29 million and counting: capital numbers and the calculative agencies of science and technology studies.

At almost every point in its history, STS has taken an active interest in what numbers do, who they matter to, and how they are made. The extraordinary diversity of numbering practices studied by STS and affiliated fields include enumerations (Verran, 2001), measurements (Latour, 1987), prices (Callon & Muniesa, 2005), risks and chance (Porter, 1996), statistics (Hacking, 1990), indices (Nafus, 2014), metrics (Burrows, 2012), percentages (Guyer, 2014), model parameters, or even computable numbers (Mackenzie, 1997), as well as problematic and contested numbers (Verran, 2012; Lampland, 2010; Miller, 2009). From this rich panoply of work, many lessons concerning the historical, contextual, material and political fabric of numbers can be drawn. Crucial to our elementary senses of the multiple, numbers and numberings effloresce in reef-like growths that resist any attempt to harmonise their essence in a single mathematical formalism such as set or number theory. STS work has not itself shied away from working with and producing numbers. We need only think of the many metrical components of actor network theorists in their work on co-word analyses of scientific literature (Callon, Courtial & Laville, 1991), the software that Thomas Kuhn describes in the appendices to *The Structure of Scientific Revolutions* (Kuhn, 1996) or the ongoing scientometric strands of STS research (Leydesdorff, Carley & Rafols, 2013).

The contemporary agency of STS around numbers is somewhat obscure, since STS finds itself amidst the ethical, epistemological, infrastructural and economic entanglements of contemporary data practices (Gitelman, 2013; Halpern, 2015; Ruppert, Law & Savage, 2013; Kitchin, 2014), as well as the sometimes costly calculative complexities found in many scientific, governmental and commercial domains. Economists (Porter, 2008), social physicists (Pentland, 2014), statisticians (Desrosières, 2011) or data scientists (Diggle, 2015) have, it seems, exponentially greater calculative power than STS researchers in their work with numbers. The ways and byways along which STS researchers encounter numbers will be diverse. Social media (Marres, Gerlitz & others, 2015), scientific publications (Leydesdorff, 1998; Latour, Jensen, Venturini, et al., 2012), scientific databases (Bowker, Baker, Millerand, et al., 2010), official statistics, and web-scraping (Rogers, 2004) present very different, albeit increasingly interwoven assemblies of number.

If we heed STS's own accounts of what happens in knowledges, devices, practices of enumeration and calculation, then we can expect numbers to multiply, or to become multiply multiple. (For instance, if the body is multiple (Mol, 2003), then numbers, the figure of the multiple, might also be multiple.) Numbers occur amidst what Michel Callon and Fabian Muniesa call 'calculative agency' (Callon & Muniesa, 2005). Could the study of numbers borrow or imitate some of the operations of calculative agency that crystallize valued numbers in other settings in order to practice different versions of number? Callon and Muniesa describe the 'costly activities' of calculation in terms of three operations:

> Isolating objects from their context, grouping them in the same frame, establishing original relations between them, classifying them and summing them up are all costly activities that raise the question of calculative power.(Callon & Muniesa, 2005: 1232).

The operations of isolating, grouping together, establishing relations, summing up are 'costly' because they need to be configured as a form of agency that draws on tools, devices, goods and services, and multiple expertises (in relation). As Callon and Muniesa observe, 'calculative agencies are as numerous and diverse as the tools they use and the hybrid collectives to which those tools belong' (Callon & Muniesa, 2005: 1245). While STS researchers might well eschew calculative agency (since so many sciences and technologies tenaciously leverage, marshal and at times monopolise numbers), the social study of science and technology will often be dogged by the necessity to make sense of, enumerate, count, figure and re-configure numbers.

Numerical encounters may well lack the ostensible precision of some forms of calculative agency. Indeed, other STS approaches point to the messier but perhaps more challenging numerical responsibilities. Rather than calculative agency, Helen Verran, for instance, suggests that any work with numbers, including our own as researchers, is bound to involve political differences, and that these differences will affect any work done with numbers:

> In any practical going-on with numbers, what matters is that they can be made to work, and making them work is a politics. Yet is a politics that completely evades conventional foundationist analysis (Verran, 2001: 88)

Calculative agencies associated with numbers, in all their diversity, will always require, from this perspective, that we think of numbers in a double sense as both an engagement with multiplicity and as a device-specific negotiation of differences and becomings. This double sense of numbers as both a calculatively ordered cutting-framing-summing and as materially specific figuring of differences could be also understood in terms of Lucy Suchman's 'configuration' device:

> the device of *configuration* has two broad uses. First, as an aid to delineating the composition and bounds of an object of analysis, in part through the acknowledgment that doing so so is integral not only to the study of technologies, but to their very existence as objects. And second, in drawing our analytic attention to the ways in which technologies materialize cultural imaginaries, just as imaginaries narrate the significance of technological artefacts. Configuration in this sense is a device for studying technologies with particular attention to the imaginaries and materialities that they *join together*,

> an orientation that resonates as well with the term's common usages to refer to the conjoining of diverse elements in practices of systems design and engineering (Suchman, 2012: 48).

If numbers are configured, they have bounds and a composition not unlike other things or devices. At the same time, numbers inevitably numerically materialize a cultural imaginary of inclusion, value, difference, or change. Numbers, or figures in the numerical sense of the term, are *configurative.*

Re-counting a capital number on a platform

The implications of configurative numbers could be explored in many places. I report here on a deliberately naive attempt to count a single number: the approximately 29 million code repositories on the social media platform Github at a particular point in time (late 2015). The number 29 million is the case study and perhaps the principal actor here. The number appeared in a research project focused largely on transformations in the social life of code. But the project also explicitly sought to experiment with data-intensive methods for analysis of re-assemblings and re-associations typical of places such as Github.[1] On Github, two numbers have capital importance – the number of people using the platform, and the number of different code repositories stored there – but I deal in this paper only with the latter. The counting I undertake will focus on problems of copying, duplicating, and imitation that render large numbers a moving substrate.

