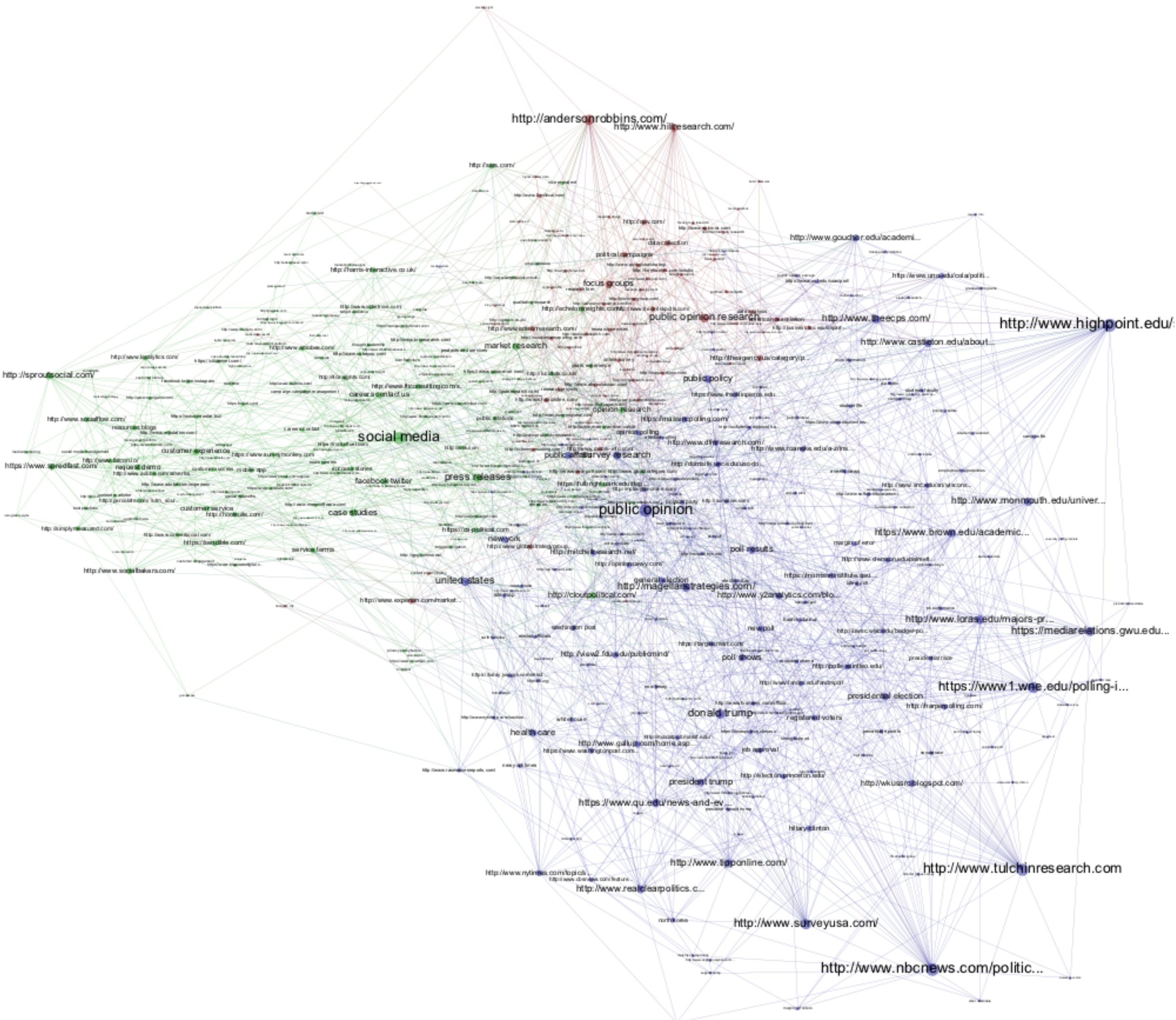
Although the group ran out of time, such a network could be used to help a (manual) textual analysis of how these companies construe the object of politics. In contrast to the way the first network seemed to 'speak for itself', this map does not seem to have clear clusters or focal points (except for election) but it can be used like a map of an unknown territory. It might pique our curiosity about certain partner's websites or suggest, but not over-determine, certain avenues for browsing through the data.

Micro-Targeting the Micro-Targeters

Another group looked at the same companies from a different angle, to see if any computational processes could spot patterns within the behaviours of the very companies using these techniques. This once again relied on making lists of companies. In addition to the self-identifying political marketers identified in the previous step, this group decided to incorporate other competing definitions of the field from different sources. This included a manually collected list of companies which had appeared in press coverage; a search using company database Factiva (linked to Lexus Nexus) and Angle.co.list, a list of start-ups for potential investors. They contrasted these companies with industry lists of 'traditional pollsters', from popular podcast, the pollsters (US) and the British Polling Council (UK).

The group then extracted the full text of the front pages of these company websites to see if anything could be gleaned about the differences between the pollsters and micro-targeters, or differences within each group, in terms of the language they use. In effect, they wanted to micro-target the micro-targeters, in a somewhat playful way, with much less sophisticated computational techniques. Co-word analysis is a technique developed in STS for helping to understand transformations in scientific fields though corpuses of scientific abstracts. Key-words in these abstracts, which might be scientists, institutions, natural phenomena or equipment are seen as linked when the occur together in abstracts; the more times they occur, the more strongly associated they are deemed to be (Callon et al., 1986). These can then be represented as a network of connected words in which the lines connecting them become thicker the more they co-occur.

Using a programme called Cortext, the group produced a network of marketer's webpages and key words occurring in them (extracted using Natural Language Processing algorithms): allowing the group to spot which marketers used similar language.



Co-word of companies and key terms, processed by Cortext, coloured by cluster

This graph was somewhat disappointing because the spatialisation algorithm did not yield obviously distinct clusters. We ran another algorithm called 'modularity' in the Gephi interface which, put simply, *forces* the program to find clusters in the network and identifies them with colour coding. The two main clusters, as determined by the algorithm, seemed to correspond to pollsters and micro-targeters, but they were far more inter-mingled than one might have guessed. Also, a third cluster, between them in red, seemed to correspond to neither group. It is important to remember that co-word was developed for scientific abstracts, which are well formatted, written in a fairly uniform style and relatively short – there is no guarantee that it will show the same sorts of things when applied to different empirical material, in this case, promotional website texts.

Discussion

This workshop was unique in two ways. Firstly, the participants were very adept at using these tools, particularly network graphs, so we more-or-less leapt into the task without much discussion. The participants were also all accustomed to working with web data and particularly social media data for which there were established research protocols, but less so with spending registers and company websites. This meant that we were confronted by fundamental methodological questions which are not always present in social media discussions.

For example, Aaron Cicourel (1964) long ago argued that the very possibility of measurement in the social sciences is predicated on agreed upon or banal concepts, which are themselves part of the phenomena being studied. Quantitative surveys but also qualitative interviews are premised on everyday understandings of how people talk, or what words mean to people, which are by no means clear in any given interaction. In STS analyses of social media data, this problem is often addressed by, whenever possible, relying on 'members terms' and categorisations which exist in the field. We might try to use a particular organisation's demarcation of 'politics', rather than our own, or make graphs using 'hashtags', which are user created. But in this workshop we were forced more often, it seemed, to make banal decisions about what counts as a singular company or how companies identify themselves as being relevant to politics. For this reason, there was an interesting relation between the graphs and our assumptions. All our efforts could be boiled down to tacit hypotheses: that there are discernible patterns in spending between campaigns or that different types of organisations talk in different ways. So, on one hand we needed the graphs to confirm some aspects of our tacit understandings of the field – *they needed to be matched to the image of the data in our heads* – but at the same time we were only excited by the graphs when they in some way went against our expectations. We were disappointed to find, for example, that pollsters and micro-targeters appear to express themselves in fairly homogenous ways. Yet many of these assumptions are only materialised *as a consequence of* making the graph and remain unspoken in advance.

Secondly, the topic area was marked by secrecy and proprietary knowledge and thus our attention became directed toward making visible phenomena which were, not hidden per se, but jumbled and inaccessible. Ironically, the abundance of data creates a sense of lack or 'casts shadows' (Leonelli et al., 2017). Because of this distance from the topic, rather than asking traditional sociological questions about how and why, we became limited to more basic questions like what and who?[8] Who were the companies intervening in this area and what were their claims? We became more like journalists or detectives looking for a 'smoking gun'. Only near the end did we get the opportunity to analyse the text of their websites and consider how these companies position themselves and the implications of this positioning for the campaign industry and politics in general.

However, while this workshop in certain ways yielded less sociologically rich material, these graphs and data sets laid some important groundwork for future investigations. Rather than endpoints, they became starting points, for further qualitative work. It also yielded visualisations and databases which might be more straightforwardly useful to informants.[9]

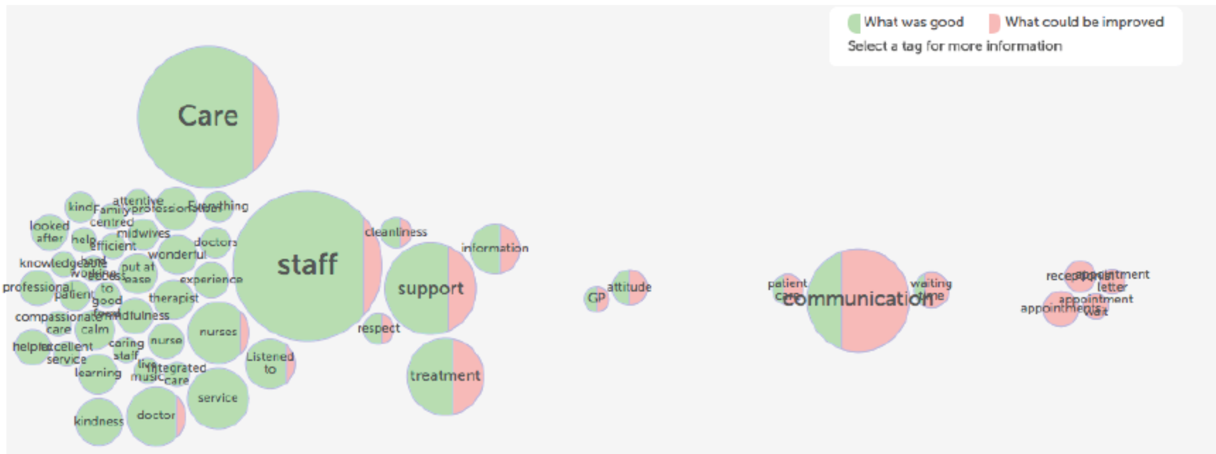
[8] It has interestingly been argued that so called maco-analyses are not directed at an ontologically separate layer of society but merely a consequence of analytic distance and detachment (Knorr-Cetina, 1981).