Coding practices have changed over the last decade or so in fairly profound ways. The organization of software development has shifted away from massively centralized software development (the kind epitomised by Microsoft or IBM during the latter part of last century (Campbell-Kelly, 2003a; Ensmenger, 2010)) to a much more de-centralized process closely woven into communication and network infrastructures. The geography of coding still centres on well-established urban centres such as San Francisco and London (Mackenzie, Fuller, Goffey, et al., 2016), but by virtue of a much wider distribution of coding tools and skills, occurs in many other places as well. Added to this global geography, changes in digital devices and infrastructures, ranging from proliferating mobile devices and sensors to mushrooming data centres have widened the reach and variety of coding. Code addresses much greater parts of the world today than it did at the turn of the century. The lines and borders separating the

[1] The study sought to describe 'meta-community' practices in the code-sharing commons by examining patterns of imitation and copying across a large number of software development projects. The project made use of data from the platform (via various devices discussed below). While the research almost entirely focused on this platform, it also participated in that platform. This manuscript and others relating to the project, as well as the many scripts, visualizations, queries and other pieces of work with the data can all be found on the platform at [URL removed for anonymous review]. Versions of the manuscript found there include code and queries used to enumerate or re-count the various numbers I discuss here. This manuscript and its associated files is one of the 29 million.

production, distribution, and use of code have thoroughly blurred. Code flows in much more diverse ways through infrastructures and devices. For instance, 'devops' (development-operations) deploys frequently revised code directly into the ongoing transformative-maintenance of information infrastructure through continuous deployments, often directly from repositories on sites such as Github (Mackenzie, 2016). Across all of these changes, coding has become a much more *public* activity (see (Kelty, 2008)), and the *publicness of code* has become a major concern for business, government and science, albeit for different reasons (for instance, in scientific work, open data has been accompanied by a rapid proliferation of techniques, resources and tools for publishing code written for scientific purposes; for instance, in the many attempts to train school students and others to code as a putative 'basic life skill').

The social media site Github, started in in 2007, epitomises and has indeed been central – as the suffix 'hub' suggests – to these changes. Like many social media platforms, it has grown tremendously in the last 7 years to include around 31 million software projects (February 2016). This growth flows from a tremendous variety of processes that are difficult to summarise or classify partly because the actors, topics or domains of coding are so diverse and partly because much of what flows through Github is both technically and socially 'innovative' in the sense that Bruno Latour uses the term: '" innovative" means that we do not know the number of actors involved in advance' (Latour, 1996: 72). Github's growth is not unique. Online code repositories have existed for several decades (most famously SourceForge), although more recent arrivals such Github, BitBucket, GoogleCode (which has closed and migrated to Github), GitLab, or Gitorious tend to invite and promote collaborative coding on a larger scale.

Like many contemporary assemblages, Github seeks to describe itself through large numbers, and these numbers in their sometimes daily changes and updates attest to an investment in being innovative or open-ended enumerations. In June 2014, the Github homepage showed a photo of a women working in an urban office location, with lights and professional camera focused on her work. It described '6.1 million people collaborating right now across 13.2 million repositories … building amazing things together' (Github, 2014b). In late November 2015, Github has a slightly more functional description, and the image seems to be a press conference in China:

> GitHub is how people build software. With a community of more than 12 million people, developers can discover, use, and contribute to over 29 million projects using a powerful collaborative development workflow. (Github, 2015)(https://web.archive.org/web/20151216055610/https://github.com/about

The numbers change over time alongside the composition of the actors involved (women coding and working at Github, a sore point for the company; Chinese developers using Github, but also political activists, leading to a denial of service attack on the Github, allegedly by the Chinese government[2]). But the chang-

[2]Neither image – the woman at work in a funk urban office or the press conference in China – is arbitrary, again suggesting that enumeration is always also configurative in the sense of

ing numbers, alongside the sometimes rather sublime scales of community (a 'community' of 12 million people?) certainly form part of Github's description.

How might we treat numbers such as 12 million and 29 million? They have a specific texture and consistency as counts of actors associated with platforms, devices and infrastructures. In important ways they are part and parcel of platform politics and market economies (Gillespie, 2010). Given the capital investments that Github has attracted in the last few years ($US100 million from the venture capital investment fund Andressen-Horowitz in 2012), these numbers affect flows of capital. They may be 'false numbers' as Martha Lampland suggests, designed to enable certain forms of rationalization (Lampland, 2010). They also have the feel of what Carolin Gerlitz and Celia Lury describe in their study of the social reputation website Klout as 'reactive numbers' meant to evoke further work, action and engagement around those numbers (Gerlitz & Lury, 2014). The numbers 29 million or 12 million are examples of widely used numerical forms: these numbers enact Github as an 'enumerated entity' (Verran, 2012: 61). While these numbers resemble others – population statistics for small nations, amounts of money spent on a new building, a research programme or a CEO's remuneration – they become visible, circulate and are seen in ways that prevent them from being easily assimilated to biopolitical or financial numbers. Their salience and pluri-potence will be described further below, but already they warrant some kind of designation and differentiation. We might call them 'capital numbers.' Capital here has the triple sense of a centre ('hub'), the capital letter of a proper noun, and an accumulated fund with potential to growth. A capital number enumerates an aggregate, imbues it with importance linked to infrastructural scale and complexity, and promises the potential to generate further accumulation.

The litany of sense-making: enumerating the aggregate through events

What is enumerated on Github? At the end of 2015, the two main enumerations are 12 million 'people' and 29 million repositories. The people are software developers or coders. `Git` repositories are collections of files mainly containing code, but also a great variety of configuration specific documents (settings, manuals, installation instructions, etc.) stored in many versions and varieties. Any enumeration of Github contends with not only the sheer number of people visiting and using the platform, but the sheer diversity of ways in which

materialising imaginaries of wholeness, difference, inclusion, etc. Github was rocked in 2012 by allegations of sexism and mismanagement, particularly associated with one of the co-founders, who subsequently resigned and left Github (Github, 2014a). In early 2015, Github itself was subject to a massive Denial oF Service attack, emanating from China. Reports alleged that the DoS attack targeted pro-democracy repositories on Github. Unlike its regulation of other social media platforms, China does not block Chinese citizens' use of Github because Github is seen as economically and commercially important for the software industry (Marczak & Weaver, 2015–2015-04-10T05:55:18+00:00).

they associate with each other through repositories using one or more of several hundred different programming or scripting languages. It is difficult to say which capital number matters more – the head count or the repository count. On Github, both normally appear in tandem: '2,641,337 people hosting over 4,442,708 repositories' announces the Github home page on 27 November 2012 (Github, 2012).

While Github features capital numbers on its webpages, those numbers re-play in other settings (for instance, in the many articles appearing in newspapers, magazines and business press about the popularity and growth of Github; see (Hardy, 2012; Meyer, 2013; Gage, 2015)). They are also re-counted in different ways. Take for instance the data published through the Github API (Application Programmer Interface). APIs were not created as a data resource for social scientists, but for software developers working in convergence cultures (Jenkins, 2004) in which connecting different platforms, devices, sites and practices together using code has become standard practice, APIs have great significance (Bucher, 2013).

In Github's case, data published through the API refers to movements of code, and associative acts connected to coding. The data is generated as developers write code and move that code in and out of code repositories using the `git` system (Git, 2014). Coding today constructs and take place in increasingly complex associative infrastructures. Github seeks to code coding as a social network practice. The collaboration or 'sharing' (to use social media keywords) coordinates both people and things. The format of the data that affords and records this associative structuring collocates people and things, although the distinction between a 'person' and a 'thing' is never entirely clear on Github. By convention, for instance, much API data from social media sites has an 'event' structure that links named actors and named entities to specific infrastructural locations (usually coded as an URL) at a particular time (the 'timestamp'). Events on the Github 'events' API (Github, 2016) are classified according to one of roughly twenty actions.


```
{
"id": "2111998059",
"type": "WatchEvent",
"actor": {
"id": 1459103,
"login": "mmemetea",
"gravatar_id": "4532d1e4885f579ca7d9aa8748418817",
"url": "https://api.github.com/users/mmemetea",
"avatar_url": "https://avatars.githubusercontent.com/u/1459103?"
},
"repo": {
"id": 14802742,
"name": "OpenSensorsIO/azondi",
"url": "https://api.github.com/repos/OpenSensorsIO/azondi"
```

```
},
"payload": {
"action": "started"
},
"public": true,
"created_at": "2014-05-23T08:40:56Z",
"org": {
"id": 5497318,
"login": "OpenSensorsIO",
"gravatar_id": "1e0218942846ec8ef59f5d679dbca782",
"url": "https://api.github.com/orgs/OpenSensorsIO",
"avatar_url": "https://avatars.githubusercontent.com/u/5497318?"
}
}
```


The relatively simple `WatchEvent` on Github shown in Figure 1 documents the act of an `actor` calling themselves `mmemetea` interested in the repository called `azondi`, a software project coordinated by the 'organisation' calling itself `OpenSensorsIO`.[3] Note that the event also has various status designations – it is a `public` event, it has a 'payload' (often much more complicated than simply `started`) – and includes various indexical references or `ids` that link the event to other groups of people, organisations, repositories and images (`gravatar_id`). The intricate syntax of this data – many brackets, inverted commas, colons, commas – attests to a complex social coordination which aligns actors, actions, places and times in discrete events. The discreteness of such events attests to the potential for the precise configuration of actors around a given repository to change or mutate in some way. New actors might be added; relations might appear between entities; the location of entities might shift, and forms of association ('organizations') might subsume or grow out or around all of this.

Numbers such as 29 million or 12 million capitalize the capillary flow of associations in and through named entities as repositories and people. They also suggest the possibility of navigating the capillary flow differently. A vision of labile traceability is certainly offered in the Github data. Does the public availability of data produced and used in associative practices allow us to re-count capital numbers differently?

Imbuing numbers with importance

A database purporting to contain the whole associative fabric of Github suddenly appeared in mid-2012. Ilya Grigorik, a 'Web Performance Engineer' at Google, launched a Github repository `igrigorik/githubarchive` linked to a website `GithubArchive.org` dedicated to amalgamating all the Github API

[3] In what follows, I use monospace font like `this` when referring to entities named in the Github data. The typography highlights entity subject to enumeration by capital numbers.

public event data – the so-called 'timeline' – in one place (Grigorik, 2012). Grigorik, or `igrigorik`not only published all the data in a cloud-based data store but transformed that data, whose formatting we have glimpsed above, into the flat tabular forms familiar in much statistical work, (Campbell-Kelly, 2003b) and made it available through Google's newly launched cloud computing service, GoogleBigQuery.[4] The Github timeline data was listed, along with all the words in Shakespeare and all the US birth name records, as one of three dataset exemplars that people could use to learn about GoogleBigQuery (Google, 2016). Like the data on GithubArchive itself, the GoogleBigQuery copy of the Github public timeline data was updated hourly. Reaching back to 2012, it comprises around 290 million public events.[5] Table 1 shows the kinds of categorizations, valuing and aggregations of practice that become visible through this dataset. The table results from the simple query `select type, count(type) as events from [githubarchive:github.timeline] group by type order by events descending` and such queries are typical of the enumerative practices that such data attracts.

type	events
PushEvent	140,739,368
CreateEvent	35,483,741
WatchEvent	26,101,411
IssueCommentEvent	25,056,189
IssuesEvent	16,166,541
PullRequestEvent	11,208,920
ForkEvent	10,037,523
GistEvent	4,816,399
GollumEvent	4,253,087
DeleteEvent	3,680,053
FollowEvent	3,435,804
PullRequestReviewCommentEvent	3,066,117
CommitCommentEvent	2,493,741
MemberEvent	1,492,529
ReleaseEvent	418,180
DownloadEvent	302,247