[9] We are currently working on cleaning up our analyses so that it can be shared with journalists currently investigating potential abuses by tech companies in recent elections.

Workshop 2: Big Health Data

The second workshop also focused on an industry starting to wrangle with the possibilities and pitfalls of big data. Machine learning and artificial intelligence are increasingly being used in both private and nationalised health services. To give one famous example, Google's artificial intelligence subsidiary DeepMind was contracted by the Royal Free Trust in the UK's National Health Service to see if their algorithms could be used to predict instances of acute kidney injury. Although they had some success with a similar process for processing scans of eyes to detect various problems, this initiative was derailed when it was revealed that Deepmind was given access to much more data than originally reported (5 years worth of data on every patient).[10]

A somewhat different example of 'big' health data is online patient feedback. Various forms of feedback are a crucial tool of nationalised healthcare systems for identifying problems and potentially distributing resources. However, as Farzana Dudwalla of Oxford University's Inquire project pointed out, there is a contradiction between the needs of patients to give feedback anonymously and without fear of repercussions, and the doctors' and healthcare providers' need to assign specific feedback claims to specific hospitals. One company addressing this problem is Care Opinions, a popular website (separate from but linked to the NHS), which solicits feedback from patients which are anonymised[11] but also linked to specific Trusts (collections of hospitals), though not specific hospitals or wards. These 'stories', as they are called, are searchable through their public interface, but professionals in healthcare can perform more complex filters, create alerts for new stories and visualise data sets in various ways. The below visualisation, available to healthcare providers in the backend, uses story 'tags' and a form of sentiment analysis (discussed later) to determine the percentage of 'positive' and 'negative' stories containing each tag.



Story tags sized by number and coloured by proportion of negative and positive stories

We were kindly provided access to the backend of Care Opinions' database by the owners,

---

[10] See for example: https://www.newscientist.com/article/2086454-revealed-google-ai-has-access-to-huge-haul-of-nhs-patient-data/

[11] Anonymising data, in the minimal sense of removing or obscuring names, must also be accompanied in cases like these of attempts to remove other incidental details, names, places and events which could be used to re-identify individual later.

giving us a repository of stories from two different websites with slightly different formatting. Each story had a created and published data, sometimes the administrative location (the hospital trust), whether or not there were responses and the text of those responses.

Researchers in the Oxford project, which included social scientists, doctors and nurses, had started to experiment with machine learning to process feedback, to potentially group it by topic and identify anomalies. Many qualitative researchers are sceptical of using machine learning on textual data because texts are still one of the hardest types of data to process automatically. Forms of sentiment analysis, like the above, still have yet to master basic problems like 'sarcasm', let alone extract subtle meaning-making and subtext. Also, many automated techniques side-step questions about what feedback is for. Is it for eliciting specific bureaucratic changes; or wider changes to the professional culture of medicine; eliciting pity; or a cathartic howling into the void (or combinations of the above)?

For the workshop, we hoped to reverse this process: rather than replacing qualitative reading with automated reading, how could automated techniques, including visualisations, be used to foreground qualitative accounts or assist qualitative analysis?

Feedback Responsibilities

It is often the case with automated forms of textual analysis that the concern is with the content of what is being said. This is the case for 'topic modelling' which is a computational process for extracting clusters of words which appear near each other in sentences. The first group started by considering not just what these feedback stories are about but what they accomplish as texts. They were inspired by a famous paper by Dorothy Smith (1978) which analyses the text of an interview in which a young women describes how she knew that her friend was mentally ill. Smith shows how the text makes the diagnosis of mental illness readable as a *fait accompli*.

In the case of patient feedback, then, we might say that what counts as 'good' healthcare, something that we might take for granted as patients ourselves, is made available through the structure of a story: through what events are foregrounded and how they are presented, including which actors are positioned as responsible in the story. Smith also talks about the use of 'contrast structures': setting up a pair of events so that one is presented as 'what is expected' so that the reader sees the second event as deviating from this.[12]

But how could they do Dorothy Smith by (semi)-automated means? One of the starting points was the identification of the 'cast of characters' mobilised through the text. It is a relatively easy task for Natural Language Processing software to extract nouns and proper names which could be grouped manually into recurring characters, something they attempted manually below.

Anyone from my local CMHT

[12] One hypothesis of the group was that negative feedback would have more contrast structures, than positive feedback.

Psychiatrist
GP
Me
'Anyone'
'no-one'
Professionals
Staff member(s)
Porter
Nurses
Member of staff
Technicians
Doctors
(my) Mother
Patient
Someone I love
'Robot-like' radiographer
People and Families

This was an interesting exercise which could be used to identify recurring words like 'mother' to extract lists of stories containing this 'character'. Yet this brought up the dilemma mentioned in the previous workshop: are we presuming to know how people refer to their mothers? What about different types of mothers? Also, how an entity is mobilised and how responsibilities are assigned to them are only perceptible within the text as a whole.[13] Here is one example story:

> Forts of all I would like too say my review today is nothing to do with the staff as they were brilliant when we eventually got seen. My 1 year old child had taken a fairy non bio wash tablet out of my washing machine and popped it in his face after rinsing his face for around 10 minutes I rang 111 on the way too the hospital the receptionist informed me he would need to be seen within the hour as we arrived at the hospital I was told to wait in line and I would be booked in soon as I got to the black of the line the lady infrount of me told be she had been waiting 30 minutes to be even booked in with her child which simply is unacceptable these are children in pain yet nothing was done my baby's eyes were getting worse and worse I called a nurse over and asked too be seen to be told yet again that I would have to wait I then proceeded too get extremely upset about this so I was then seen then to be told oh the staff on 111 are not medically trained but I am pretty sure the doctor at 111 conferred with is medically trained so after this I was sent to wait in the children's waiting room where I continued to wait for over a hour my sons discomfort increasing quite quickly too which I called a nurse yet again then to be told with a room full of children there is only one pediatrician and they are currently with a child in recuse how is it possible in the whole hospital there is only one paediatric doctor so I continued to wait by the 2 hour mark my son is now crying uncontrollably in pain latterly screaming in distress and vomiting still no one too see him after just over 2 hours a man was brought in by police in hand cuffs next to the child's waiting area but that's not the worse thing he was then seen straight away ! How is it that this man was seen before a room full of children I was disgusted with this so yet again called another nurse who then said we should be seen soon so after around 3 hours wait the my baby in extremely bad pain we were then seen this is the worst experience in a hospital I have ever had under staffed and over worked

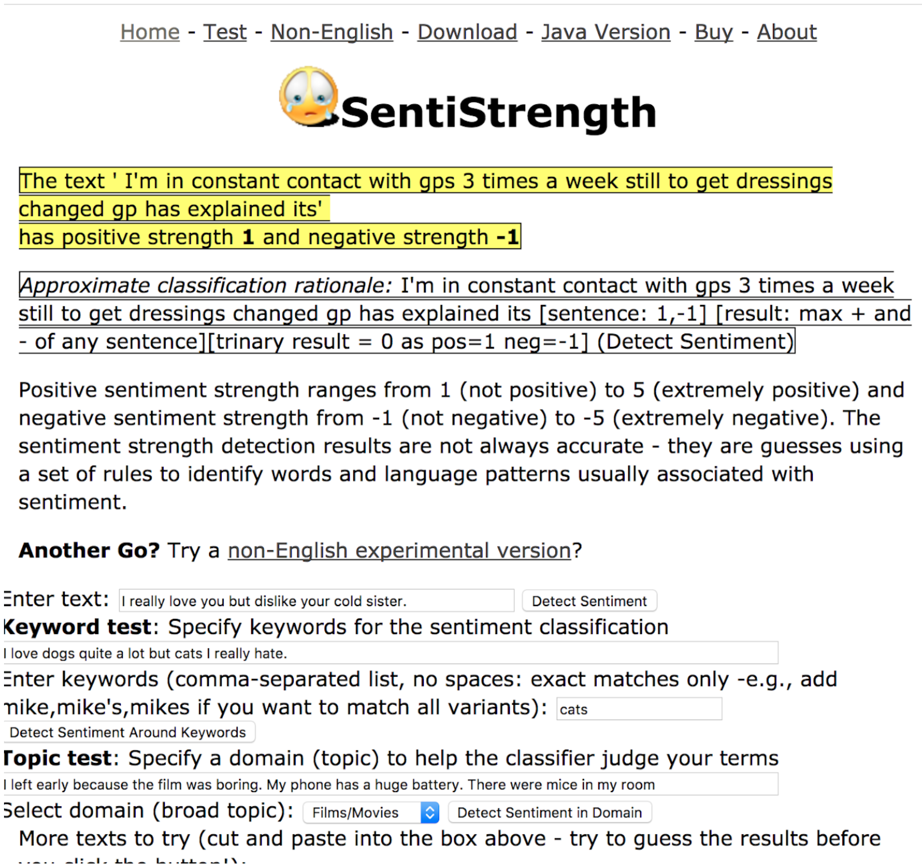
The first sentence already sets up the negative valence of the account by absolving the staff of responsibility, sensitising the reader to view the actors that follow as potentially culpable. The contrast between children, frequently evoked, and others, particularly 'the man brought in by police' presents us with a moral order, presented as self-evident, that a room

[13] Some scholars would argue that the text only becomes sensible with reference to the actual situations being described or any number of other texts (Knorr-Cetina, 1981).

full of children should be seen before men brought in by police. Although some responsibility is assigned to the nurse deciding who is seen (wrongly in this woman's eyes), most readers would understand the real problem is the 'under staffed and over worked' hospital with 'one paediatric doctor'. The point is that while we can detect actors and their arrangement in the text automatically, we can only see what the story is doing by taking the story in as a whole.

When we zoomed out and started thinking like a computer, this plan to detect 'responsibilities' and 'characters' seemed reasonable enough but when we analysed an individual story, the complexity of the text and the social world it references was overwhelming. Part way through day two of the workshop, the group was also getting a craving for visualizations: the other group had been using the projector and producing all sorts of colourful maps (see below) and the group wanted the endorphin rush that comes with seeing an, even faulty, panoramic view (Latour, 2005).

The group decided to switch tack: instead of trying to approximate qualitative readings in a quantitative way or scaling up qualitative work, they tried starting with a quantitative approach and modifying it. We took the example of sentiment analysis, mentioned earlier, which Care Opinions already used in the backend of their website.



Sentistrength Website http://sentistrength.wlv.ac.uk/

Sentiment analysis, in its simplest form, works using a library of words that are inherently deemed to be positive or negative (ranging from -5 to +5). The words in a sentence are added up to produce a sentence score (taking account of basic modifiers like 'not ____' and other rules). This approach of course ignores most relationships between words or between sentences, and as mentioned earlier, sarcasm. 'Murder' is an extremely negative word even if the sentence was 'I could murder a pizza'. The object then became to come up with a system which mirrored Senti-Strength but improved on it by making it either more sensitive, more nuanced or more targeted to the specific problem of analysing patient feedback.

What they proposed was that entities assigned responsibilities in the text could be conceived of on a spectrum of more specific to more general. A 'nurse' or 'the nurse' was more specific than 'the staff', 'the hospital', 'the NHS' or the practice of medicine in general – and this has very different implications for how responsibility is being distributed. They made a brief library of common nouns and pronouns and then, somewhat arbitrarily, assigned them 'generalised responsibility scores' (from 10 – 0 because it was deemed that one couldn't get more specific than an individual). Knowing how to rank the 'generality' of words is of course also a background assumption which we might consider to be achieved through textual utterances, but such a provisional metric is still a more compelling way of sorting texts than whether they are positive or negative.

Collectives		**Individuals**
The hospital	(6)	nurse
The ward	(3)	receptionist
No-one	(10)	Doctor
The department	(4)	GP
Centre	(5)	Secretary
suite	(3)	Worker
staff	(2)	consultant

To begin with, they tested this with the frequently occurring terms 'nurse' and 'the hospital' automatically assigning a value for each story based on the number of appearances of each multiplied by their scores.

Switchcoard staff	The switchboard staff are rude and unhelpful - I rang 3 times to speak to a general surgery secretary and was cut off 3 times after	0	0	203	-2030	08/09/2017 12:38:53	08/09/2017 14:3
Stay away	went in to see doctor. doctor was 45 minutes late, eventually when they called me in they rushed the appointment and said a	0	0	68	-680	08/09/2017 12:33:20	08/09/2017 14:3
Excellent	My physiotherapist was extremely helpful and in only two visits and lots of advice on exercise movements	0	0	0	0	08/09/2017 12:03:27	08/09/2017 14:3
Some concerns	I was discharged and my Care coordinator was not notified, it was only that I had told them myself that they knew. Not a safe or	0	0	106	-1060	08/09/2017 12:25:42	08/09/2017 14:3
Day surgery - excellent care	My son stayed in the hospital for day surgery to repair a perforated eardrum. All the staff without exception were professional, kind	0	0	0	0	06/09/2017 22:49:49	08/09/2017 15:3
St Helens dermatology	I have been going to St Helens Dermatology now since 2009 where i was diagnosed with a serious condition called reoccurring	0	0	448	-4480	08/09/2017 01:37:59	08/09/2017 15:3
bleeding in the stomach	We've visit Sheffield from Corfu, Greece to be at the wedding of some friends, through Manchester airport.	0	0	0	0	08/09/2017 09:37:12	08/09/2017 15:3
Does a good job in challenging	Admitted to A&E late Monday afternoon with suspected facial cellulutis. Waiting and triage area crowded and far too small. Once	0	0	0	0	08/09/2017 09:26:27	08/09/2017 15:3
Brief stay in Ward 14	I was admitted to Ward 14 to assess a problem with stomach pains. My GP examined me in the morning and thought I might as a	0	0	0	0	08/09/2017 09:32:07	08/09/2017 15:3
Disgusting service	. My son was admitted to ward 21. Firstly the doctor already made a diagnostic when we got there but they wanted us to stay over	0	0	102	-1020	08/09/2017 09:45:12	08/09/2017 15:3
Bury Integrated MSK Service	My appointment was for an injection for carpal tunnel. the appointment was on time and the person my appointment was with was	0	0	162	-1620	08/09/2017 08:45:28	08/09/2017 15:3
Brilliant cataract surgery	It takes some effort to comply with Benenden's requirement that you should go fully through the NHS referral system before they	0	0	124	-1240	08/09/2017 10:05:44	08/09/2017 15:3
The breast unit	I have nothing but praise for all the staff	0	0	0	0	08/09/2017 10:49:35	08/09/2017 15:3
A&E visit	Went because of twisted knee. Was in waiting room for just 10 minutes before being seen by a very helpful and kind triage nurse. I	0	0	0	0	08/09/2017 09:45:40	08/09/2017 15:3
Financial waste	What is the point of sending a letter for a primary out patient appontment without allowing for processing and postage times?	0	0	397	-3970	08/09/2017 10:59:55	08/09/2017 15:3
Replacement knee fantastic	13 weeks on and knee is amazing. Staff courtesy, care and attitude is excellent throughout, and makes you feel as comfortable as	0	0	0	0	08/09/2017 10:07:32	08/09/2017 15:3
SEAL Unit Arrowe Park	I attended the SEAL unit for a gallbladder removal on Monday 4th September, unfortunately due to surgeon illness this had to be	659	0	0	6590	08/09/2017 11:02:45	08/09/2017 15:3
Ct scan 7/9/17	Today I had an appointment for a ct scan . I was called thro to the waiting area early for my appointment.	0	0	0	0	08/09/2017 09:29:27	08/09/2017 15:3
Day surgery tcre/trcf	Lovely staff, nice surroundings and caring medical team. I also felt fully informed throughout	0	0	0	0	08/09/2017 11:02:51	08/09/2017 15:3
Double hernia	From consultation to completion, I was handled with curtesy and respect. At all times I was kept fully in the picture, without me	0	0	0	0	08/09/2017 11:19:19	08/09/2017 15:3
Efficient and Courteous Care	The service given to my mother from the paramedics, A/E and then Cedar Ward was outstanding. From dialling 999 at her home to	0	0	0	0	08/09/2017 12:01:24	08/09/2017 15:3
Day case laparoscopy	Staff were very friendly helpful and respectful at all times.	0	0	0	0	08/09/2017 11:09:02	08/09/2017 15:3
Fantastic experience of hip	After quite a frustrating drawn out process of getting a final recommendation for a hip replacement through the bureaucratic NHS	360	0	1603	-12430	08/09/2017 12:02:05	08/09/2017 15:3
endoscopy upper GI team	The buses and care staff was very help full and friendly. They explained everything in a way it was easy to understand and left me	245	0	59	1860	08/09/2017 11:26:08	08/09/2017 15:3
Short stay after op	From the moment I arrived to the moment I left I recieved an exceptional service.	903	0	0	9030	08/09/2017 11:10:49	08/09/2017 15:3
Do not bother phoning	I have been trying to book an appointment for half a month, no answer whatsoever. Great, I need my coil removed in the next 2	0	0	570	-5700	08/09/2017 11:26:03	08/09/2017 15:3
Mental health	My partner was admitted to the Taylor psychiatric ward in Tameside on the 1st of June 2017, a week or so later she was moved to	0	0	878	-8780	08/09/2017 10:13:48	08/09/2017 15:3
They couldn't care less	I was referred to this place about a year ago by my GP. Nobody seems to listen or to care about the patient's needs. Most of the	0	0	0	0	08/09/2017 15:15:15	08/09/2017 15:3
The Cedars Centre,My Perspective :)	Ok,I've been a service user at The Cedars Centre now,for a very long time indeed,as I'm a chronic depressive,with associated	0	0	2029	-20290	08/09/2017 14:55:41	08/09/2017 15:3
Is that good enough?	shocked at the lack of communication have been waiting months for tests on my hiatus hernia to see if it can be repaired. I feel	0	0	0	0	08/09/2017 16:11:31	11/09/2017 11:1

Stories with cumulative generalised responsibility scores

This method, if properly implemented, with a more carefully thought-through list, could be used to make scatterplots of the stories and identify stories with extreme patterns of entity usage: many generalised entities, many specific ones or peculiar combinations of both.

Milligrams and Angry Nurses

The second group, was also interested in generality and specificity but in a somewhat different way. They identified a potential tension between stories that perform detached expertise and those that lay bare emotions. While this might seem like an odd pairing, the ways in which laypeople are written off as having emotional responses while scientific experts are seen as purely rational beings, has been a longstanding topic in STS (Wynne, 2011). The question became, do patients who perform expertise as opposed to performing emotionality, get more or less responses and how did these responses differ? This group was also interested in feedback as a process, not just the individual texts but the texts as part of a series of interactions, which might include multiple responses.

For expertise, the group started with an interesting hypothesis: accounts which used measurements might appear more authoritative and specific. Stories which used numbers and specifically 'mg', (milligrams) would be referring to highly specific issues about dosage. There are plenty of other ways in which expertise could be performed through an endless list of medical terminology but numbers and recurring words like mg had the advantage of being easily traceable. For emotions, seemingly obvious words like 'angry', 'feelings', 'felt' etc., were tried but interestingly, one key word which seemed to crop up through qualitative analysis was 'nurses'. Nurses, rather than doctors or other types of staff members, seemed to be the locus of highly charged, mostly angry reactions, something which chimed with the participants' own experiences in the field.[14]

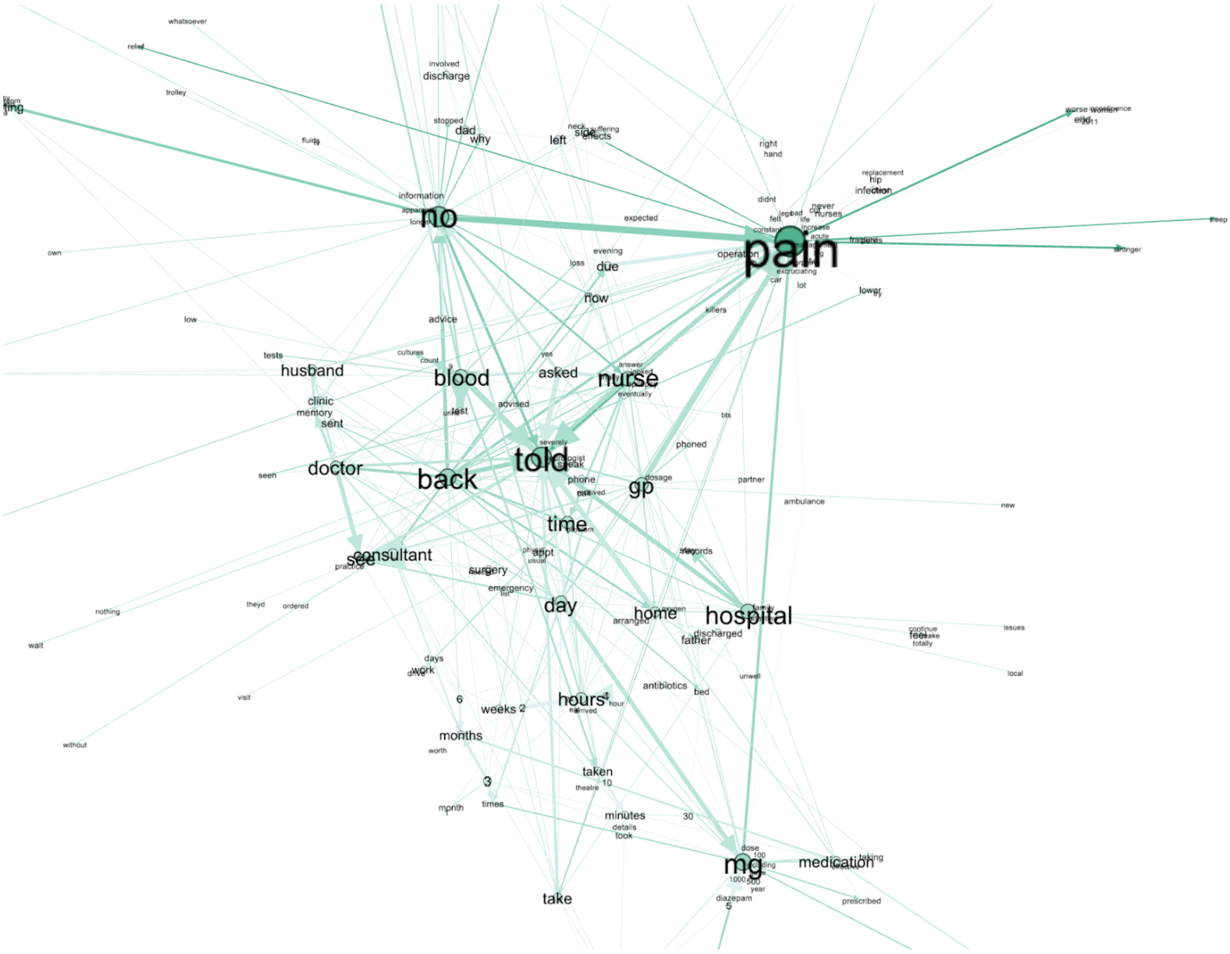
It should be noted that these two groupings are not in any sense mutually exclusive, but these seemingly arbitrary queries seemed to hang together as the analysis progressed: they consistently yielded the sorts of stories *the group was expecting to find*.