[4]GoogleBigQuery, it seems, is the commercialisation of an internal infrastructure called `Dremel`, that Google has since 2006 used in many different ways to manage its own platforms. - Analysis of crawled web documents. - Tracking install data for applications on Android Market. - Crash reporting for Google products. - OCR results from Google Books. - Spam analysis. - Debugging of map tiles on Google Maps. - Tablet migrations in managed Bigtable instances. - Results of tests run on Google's distributed build system. - Disk I/O statistics for hundreds of thousands of disks. - Resource monitoring for jobs run in Google's data centers. - Symbols and dependencies in Google's codebase. (Melnik, Gubarev, Long, et al., 2010)

[5]In late 2015, the main GithubArchive timeline dataset was frozen. The last events date from October 2015. Data after that date flows into new datasets named by the month. e.g. https://bigquery.cloud.google.com/table/githubarchive:month.201107 points to the data for July 2011. This re-tabling of the data attests partly to the increasing popularity of Github. For the purposes of my argument, and to simplify code slightly, I only make use of the main Github timeline dataset covering 2012-2015.

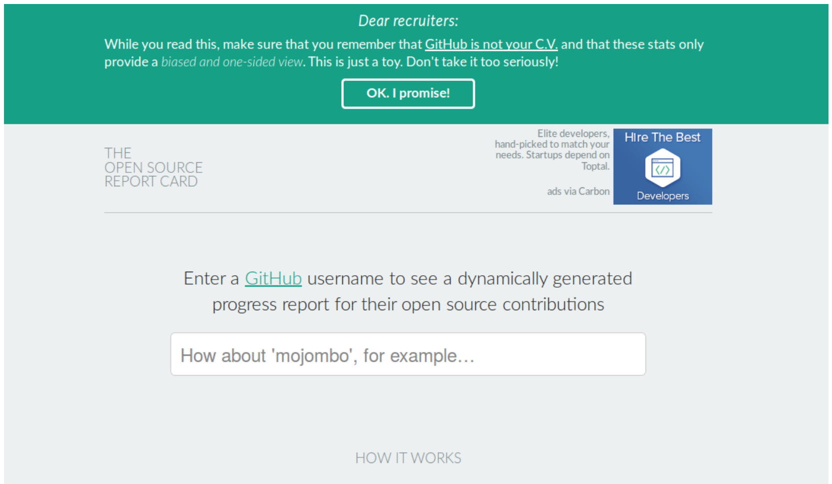

Figure 1: The Open Source Report Card

type	events
PublicEvent	261,821
TeamAddEvent	175,909
ForkApplyEvent	5,628
Total	289,195,208

The archiving, copying and transformation of Github event data drew the attention many people, including Github itself. A succession of 'Github Data Challenges' ensued in the next three years (2012-2014) The fact Github itself organises 'data challenges' (Github, 2013) suggests that people doing things with this data is part and parcel of the public life and publicity of the platform or device. These data challenges are what Anne-Lise Riles calls 'momentary apprehensions of depth' showed that drawing networks has long been part of networks (Riles, 2001: 184). Whatever the reason for these data challenges (they are increasingly common in many places today in the form of hackathons, data competitions, etc.), they occasion practical re-countings of the data that imbue the capital numbers with greater mobility, value and media or inscriptive form.

For instance, the 'OpenSource Report Card' (http://osrc.dfm.io/) or `dfm/osrc` by Dan Foreman-Mackay (Foreman-Mackay, 2014), is a prize-winning use of the timeline data (see Figure 1). It ingests all the data from the Githubarchive, counting what developers do, when they do it, and using what programming languages. With this data stored, it then builds a predictive model that allows it to both characterise a given Github user, and to predict who that Github user might have affinities with. Here the mass of events in the Github timeline are brought to bear on finding similarities between people. An admonition from Foreman-Mackay – 'Dear recruiters: While you read this, make sure that you remember that GitHub is not your C.V. and that these stats only provide a biased and one-sided view. This is just a toy. Don't take it too seriously!' – suggests that even playful application of the data attest to its capitalisation. For coders, programmers and software developers, the profile of their contributions to Github repositories can be part of getting work.

In response to the Github Data Challenge in 2012, people looked for feelings or 'sentiments' in the timeline data. Feelings associated with coding were mined by counting emotional words present in comments accompanying the Github events (http://geeksta.net/geeklog/exploring-expressions-emotions-github-commit-messages/). The presence of words in these message can be cross-linked with programming languages in order to profile how different programming languages elicit different emotional reactions. This emphasises how software developers or coders feel in relation to different kinds of work. Certain languages attract exasperation and others pleasure (see (Coleman, 2012) for a broader account of these feelings). The enumeration of affects around code animates the capital numbers with an underpinning affectivity associated with coding in both its vicissitudes and its unevenly shared privileges of participation and control (Chun, 2011).

The capital numbers associated with Github also begin to acquire a geography through the GithubArchive data. People began to map coders and repositories by geographic location. The mapping of Github contributions by location performed by David Fischer (http://davidfischer.github.io/gdc2/#languages/All) is typical in that it too counts events, but this time puts the emphasis on the geography of the 'top' repositories, coders and their programming languages. As the David Fischer puts it 'this data set contains contributions to the top 200 GitHub repositories during the first four months of 2013 and plots the location based on what the contributor provided' (Fischer, 2013).

People also make live dashboards for Github. `Octoboard` (http://octoboard.com/) animates changes on Github using the timeline data (Roussell, 2015) (see Figure 2). `Octoboard` ornaments the capital numbers with a range of peripheral live enumerations that point to the liveliness of events on Github. It presents a summary of daily activity in major categories on Github – how many new repositories, how many issues, how repositories have been 'open sourced' today. It offers realtime analytics on emotions. Like many other dashboards associated with social media analytics, `octoboard` suggests that the constant change in associations and projects in software development can no longer be known through leisurely rhythms of analysis, but is increasingly framed as a problem realtime awareness. Significant shifts, trends, hotspots of activity, improbably important marginal developments or breakdowns need to be brought into immediate view through 'stream analytics.' That is, not also does the data stream, but the analysis is meant to stream as well in order to be timely, lively and responsive to change.