	'mg'	**'angry' and 'nurse'**
	47 posts	35 posts
Number of responses	30 posts (63%)	21 posts (60%)
Total length	19 000 words 400 words/post	11 600 words 330 words/post

Stories in our dataset with 'mg' and 'angry' + 'nurse'

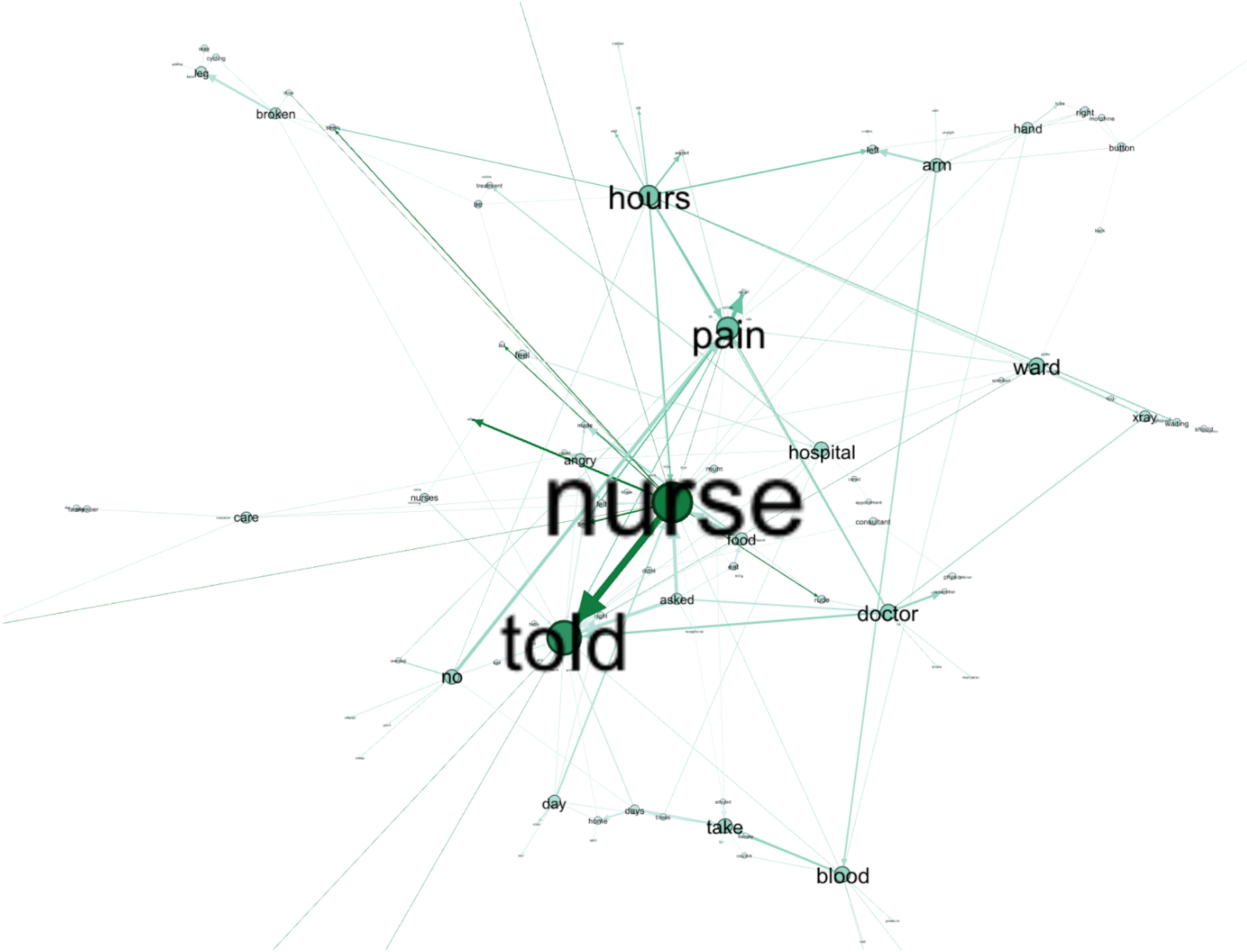
In the 'mg' group of stories, there were some very detailed accounts mentioning particular procedures, dosages, waiting times etc. and in the 'angry nurses' there were more obvious displays of what might be read as emotional language. For each of these two corpuses of documents the group used co-word analysis (this time using a program called Wordij) to produce a network of words which appear next to each other, within a particular distance of words on either side. The texts are first prepared by removing overly frequent works like 'a' 'the' etc. using a generic stop list.

[14] Also, as Teun Zuiderent-Jerak pointed out, for patients not well versed in hospital hierarchies, every staff member from receptionists to anesthetists might be perceived as a nurse.



‘Mg’ - Wordij network word distance 5.

In the first version of this network, the word ‘NOT’ completely dominated the graph because it appeared connected to almost every other word. Deleting nodes which are too common or not common enough is often standard procedure with network diagrams – but this is done in the service of making them legible, rather than for analytic reasons. It is tempting to read into the centrality of ‘not’ or indeed ‘no’ as indicating negative sentiment or something like contrast structures but it is impossible to say without delving into the texts.

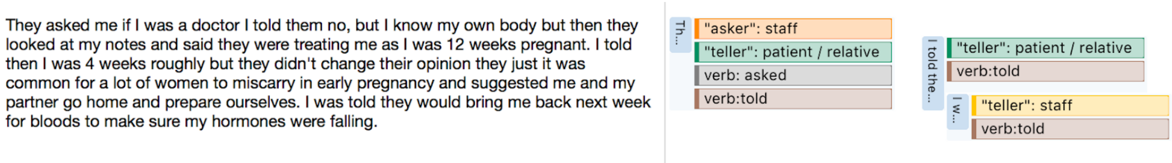


‘Angry’ and ‘nurse’ Wordij network

Looking at the two networks together we might note lots of similar words like pain, hospital and doctors but there are also other surprises like the frequency of hours and other measures of time. Often it was not so much what happened but how slowly or fast it happened. In terms of the differences it might be said that the ‘angry nurses’ network is dominated by recurring formulations of ‘nurse’, ‘asked’ and ‘told’. This became a compelling point of focus perhaps because ‘asked’ and ‘told’ were not part of the query nor were they obvious aspects of their mental image of emotionality.

This surprise ‘discovery’ led to a follow up question: who does the asking and who does the telling? One of the participants extracted stories which contained variations on asked and told and manually coded them for which type of actor was the subject and object.

> ‘ I was however concerned when **she asked me** which antibiotic I took and what strength and how many mg of prednisolone I took a day.’ (‘asker’ staff)
>
> ‘ I said what dosage I was on, and **this doctor told me** to go back to my GP and say that my dosage was too high.’ (‘teller’ staff)



	'angry' & 'nurses'	'mg'	Total sentenes
'asker': patient / relative	**16**	**20**	**36**
'asker': staff	**10**	**15**	**25**
'teller': patient / relative	**13**	**4**	**17**
'teller': staff	**36**	**62**	**98**
Patients 'asker' & 'teller'	2	0	2
Patients 'asker' & staff 'teller'	4	4	8
Staff 'asker' & patients 'teller'	1	0	1
Staff 'asker' & 'teller'	2	2	4
Sentences	**65**	**95**	

Table of manual coding results: subject and object of asking and telling

It appeared that more often, the staff were 'telling' people in the mg stories and in the angry nurse stories, it was the patient doing the 'telling'.

In terms of responses, it was interesting that while there were many stories which included specific expert-like statements, these rarely received specific answers.

> *Thank you for taking the time to share your experience. I was very sorry to hear about the injury you sustained whilst playing football. I was also sorry to hear that the treatment you received whilst in accident and emergency was less than satisfactory*

The above response, for example, reads like a 'boilerplate' in which the specifics have merely been inserted into a standardised template. Those responses that do engage in specifics, and felt more tailored, tended to do so at the level of service delivery rather than the actual medicine being practiced: was it on time, pleasant, affordable, clear, rather than was it medically correct.

Discussion

This workshop was different from the first for a couple of reasons. Firstly, because the participants in this workshop tended to self-identify more as 'qualitative' researchers, and partly because of the breakneck speed of the last workshop, we tried to slow things down this time and focus more on the close reading of the texts. Second, the participants this time were far more embedded in the topic, having worked in or observed hospitals themselves, and had a greater awareness of lived experience on the ground and what was at stake. For whatever, reason, there was a noticeable slump in the afternoon of the second day in which a sense of palpable frustration hung in the air when we discussed our 'progress'.[15] One of our visualisation experts then remarked 'since time is short, shouldn't we focus on what we *can* measure'. One of the main objectives of these workshops was to constantly ask ourselves what we *should* measure or what is important to measure, but this move may necessarily strain relationships with programmers and scientists. Calls for a slow science (Stengers, 2011) are all well and good but they need to contend with the routines and pressures imposed on scientists who have even more frantic publishing regimes than our own.