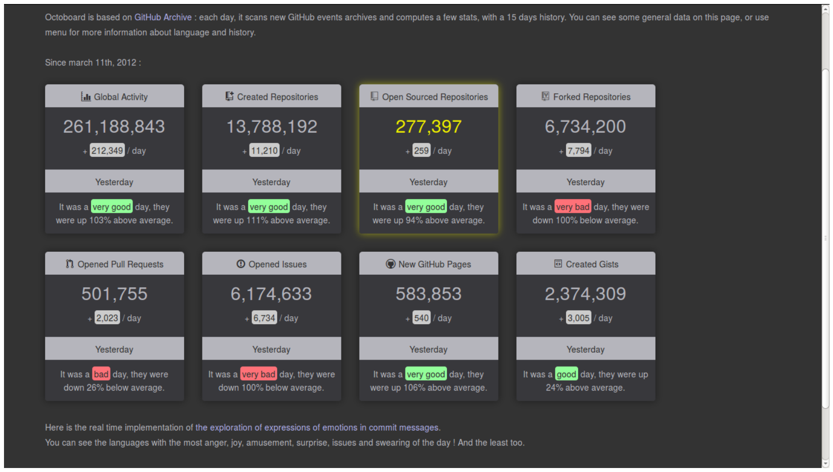

Figure 2: Octoboard: Github activity dashboard

Looking slightly more widely, the Github timeline data has quickly become a favourite training tool for data mining textbooks books that configure and convey the calculative agencies characteristic of capital numbers. In *Mining the Social Web: Data Mining Facebook, Twitter, LinkedIn, Google+, GitHub, and More,* Matthew Russell makes use of the Github timeline to demonstrate ways of uses social network analysis to highlight the important nodes and links between repositories and users (Russell, 2013). Again, the propensity to apply network analysis approaches is widespread and endemic to the data itself, given the way that the event format is already implicitly framed by a network or 'social media' understanding.

Finally, for academic researchers in computer science and certain parts of organisation studies, Github has been a boon because they study technologically and economically important practices software development in the wild much more easily. Academic researchers in fields such as software engineering do social network analysis in order to gauge productivity, reuse, efficiency and other engineering and management concern (Thung, Bissyandé, Lo, et al., 2013). Like the many Github-hosted projects discussed above, they analyse sentiment (Guzman, Azócar & Li, 2014), collaboration and productivity (Dabbish, Stuart, Tsay, et al., 2012), and geography (Takhteyev & Hilts, 2010).

The diagramming, counting, measuring and statements generated by the many different treatments of the Github timeline data, particularly although not always in its transformation to a 'big data' queryable dataset on GoogleBigQuery, affect the capital numbers that Github enumerates. They enliven, animate, reactivate, localise and qualify those the numbers in ways that section off fragments from the 'natural multiplicity' (Badiou, 2008: 211) of any aggregate. They potentialise the numbers in terms of expansiveness, liveness and further accumulation by summing them up in different ways (realtime, networks of connections, geographies of work and affect), and by performing acts of calculation (social network analysis, machine learning classification and prediction). Github's capital numbers become a point of enucleation for calculative agencies (Callon & Muniesa, 2005: 1245) associated with the calculable goods embodied in code and software.

Acts of infrastructural imitation

The calculative and diagrammatic projects centred on the Github data in some ways resonate with recent STS-orientations to numbers and enumerations more generally. Recent ANT work, such as Latour, Jensen, Venturini et. al's 'The Whole is Smaller than the Part: How Digital Navigation May Modify Social Theory,' argues that given a 'sudden proliferation of digital databases' (Latour, Jensen, Venturini, et al., 2012: 18) and tools that allow us to move between different scales and locations in social fields, we no longer need the long-standing structural oppositions between structure and agency, between macro and micro. We instead begin to inhabit a space of continuously transforming scales in which every given standpoint stabilises scales and boundaries of the whole for a particular actor in a perspectival variation. They too calculatively synthesize diagrams that invite visual navigation to identify the ebb and flow of associations: 'it might now be possible to account', write Latour, Jensen et. al 'for longer lasting features of social order by learning to navigate through overlapping 'monads' instead of alternating between the two levels of individual and aggregate' (Latour, Jensen, Venturini, et al., 2012: 3). While the capital numbers consistently split social order between people and things, subject and objects, the data visualizations offer different levels of navigation, sometimes sub-individual ('sentiment') and sometimes supra-individual (global geographies of coders and repositories). The calculated diagrammatic enumerations function 'oligoptically': they offer a limited view of a whole (Latour, Jensen, Venturini, et al., 2012: 11). While capital numbers emphasise the whole, STS accounts of the making of views of the whole through enumeration attest to an always partial grasp of the whole, to the calculative stabilisation and discretisation of entities in a single frames.

How could we re-count capital numbers in order to both further highlight their dynamic composition through flows of association and to highlight some of the highly reactive, performative work they do in enumerating commercial, financial and social value? Could those numbers in both their compositional and imaginary-materialising configurative multiplicity be counted? Re-counting of capital numbers might begin with the seemingly trivially abundant yet absolutely basic entity in code and on Github more generally, names. As seen in the sample event above, many entities in the Github timeline data and on the platform more generally are named. Names refer to people, organisations, repositories, platforms, code libraries or packages, and a very wide range of devices and places. If Github's capital numbers count people and repositories ('code'),

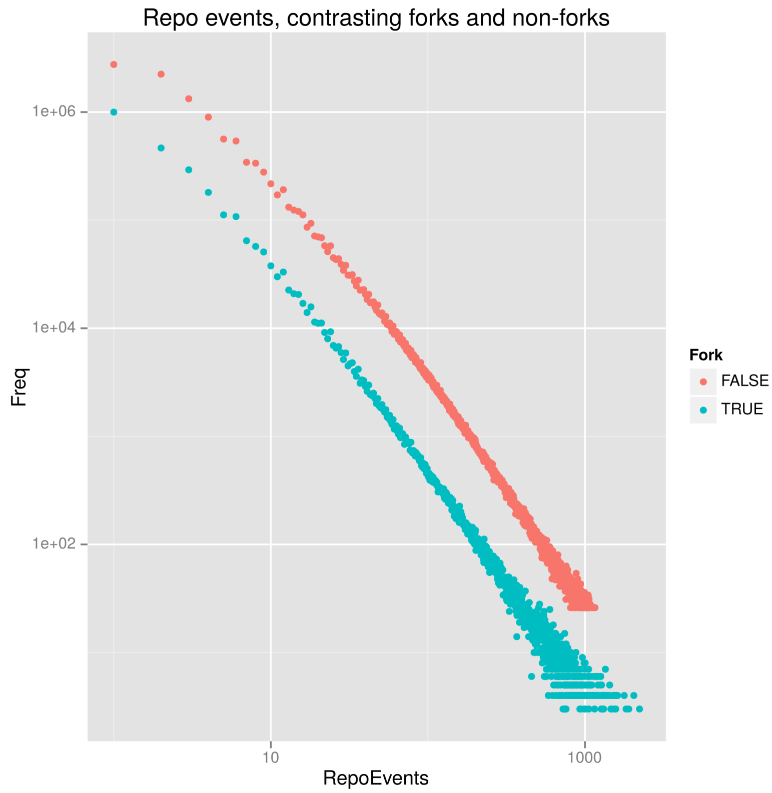

Figure 3: The power of events over time

then what would happen if we counted in the timeline data as published on GoogleBigQuery associations between people and repositories by tracking the popularity of different names?[6]

Such a counting, undertaken in a very crude but heavily computational fashion, appears in Figure 3. This simple graphic suggests that the vast majority of repositories on Github are highly ephemeral. They attract one or two events. A small number of repositories, counted on the lower right-hand side of the plot capture many thousand events.

A great proportion of events in the timeline data were absorbed into ephemeral repositories (if that is not too great a contradiction in terms). Millions of repositories flash into visibility on the timeline for a brief period before falling back

[6]A host of methodological questions cluster around the question of counting names. For the purposes of this discussion, I leave aside the many sidetracks, dead-ends and catastrophic failures associated with working on a several hundred gigabyte dataset. Rather than pursuing statistical accuracy, I am concerned with the practical calculative agency of STS research in such settings.

into obscurity. The diagram shown in the Figure 3 encapsulates this imitative flux. On the left hand side, millions of repositories receive less than five events during the 18 months. On the right hand side, less than 50 repositories receive more than a thousand events. A similar pattern appears in the other capital number: while some 'people' (a person is not always easy to distinguish from a robot on Github, and some people appear to organise the work of many others) emit many thousands of events, others only trigger a few.[7] Already, then, the capital number of repositories on Github takes on a different complexion.

Given that repositories are receptacles for all the work of software developers and coders, it seems strange that the vast bulk of them are so ephemeral. But rather than see this ephemerality as waste, noise or something to be discarded, we might trace the working of associative processes through it. Tracking the patterns of imitation, rather than looking for the events that define repositories as important, allows something different to appear.

Every event in Table 1 (Events in the Github timeline) names at least one sometimes two repositories, and at least one actor who might be a person or who might be a robot acting like a person. Examining just the names of repositories in the timeline data, counting how often and when they appear in events begins to show how associations are configured on Github. In particular, the movement of waves of imitation begins to show something of the dynamics that animate the growth of capital numbers. These take two forms in the event data. We have already that almost one half of the repositories in Github trigger only one event in the timeline data stream. What is this event? Many repositories arise from the act of copying ('forking') another repository. People 'fork' other repositories frequently. Table 1 shows 10 million ForkEvents, suggesting that around one third of the total 29 million repositories are direct copies. Acts of copying occur on many scales and at various levels of infrastructural and associative complexity. This copying is vital to the 'sharing' practice of Github coding. Of the 290 million events in the timeline dataset, approximately 95 million result from copying, watching, commenting on or directly participating in other repositories. (The 140 million PushEvents shown in Table 1 might also be highly imitative, but their imitative character is less immediately visible in the timeline data.)

The figures shown above show some localised aspects of this propagation of imitation. Both Figure 4 and 5 count imitations in the form of copying code, but in ways that go beyond the formal copying mechanisms offered by Github itself (for instance in the 'Fork' button that appears on the top right hand side of any repository on Github; for example https://github.com/twbs/bootstrap). Imitations appears in two different ways in these figures. The broad bands of color rippling horizontally across the middle of the figures graph the counts of copies being made each day on Github of popular repositories using the Fork

[7]In all this re-counting, the data relates to the public repositories and actors on Github. An unknown but substantial number of repositories are private. Given Github's financial valuation at $US 2 billion in 2015, this number may be quite large. (Gage, 2015)