In this workshop we ended up with two interesting strategies for working more constructively between quantitative and qualitative techniques. The first, we might refer to as an 'ironic' orientation to building tools – the goal was not to produce a digital approximation of qualitative work but to in some way critique an existing quantitative approach, by creating something better, rather than something perfect. The second was to use visualisations to generate surprises. What was interesting about the second group was that they used the visualisations as part of an iterative process. A seemingly arbitrary search query, 'milligrams' and 'angry nurses', yielded further collections of words which sparked further questions. These did not lead as in the case of triangulation or Grounded Theory type approaches to *inductively* converging on some underlying reality, but *abductively* generating new ideas, relationships and contrasts. However this strategy might only be possible when there are topic experts who really know the data.

[15] In the world of data sprints and hackathons, there is almost always a slump halfway through which turns to a manic and productive energy in the final few hours.

Workshop 3: Academic Rankings and Metrics

The third workshop shifted the focus to transformations in our own industry. Academia has for a very long time been producing data about itself, mainly through the cataloguing of journal articles, authors and citations. Indeed, academic data has long provided a testing ground for data-driven techniques because it offers (reasonably) well formatted data and stable data sets over time. For example, the technique of co-word analysis, discussed in an earlier post, was first made possible because of the standardisation of scientific abstracts. As we discovered in the other workshops, techniques developed for one field often struggle when imported to another.

Yet academia is no less susceptible to hype around data and has recently come to consider 'alt-metrics', things like social media traces and download statistics which supplement traditional measures like citation analysis. This are also part of increasing drives to measure and make academic research more accountable. In the UK for example, the complex and opaque process of ranking the research outputs of universities (The Research Excellence Framework or REF) is actually used to distribute government funding. While it is unclear if this has resulted in more robust research, it has certainly created a legion of consultants who help academics package their research in more visible, measurable ways. In an earlier keynote presentation, Wendy Espeland from Northwestern University explained how these ranking systems create new anxieties and ways of gaming the system.

Our goal for this workshop was relatively clear: we wanted to problematise these drives to measure academia by proposing concrete alternatives. Rather than rejecting measurement all-together, we were interested in how visualisations could be used to locate what certain measurement systems miss or show how these measurement systems drive behaviour (such as gaming and collusion) instead of just passively describing academic activities.

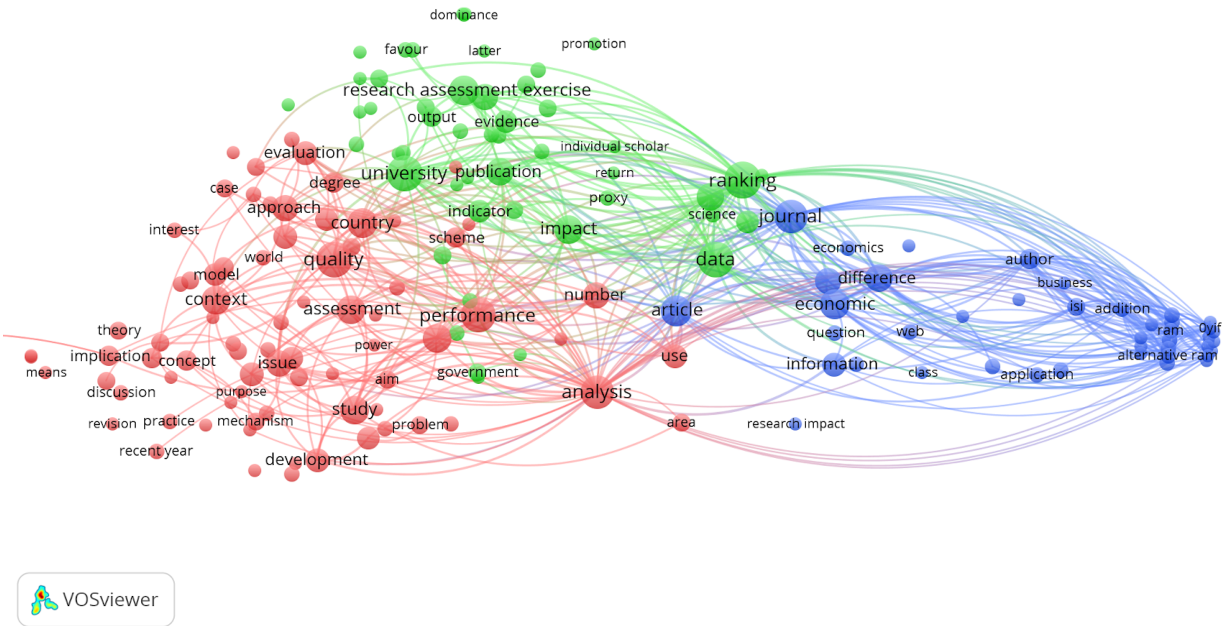
Reflexivity about Measurement

One group started by focusing on how citation practices and other related metrics work differently in different disciplines. For example, how does citation analysis understand emergent versus established communities of researchers? What kinds of phenomena are not visible to citation analysis and do these silences look different in different academic areas? One of the researchers had been working with theologians who published academic papers in journals just like social or natural scientists. However, the way they understood academic 'impact' was noticeably different. These researchers were happy when their ideas made their way into sermons or their initiatives became visible in local newspapers – in other words different media other than journal articles. Networks, including citation networks, as noted earlier, incline us to make particular traces equivalent (one citation is worth the same as any other citation) and exclude other types. How could we foster more 'heterogeneous couplings' between different types of materials?

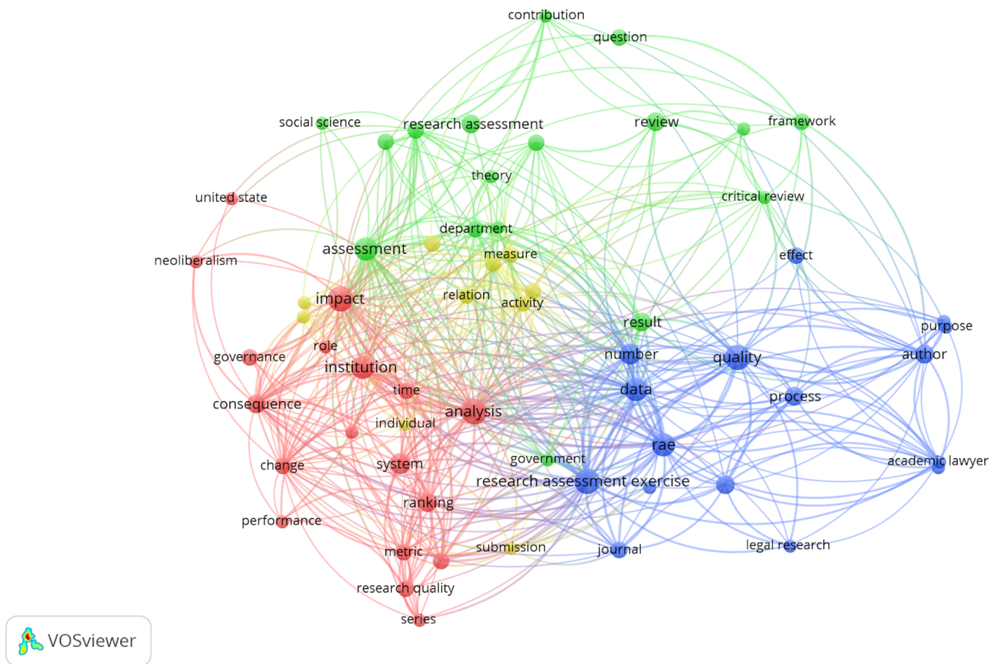
Another question which emerged was: in what ways do disciplines react to the condition of being measured. Are they reflexive about academic measurement in different ways? Do they focus on different types of overflows (what exceeds or falls outside metrics)? As ever, these investigations started with a database and some search queries. Web of Science provides one freely available resource for articles and citations. The group decided to search for the following terms in journals tagged as Economics and Sociology in Web of Science.

> TS='academic evaluation*' OR
> TS='research excellence framework' OR
> TS='Norwegian system' OR
> TS='Performance based funding'
> TS='research assessment'

The resulting list of articles, or more specifically their abstracts, were then visualised as co-word networks using the programme [VOS viewer](), developed by the University of Leiden.