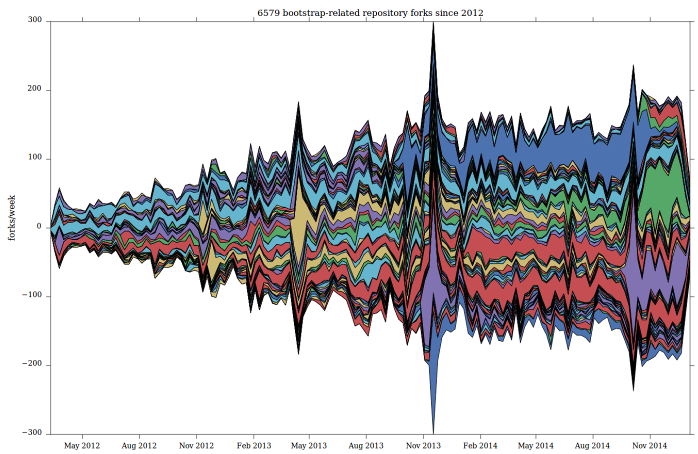

Figure 4: ‘bootstrap‘ repository forks

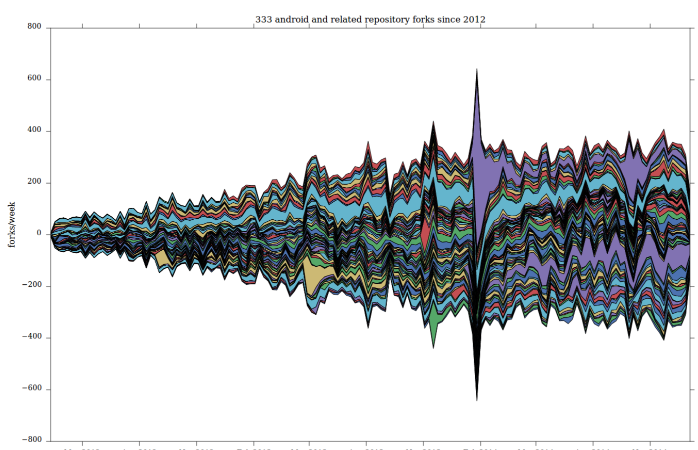

Figure 5: ’android‘ repository forks

action. (In forking, the repository name remains the same.) But the much more dense striations on either side of the central bands, seen for instance in Figure 5, count repositories whose names incorporate the important repository, but vary it in some way. These repositories have some association with the `android` repository, but diverge from it in a multiplicity of ways. A repository may, for instance, relate to the popular Twitter `bootstrap` repository yet combine it with a range of other platforms and devices such as `android` or `jQuery`. Visually, these marginal striations in the diagram still amount to imitations, and they generate large numbers of events in the timeline dataset, but less homogeneously, and in greater multiplicity than the official copying mechanisms configured by Github itself as a social media platform.

In striated zones of associative imitation, it is possible to count something different. Imitations augment the average everydayness of coding work with forms of associative investment and affective identification. In this work, important repositories act as high visibility markers around which forms of identity take hold and multiply. Unlike the capital numbers with their total aggregates of people or repositories, the imitative fluxes that bulk out and pad in these diagrams have diverse network-making configurations. As relatively recent accounts of invention drawing on Gabriele Tarde's work have suggested, passionate imitation provides a powerful account of the operation of markets, economies, technical and expert knowledges and media (Barry & Thrift, 2007; Barry, 2005; Borch, 2005; Latour & Lepinay, 2008). It may be that the flows of varying imitation we see in these figures amount to something like 'encounters and interferences of imitative rays' (Latour & Lepinay, 2008: 530) on Github. Imitative repetition is write large here as the primary social process (Tarde, 1902: 20) or as the ineluctable configuration of any sociality (G Tarde, 1895: 11), generating both habits and stable organisations that might solidify into devices or systems, or might precipitate the restless dynamisms that capital numbers figure as an accumulation or a resource. In the recombination of different names, we see some of 'the network effects that come from the imitative dynamics of actors' (Arvidsson, 2011: 141) or the 'multiplicity of imitative fluxes' (Lazzarato, 2002: 132). These figures suggest a way of re-counting capital numbers and their tendency to expand. This re-counting entails requires a form of calculative agency that works to re-animate the numbers in terms of the 'repetitive imitations' (G de Tarde, 1895: 25) that weave the associative fabric of social process through, in this case, the naming and invocation of named entities.

De-capitalising numbers

The quasi-Tardean counting of names in their repetition and variation has its own limitations or configurative specificity. A purely imitative understanding of the growth of capital numbers does not help us grasp how these numbers or figures 'populate our sociomaterial imaginaries' (Suchman, 2012: 58) – and *vice versa.* The very act of counting these imitations has a device-specificity, in its reliance on the Github platform API, the GithubArchive timeline dataset and the GoogleBigQuery cloud analytic's platform, and indeed Github itself as the repository of the code and text comprising this article (Metacommunities, 2016). Counting these associations, I am suggesting, differs from navigating a path through a network, as recent work on traceability or Tardean-inspired accounts of methodological transformation through database traceability might suggest (Barry & Thrift, 2007: 517). Unlike the network-tracing approach, it can perhaps take into account the ways in which the platform itself works to shape networked associations. Whenever we count or calculate in any form, device-specific configurations materialise in the numbers. The expansively contagious nature of imitation inevitably encounters device-specific configurations.

One sense of how numbers embody device-specific configurations can be seen by examining high event-count repositories on Github. The top 1000 repositories attract around 16% of all events in the timeline period. The query de-capitalises the repository names for the purposes of aggregation (e.g. a repository called `DotFile` will be counted along with `dotfile`). Given the flattening of the names to lower case produced by this query, we cannot readily see how those events are distributed. But these highly frequented repositories, which only comprise a tiny percentage of the 29 million on Github, absorb many events. What happens in these high event count repositories?

Several major traits appear in these high event-count repositories. Many events come from off-platform. Github, it seems, functions as a mirror for other sites and places. For instance, the repository `eclipse.platform.common` accounts for almost 2 million events in the timeline data. A single repository attracting two million events (or almost 1% of the total event count in the timeline data) suggests something highly significant in the geography of coding work. Perhaps the fact that `eclipse` is itself part of the Eclipse Foundation, 'an amazing open source community of Tools, Projects and Collaborative Working Groups' (Eclipse Foundation, 2016) with almost a thousand of its own projects, might help explain the large number of events. More significantly, the high-frequency event trade in `eclipse.platform.common` is an example of how Github itself functions as part of a configuration. The `eclipse` repositories are not actively developed on Github. They are mirrored or themselves copied from the hundreds of `git` repositories found at (git://git.eclipse.org)[git://git.eclipse.org]. The work and much of the associative fabric of `eclipse` does not reside on Github, but is copied there.[8] The many events attracted by `eclipse.platform.com` are copies of associative acts occurring elsewhere. Like many other significant repositories (`linux`, `android`, `mozilla`, `apache`), high event counts attest here to Github's function as a capital. At the same time, their visible presence in the timeline data suggests that we might also re-count the large capital numbers of repositories and 'people' in terms of what in them is a mirroring or reflection of

[8]The presence of `eclipse` on Github suggests another important analytical problem to consider when enumerating: how do know for a given setting or place that what we count belongs there? Sometime repositories on Github mention they are 'mirrors' of something else. Other times they may not.

things happening elsewhere.