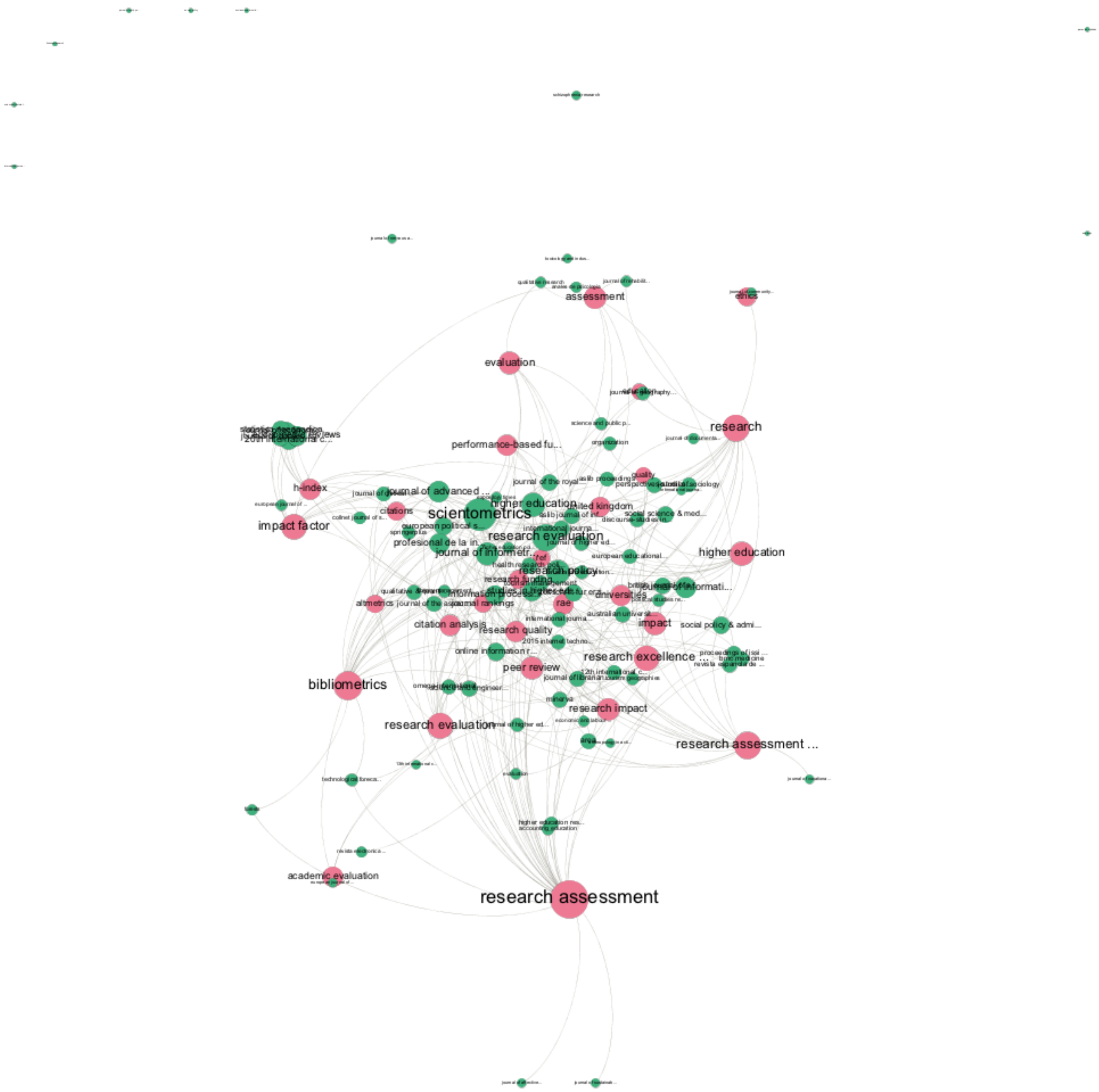


Co-word of article results in economics



Co-word of article results in sociology

These networks showed different key words used by the different disciplines but this presented a methodological problem: using these disciplinary tags meant accepting the definitions of the disciplines provided by Web of Science when they wanted to see how the disciplinary divisions were manifested *in practice*, around this very particular topic. One way out of this was to make a bi-partite network of journals and keywords, allowing for the possibility that journals, which often have multiple disciplinary identities, would cluster together in different, less obvious ways.



Bi-partite network of keywords, sources for all data (filtered for occurrences of 10 terms) Green: Journals, Red: Keywords

This result indeed seemed to show a less obviously disciplinary division in terms of the language chosen. Rather than economics and sociology, this network clustered around empirical topics and particular techniques of academic assessment. These networks did not immediately make sense on their own so the group decided to qualitatively analyse the articles and look at how they construed the relationship between disciplinary specificity and generic accountability. It should be noted that, although participants frequently talk about their lack of technical abilities, the words 'qualitative' and 'quantitative' were rarely mentioned in the course of these workshops, except when switching between types of analysis as a contrast.

Economics	Sociology
'The 'centres of excellence' policy implicitly pursued through the RAE is an optimal allocation strategy only if all departments in all disciplines are of the generalist variety, i.e. each pursues a research path through all its stages. Conversely, the RAE-induced research allocation minimizes efficiency if applied to specialist departments, when resources are concentrated on one specific research obstacle.' p 637 La Manna, M. M. A. 2008. 'Assessing the Assessment or, the Rae and the Optimal Organization of University Research.' Scottish Journal of Political Economy 55 (5):637-653.	'An exploratory modelling exercise using these variables to predict RAE 2008 revealed that despite what we might like to think about the subtle nuances involved in peer review judgements, it turns out that a fairly astonishing 83 percent of the variance in outcomes can be predicted by some fairly simple 'shadow metrics': quality of journals in the submission, research income per capita and scale of research activity.' (p. 130) Kelly, A., and R. Burrows. 2011. 'Measuring the Value of Sociology? Some Notes on Performative Metricization in the Contemporary Academy.' Sociological Review 59:130-150.
'For that reason, an alternative ranking method is developed as a quality indicator, which is based on membership on academic editorial boards of professional journals. This ranking method constitutes a good approximation of the appreciation, hence the quality, attributed by professional peers.' p3 Frey, B. S., and K. Rost. 2010. 'Do Rankings Reflect Research Quality?' Journal of Applied Economics 13 (1):1-38.	'Transdisciplinary dialogues are essential if we are to establish common ground and it is with this objective in mind that we offer a conceptual meta-framework for assessing arts-based works.' (p. 321) Lafreniere, D., and S. M. Cox. 2013. ''If You Can Call It a Poem': Toward a Framework for the Assessment of Arts-Based Works.' Qualitative Research 13 (3):318-336.

Examples of key quotes from economics and sociology papers

One hypothesis was that economists, who are closer in certain ways to the methods of measuring academia, might articulate the problem in more standardised ways. They selected a handful of texts (19 from Sociology and 50 from economics) and skimmed the abstracts for topical focus – to weed out some of the articles which, despite the key word searchers, were not talking about the efficacy or problems of evaluative metrics. They then read a handful of these articles, trying to pick out particular passages which spoke to the author(s) orientation to ranking.

The group found, perhaps unsurprisingly, that the two disciplines had very different approaches to formulating the problem. Economics framed academic evaluation as a technical problem (the measurements are wrong) while most sociologists (Kelly and Burrows being the exception) treated it more like a threat to academic practice. Both used lots of jargon but the economic jargon was more technical while the sociological jargon was theoretical.

It was interesting to hear that the group found the textual analysis helped the earlier co-word maps 'make sense'. In other words, they did not know what they were looking at until they switched methods. Rather than visualisations acting as an orienting device, *they only became sensible through another analysis*. They described this as the 'joy' of two different methods, visualisations and qualitative readings confirming their prejudices (economics has narrow problem-framings). Just because different methods produce different realities, does not mean that these realities cannot converge, however often this convergence seems to happen around relatively uncontroversial claims – such as about the differences between economics and sociology.

Invisibilities of Business Schools

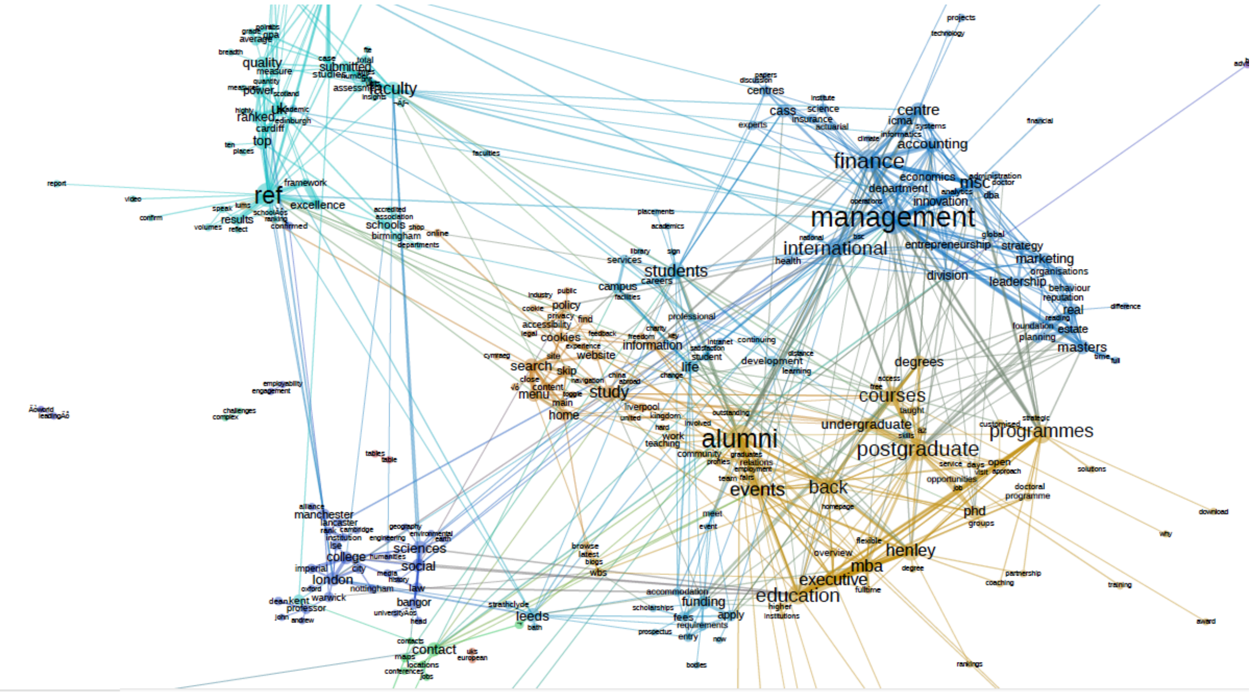
The other group, in a somewhat different way focused on the invisibilities created by metrics. For example, the effects of measurement systems (gaming, reorienting, anxieties etc.) are manifested in other media like online comic strips, personal correspondence and blog posts in ways that do not emerge in journal articles. It was noted that sometimes invisibilities are strategically necessary – many academics have protested the drive to measurement by withholding data – but it could also be strategic to make visible these detrimental effects of metrics.

The group focused on business schools as a special case. Business school managers seemed to take rankings more seriously than other disciplines, while at the same time deftly sidestepping their implications when it suited them. In a similar way to the other group, they became interested in reactions to measurement, particularly the Research Excellence Framework (REF) in the UK. Schools often issue press releases reacting to the REF, training employees to be 'REF ready', reframing the results and trying to game them in advance.

The group gathered a list of business schools in the UK mentioning the REF on their websites. They wanted to think of different ways of creating typologies of reactions to the REF and then matching these up to their previous ranking results, exploring Espeland's idea that that organisations which are on the cusp (e.g. barely in the top ten or almost in the top third) will be more anxious about rankings than schools comfortably situated in the middle of a particular tranche.

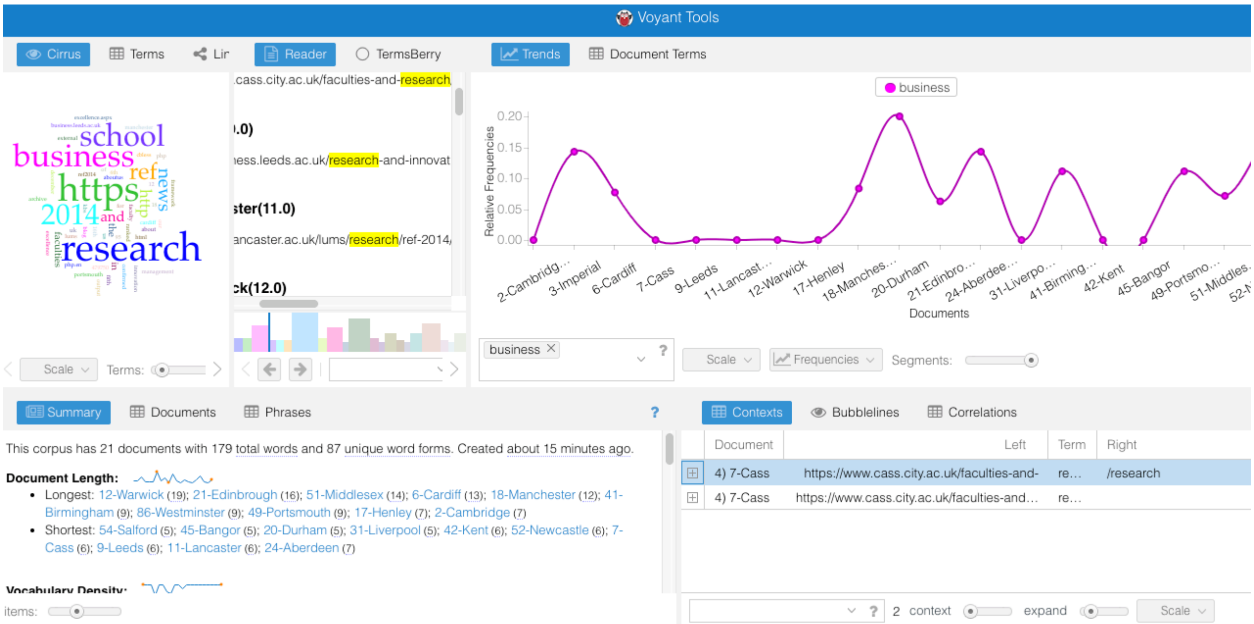
> Since 1990, Cambridge Judge Business School has forged a reputation as a centre of rigorous and high-impact thinking and transformative education. Our School strategy is not dictated by rankings, as rankings do not reflect all the things that we want to achieve. Our aim is to create a truly transformative experience for students, to challenge our students to see their careers through a wider lens and encourage them to pursue careers that they can proudly look back on in 30 years and feel they have made a positive difference to their communities and the wider world.
>
> While external recognition of our Business School, programmes and research don't define our mission and strategy, they are a useful external view of the steps we are taking to create one of the world's top business schools.

Because Cambridge, above, is highly ranked they can perhaps afford to dismiss rankings, but if they dismiss them too much, they might undermine their success. The group was, however, confronted by the problem that the REF results themselves were in PDF form and not immediately convertible to a table. This is another example of how open data is not always as straightforwardly open and available as it seems. There were also questions about which numbers or rankings were most important: the top level school ranking or specifically research outputs?



Co-word network of business school response texts

As usual, they started with co-word analysis (using Wordij), which analyses the corpus of texts together through word distance, but this did not bring the schools as entities into the analysis, nor did it allow us to think about the relationship between assigned rank and reaction.



Voyant tools – with business schools arranged in rank order

They instead turned to a simpler tool called Voyant which allowed them to keep the texts separate. However, this tool only took into account the frequency of particular words, rather than their relationships through co-occurrence. It took them some time to load the texts into Voyant in rank order and when they did there seemed to be no patterns which made sense to them. One of the limits of such frequency measures is that one needs a large enough corpus of texts to have big enough numbers of words to compare, but the large corpus then precludes easy manual analysis of the content.

Later, they considered that it might be equally interesting what business schools *did not say* as what they did say. Making a list of recurring items which appeared in these reactions, they then queried the text to see whether or not term was present. For example, did schools mention hiring rates as an alternative metric of the success of their students? Did they mention how many submissions they had – a key aspect of how the score was calculated? Frequencies of these terms would not tell us very much but a matrix which highlighted *the lack of particular terms* would allow one to see what was conspicuously omitted from the texts but present in others. A similar process could have been used to figure out which schools mentioned each other in the texts – what comparisons did they choose to strategically make – since rankings are, importantly, a relational phenomenon.

Business School			REF	Overall	Impact	Acceptance Rate	Hiring Rates	Submissions	GPA
2-Cambridge	https://www.jbs.cam.ac.uk/a	Research Excellence F The REF assesses the For the purpose of the I	5	1	0		0	0	
3-Imperial	https://www.imperial.ac.uk/b	REF 2014 | Imperial Cc	10	1	1	1	0	1	
6-Cardiff	https://www.cardiff.ac.uk/ne	Cardiff Business Schoo	6	1	0	1	1	1	
7-Cass	https://www.cass.city.ac.uk/r	Research at Cass | Cas	7	1	0	0	0	0	
9-Leeds	https://business.leeds.ac.uk	REF 2014 Results - Lee	3	1	1	1	0	0	
11-Lancaster	http://www.lancaster.ac.uk/lu	) REF 2014 | Lancas	6	1	1	1	0	0	
12-Warwick	https://www.wbs.ac.uk/news	WBS ranked fifth for res	14	1	1	1	0	0	
17-Henley	https://www.henley.ac.uk/ne	REF 2014 results speal	7	1	0	0	0	0	
17-UEA	https://www.uea.ac.uk/norwi	Norwich Business School	10	0	1	0	0	0	
18-Manchester	https://www.mbs.ac.uk/news	Manchester confirmed a	13	1	1	0	0	0	
20-Durham	https://www.dur.ac.uk/busin	Durham University Bus	5	1	0	0	1	1	
21-Edinbrough	https://www.business-schoo	Business School ranks	0	0	0	1	0	0	

Mock-up of matrix of recurring words (with dummy data) – absences in grey.

The group also wanted to capture the 'plot lines' of the reactions – what order certain elements appeared in, in crafting the story. But as usual they were transfixed by what these techniques could not capture – *how* different schools are mobilised in the text. The possible substitution of one ranking for another 'We are in the top 20 in _____ but top5 in ____' or 'Top 5 schools in Yorkshire' performs a particular audience (or possible students) through the text but this is not available to us through the appearance of particular rankings or not. Perhaps more glaringly these textual analysis tools, ignored one of the more important aspects of this particular case – the graphic presentation of images, tables and numbers. Managing expectations in relation to the REF was a visual and aesthetic as well as rhetorical achievement.

Discussion

One interesting aspect of this workshop was that the participants were more hybrid than the others, or at least came from more hybrid departments. Rather than situating themselves in either qualitative and computational traditions many of the participants came from scientometric and library science departments and were comfortable in both ways of working, because these topics have been hybrid for some time.

However, in the recap at the end, the groups also felt that they remained in their silos: those who were adept at making visualizations were busy demonstrating things while the self-identified ‘qualitative’ researchers would busy themselves reading. Still, it felt as if the discussions became more targeted. Exclamations of ‘what does this show?!’ or ‘How does this work?!’ were replaced by exchanges like:

> ‘This graph seems more meaningful than the other one’
>
> ‘I’d rather not reduce away the noise in case there’s something in there’
> ‘Well depends on if it’s a meaningful reduction or not’

The word meaningful (as well as other ambiguous words like ‘interesting’) were deployed frequently to describe graphs, the contents of data or particular practices like data cleaning. In these interactions, there seemed to be the subtle acknowledgement that some process of reduction or simplification was necessary but shifted the question to whether or not it was ‘good’. In the absence of an established research protocol, such words invite possible valuation practices – does this action yield phenomena that seem sociological?; does it speak to the collectively defined problem?; and accounting practices: ‘we like this move but can we justify it to others?’ In most cases, ‘meaningful’ was not attributed to a person ‘do you find this meaningful’ but was positioned as collective good. So, what counts as meaningful becomes negotiated collectively.

Findings

These workshops reinforced several established ideas in the literature abut computational techniques and data visualisations. Firstly, they have a tendency to shift our objectives to what is easy to measure, or incline us to asking certain types of questions (Marres and Weltevrede, 2013; Uprichard, 2013). Some of the participants were understandably surprised when, at the final wrap up, they were confronted by the research questions and problem definitions they wrote down on the first day. One participant called this 'funnelling': the problem or object of study becomes narrowed as we feel out the limits of certain types of analyses. Secondly, visualisations especially are seductive: participants talked about 'visualisation envy', or the affective 'joy' of a visualisation working. Most displays of complex visualisations were accompanied by 'oooohs' and 'aaaahhh', especially when animations made them dance across the screen. Thirdly, far from 'revealing' insights, they create phenomena which captures our attention so that other phenomena are ignored. Fourthly, they are also embedded with tacit assumptions about how the world works which do not always travel well between data sets, fields or local specificities. Furthermore, we noticed how visualisations interact with stated or unstated expectations and assumptions: we only know that they 'work' when they confirm expectations and they are only 'interesting' or 'meaningful' when they diverge from what we expect. Sometimes our expectations or assumptions are only manifested in the process not given before.

However, while there were occasional invocations of the spectre of 'quantitative' and 'qualitative' research, in practice things were far more fluid. Often participants found themselves reading texts 'like a computer' or analysing a graph 'interpretively'. In particular we realised that computational tools, particularly data visualisations are not one thing, but can perform several different roles in the research process. They can **tell stories** which then come to inflect manual readings of the data, but they can also function as **maps of an unknown territory**, giving us little directions, rather than definite courses of action. They can be deployed in a somewhat **ironic orientation** to existing techniques: to critique or undermine them, or they can be used iteratively to **generate surprises**, rather than convergence or 'saturation'. They can help emphasise **contrasts and comparisons** and highlight **absences and silences** between cases.

These offer different possible relationships between quantitative and qualitative research which are not based on triangulation, hypothesis testing or existing types of collaborations. The larger barrier to alternative ways of working, however concerns resistances and misunderstandings between researchers about what they are doing. What is interesting about our instinctual resistance to reducing a text to a set of numbers or a textually performed, relational entity to a tag is that anthropologists and qualitative-sociologists tend to take transformations as equivalent, or take them equally *seriously*. We think of analysing a text OR reducing it to a number or chart – and when we think of these as competing realities, then it becomes obvious that the later does violence to the former.[16]

It has been argued that in the natural sciences, empirical materials are often analysed through a sequential chain of transformations, normally by starting with some natural

[16] Ironically, this seems to encourage nominally qualitative researchers into a realist position – the individual text is what's *really* going on, in contrast to computational representation of it.

material and progressively extracting, formatting and separating it until it becomes a number in a table which can be used in a graph (Latour, 1999). Michael Guggenheim notes that in the social sciences we tend to make big transformations: fieldnotes into write-ups, or surveys into graphs which are, as he describes them, 'loose transformations' – in which the later hardly resembles the former and the gap between them remains somewhat mysterious. He argues that sociologists could learn something from breaking down their analysis process into smaller steps (2015). However, what qualitative researchers often criticise about the natural sciences model is the fact that these sequential transformations become irreversible – how do you return to complexity from the numbers? This is less of a problem with digital data in which switching back and forth between individual cases and aggregate patterns is much easier. So maybe qualitative researchers would be less precious about reductions, if they were part of an iterative chain in which reductions are used to make maps which help us read texts rather than *replacing* the act of reading texts. Perhaps we need to commit to some problematic assumptions in order to question other inherited ideas.

But this leads us to a bigger question about what is the point of using digital tools. The standard version is that digital tools allow qualitative researchers to scale up insights – to make defensible 'descriptive' generalisations, or taken a bit further, to make causal claims. For micro-sociological approaches to data analysis: when the social order is in some sense indexed in every interaction or set of utterances then what does it mean to compare multiple such slices of activity or generalise from them? Would we find a million competing moral orders or some regularities between them? More generally, the upshot of ethnographic or micro-sociological accounts is precisely in defending local specificities. 'anthropologists do not generalize from the particular they see the general in the particular.'[17] But while often rejecting numbers or categorical reductions, ethnographic cases often come to stand for, speak to, or be positioned as a symptom of something larger. We are also comfortable trying out heuristics like 'multiplicities' and 'contrast structures', developed for specific situations, and applying them to any manner of cases.

Tim Ingold (2008), while discussing historical tensions between ethnologists and ethnographers, offers a different idea of what generalisation might mean for anthropologists. While some ethnologists argued that they were comparing far flung tribes to uncover fundamental laws or uncover patterns and this involved starting with local particulars and moving to generalities. But others like Kroeber saw this as flawed because particulars are never particular to begin with.[18] He likened the act of situating case studies and local phenomena in a wider field to an artist painting a landscape. 'to the artists gaze, the landscape presents itself not as a multitude of particulars but as a variegated phenomenal field, at once continuous and coherent… …Within this field, the singularity of every phenomenon lies in its enfolding – in its positioning and bearing, and in the poise of a momentarily arrested movement – of the entangled histories of relations by which it came to be there, at that position in that moment.'(Ingold, 2008: 73) So the artists role is not to reduce or reveal but to draw out certain phenomena which are meaningful – something which needs to be negotiated between researchers and audiences. Ingold also argues however, that this move should not be thought of in opposition to grand theories, or the

[17] Evans-Pritchard Quoted in Ingold (2008).

[18] Isolating a set of actors, a situation or a slice of activity is an artificial, and active process.

sorts of tacit assumptions Cicourel talked about, these are what sensitize us to telling certain stories over others.

We do not think that that qualitative researchers should engage with quantitative techniques because they can help make generalisations from particulars, though that might be a useful way of enrolling quantitative interlocutors. As long as we view visualisations as part of the research process, then they can be thought of adding to qualitative readings rather than as a pale imitation of them. Also, qualitative researchers should use visualisations *because* they are seductive, not more or less seductive than say, a flowery quote from an anthropologist, but seductive in different ways which suggest different sorts of audiences, and might allow us to tell different sorts of stories.

References

Abildgaard MS, Birkbak A, Jensen TE, et al. (2017) Five recent play dates. *EASST Review* 36(2). Available at: https://easst.net/article/five-recent-play-dates/ (accessed 20 November 2017).

Anstead N (2017) Data-driven campaigning in the 2015 United Kingdom general election. *The International Journal of Press/Politics* 22(3): 294–313.

Barocas S (2012) The price of precision: Voter microtargeting and its potential harms to the democratic process. In: *Proceedings of the first edition workshop on Politics, elections and data*, 2012, pp. 31–36. ACM.

Barry A, Born G and Weszkalnys G (2008) Logics of interdisciplinarity. *Economy and Society* 37(1): 20–49. DOI: 10.1080/03085140701760841.

Callon M, Law J and Rip A (eds) (1986) *Mapping the Dynamics of Science and Technology*. London: Springer.

Card SK, Mackinlay JD and Shneiderman B (1999) *Readings in Information Visualization: Using Vision to Think*. Morgan Kaufmann.

Cicourel AV (1964) *Method and measurement in sociology.* Oxford, England: Free Press of Glencoe.

Coopmans C (2014) Visual Analytics as Artful Revelation. In: Coopmans C, Vertesi J, Lynch ME, et al. (eds) *Representation in scientific practice revisited*. MIT Press, pp. 37–59.

Fielding JL and Fielding NG (2008) Synergy and synthesis: Integrating qualitative and quantitative data. In: Alasuutari P, Bickman L, and Brannen J (eds) *The SAGE Handbook of Social Research Methods*. London: Sage, pp. 555–571. Available at: http://epubs.surrey.ac.uk/231711/2/SRI_deposit_agreement.pdf (accessed 19 May 2013).

Gitelman L (2013) *Raw Data Is an Oxymoron*. MIT Press.

Glaser BG and Strauss AL (1967) *The discovery of grounded theory: Strategies for qualitative research*. Aldine de Gruyter.

Goodwin C (1994) Professional Vision. *American Anthropologist* 96(3): 606–633.

Guggenheim M (2015) The media of sociology: tight or loose translations? *British Journal of Sociology* 66(2): 345–372.

Hammersley M (1992) Deconstructing the qualitative-quantitative divide. In: Brannen J (ed.) *Mixing Methods: Qualitative and Quantitative Research*. Aldershot: Avebury. Available at: http://oro.open.ac.uk/20445/ (accessed 27 January 2015).

Ingold T (2008) Anthropology is not ethnography. In: *Proceedings of the British Academy*, 2008, pp. 69–92.

Issenberg S (2012) *The victory lab: The secret science of winning campaigns*. Broadway Books.

Kennedy H, Hill RL, Aiello G, et al. (2016) The work that visualisation conventions do. *Information, Communication & Society* 19(6): 715–735. DOI: 10.1080/1369118X.2016.1153126.

Knorr-Cetina KD (1981) *The micro-sociological challenge of macro-sociology: towards a reconstruction of social theory and methodology*.

Kreiss D (2012) *Taking our country back: The crafting of networked politics from Howard Dean to Barack Obama*. Oxford University Press. Available at: https://books.google.co.uk/books?hl=en&lr=&id=mayTrDHJVUkC&oi=fnd&pg=PP2&dq=kreiss+taking+our+country+back&ots=x7MGHbNfvy&sig=rh71YXcsndQY9ygdxyBdBzDbBk4 (accessed 1 March 2017).

Latour B (1999) *Pandora's Hope: Essays on the Reality of Science Studies*. Cambridge, MA: Harvard University Press.
Latour B (2005) *Reassembling the Social: An Introduction to Actor-Network-Theory*. Oxford, UK: Oxford University Press.
Leonelli S, Rappert B and Davies G (2017) Data Shadows: Knowledge, Openness, and Absence. *Science, Technology, & Human Values* 42(2): 191–202. DOI: 10.1177/0162243916687039.
Lynch M (1988) The externalized retina: selection and mathematization in the visual documentation of objects in the life sciences. *Human Studies* 11(2): 201–234.
Marres N and Weltevrede E (2013) Scraping the social? Issues in real-time social research. *Journal of Cultural Economy* 6(3): 313–335.
Munk AK, Tommaso V and Meunier A (2016) Data Sprints: A Collaborative Format in Digital Controversy Mapping. In: *Digital Sts Handbook*. Princeton University Press.
Neff G, Tanweer A, Fiore-Gartland B, et al. (2017) Critique and contribute: A practice-based framework for improving critical data studies and data science. *Big Data* 5(2): 85–97.
O'Neil C (2016) *Weapons of math destruction: How big data increases inequality and threatens democracy*. New York: Crown Publishing Group.
Pasquale F (2015) *The black box society: The secret algorithms that control money and information*. Harvard University Press. Available at: https://books.google.co.uk/books?hl=en&lr=&id=TumaBQAAQBAJ&oi=fnd&pg=PP8&dq=pasquale+black+box&ots=BdgGkgOQ9x&sig=mW3V5ndGYlcPzcH-k18d3NYk81k (accessed 20 February 2017).
Rogers R (2013) *Digital Methods*. Cambridge, MA: MIT Press.
Rogers R and Marres N (2000) Landscaping climate change: A mapping technique for understanding science and technology debates on the World Wide Web. *Public Understanding of Science* 9(2): 141–163.
Ruppert E, Harvey P, Lury C, et al. (2015) Socialising big data: from concept to practice. *CRESC Working Paper Series* (138). Available at: http://research.gold.ac.uk/11614/ (accessed 24 November 2016).
Smith DE (1978) `K is Mentally Ill' the Anatomy of a Factual Account. *Sociology* 12(1): 23–53. DOI: 10.1177/003803857801200103.
Stengers I (2011) "Another science is possible!" A plea for slow science. In: *Inaugural lecture of the Willy Calewaert Chair*. Available at: http://we.vub.ac.be/aphy/sites/default/files/stengers2011_pleaslowscience.pdf (accessed 27 May 2015).
Tkacz N (2014) *Wikipedia and the Politics of Openness*. Chicago: University of Chicago Press.
Uprichard E (2013) Focus: big data, little questions? Available at: http://www.discoversociety.org/2013/10/01/focus-big-data-little-questions/ (accessed 30 September 2014).
Venturini T and Latour B (2010) The Social Fabric: Digital Traces and Quali-Quantitative Methods. In: *Proceedings of Future En Seine*, 2010, pp. 30–15. CAP Digital.
Venturini T, Laffite NB, Cointet J-P, et al. (2014) Three maps and three misunderstandings: A digital mapping of climate diplomacy. *Big Data & Society* 1(2): 1–19.
Wynne B (2011) *Rationality and ritual*. Abingdon: Earthscan.
Zuiderent-Jerak T (2015) *Situated Intervention: Sociological Experiments in Health Care*. Cambridge, MA: MIT Press.

Participants

Workshop 1

David Moats	Linköping University
CF Helgesson	Linköping University
Laurie Waller	Technical University of Munich
Jan Passoth	Technical University of Munich
Anne Helmond	University Van Amsterdam
Fernando Van der Vlist	University Van Amsterdam
Anders Koed Madsen	Aalborg University
Andreas Birkbak	Aalborg University
Torben Elgaard Jensen	Aalborg University
Sveta Milyaeva	Goldsmiths College

Workshop 2

Marsha Rosengarten	Goldsmiths College
Liz McFall	Open University
Susi Geiger	University College Dublin
David Moats	Linköping University
Steve Woolgar	Linköping University
CF Helgesson	Linköping University
Cory Robinson	Linköping University
Katerina Vrotsou	Linköping University
Lotta Bjorklund-Larsen	Linköping University
John Powell	Oxford University
Farzana Dudhwala	Oxford University
Anne-Marie Boylan	Oxford University
Veronika Williams	Oxford University

Workshop 3

Wendy Espeland	Northwestern University
David Moats	Linköping Universitet
CF Helgesson	Linköping Universitet
Steve Woolgar	Linköping Universitet
Fredy Gamez	Linköping Universitet
Jonas Löwgren	Linköping Universitet
Björn Hammarfelt	University of Borås
Gustaf Nelhans	University of Borås
Christopher Kullenberg	Göteborg Universitet
Sarah de Rijcke	Universiteit Leiden
Tjitske Holtrop	Universiteit Leiden
Daniel Neyland	Goldsmiths College