In addition, even where events do originate on Github, many of them concern Github or technical configurations closely associated with Github. For instance, the list of names of high-event count repositories includes many containing a set of terms centred on a few highly repeated items. On the one hand, the sharing of names – for instance, the many repositories that contain the term `dot` as in `dotfile` or `vimdot` – suggests the imitation or repetition that Tarde describes. On the other hand, terms like `dot`, `test`, `hello`, `config`, `doc`, `build`, `setting` or `demo` also stream through this set of repository names. Many repository names proliferate from these stems. All of these stems relate to device-specific configurations and work done to sustain the configuration of devices in particular places.

The proportion of events absorbed by repositories explicitly concerned with configuration is substantial. Of the 47173750 events, 18% belong to configurative repositories. These repositories explicitly relate to configuration work done as part of coding. A `dotfile` repository will typically contain the settings for a software developer's terminal windows, code editor, and 'shell' or command line interface (see the heavily forked `[https://github.com/mathiasbynens/dotfiles](https://github.com/mathiasbynens/dotfiles)` repository for typical contents (Bynens, 2016)). The name `dotfile` refers to the fact that some operating systems hide from normal view file names beginning with the character `.`. Many of these explicit configurative repositories stand quite close to personal actors and their configured work settings. The actor `mathiasbynens` maintains configuration relating to his coding in the repository `dotfiles.`

The substantial proportion of these very commonly used repository names concerned either with direct configuration work or with the development of configurative devices (for instance the `eclipse` platform is a commonly used coding environment) highlights the operation of device-specific practices within the fluxes of imitation that appear more generally around places like Github.[9]

Conclusion

> Number is neither that which counts, nor that which we count. This regime of numericality organises the forgetting of number. (Badiou, 2008: 51)

I have been describing a regime of numericality associated with software development and with social media more generally. Capital numbers, I have suggested,

[9] We could look further for implicit configurative work in the timeline events. A query on the timeline dataset yields around ten million push events containing either a comment or message relating to 'config,' 'test' or 'build.' Similarly, a query for repositories that use no programming language results in around eleven million repositories. While null-code repositories vary greatly in their contents, many of them act as stores for configuration-related documentation, settings and data files.

enumerate an aggregate, imbue it with importance linked to infrastructural scale and complexity, and promise to generate further accumulation. As we have seen, the capital numbers attract many further aggregations in the form of data archives (GithubArchive) and oligoptic visualization.

It takes work to get contemporary digital data and associated large numbers to do something other than feed the numericality of capital numbers and their platform-centred aggregates. STS research might develop stronger ethnographic sensibilities in relation to data event streams (such as the Github timeline) and capital numbers by moving in an ambivalent space between visualization-friendly data formats and awkward device-specific configurations. Configurative numbers, I have suggested, might be one term for the versions of enumeration that result from such re-countings. We saw how that re-counting points to imitative fluxes, to that which takes places elsewhere but is refracted in the capital numbers, and to device-specific configurations that saturate the data stream. None of this discounts the capital numbers – 29 million, 12 million – as if they were false, reactive or performative numbers (Espeland & Stevens, 2008).

There are several different ways to understand what is at stake in re-counting configurative numbers. It differs from the network-based approaches of some recent STS work using digital methods. Latour, Jensen, and Venturini ask:

> Is it possible to do justice to … common experience by shifting from prediction and simulation to description and data mining? … Let the agents produce a dynamics and collect the traces that their actions leave as they unfold so as to produce a rich data set' (Latour, Jensen, Venturini, et al., 2012: 13)

Despite the immensely useful insight that parts lend themselves to simpler wholes, there are hidden difficulties in their recommendation. Any description that ignores that investment in formatted data potentially strengthens the growth of capital numbers with their hub-like tendencies. As we have seen, what they call the 'rich data set' is pre-formatted for particular actors and agents (the platform usually), and receives much attention in relation to the capital numbers.

If we are interested in alternative calculative agencies, we have to work against the traces in some ways. Verran's observation about work on numbers might be a more productive starting point. As she says, 'in any practical going-on with numbers, what matters is that they can be made to work, and making them work is a politics' (Verran, 2001). The massive availability of data from online databases is power-laden number work. Its manifest and much vaunted always partial openness (for instance, the Github timeline dataset does not include the very many private repositories) allows a certain kind of capitalising and governing based on data to expand. The formatting of data streams tends to confirm and expand how the platform already advertises itself.

How do we re-animate the large numbers that abound in contemporary sciences and technologies? To do this, we need to find ways of working on formatted data that goes against its pre-formatted accounting. Device-specific research that attends to the ways in which data is formatted, and how that formatting itself materialises imaginaries of inclusion, relation and wholes that pertain more broadly. At the same time, paying attention the configuration or device-specific composition of numbers, we can construct a view on what remains irreducible to the formatted imaginaries of the data-stream.

The calculative agency of STS and the implicit cuts of the multiple that numbering entails directly contest the value-measures associated with certain large numbers, showing that capital numbers hide many inflationary, distributed and configurative processes. In counting repositories in terms of names and their variations, the configurative numbers presented in the preceding discussion deactivate any essential link between working with digital data and ontological invention, suggesting instead that any ontological epistemic transformation associated with significantly large numbers – for instance, in claims about the epistemic potentials of data – pivots on the crafting of new calculative agencies for fragile, weak and diffuse collectives.

Acknowledgments

The research reported on here was funded by ES/K007912/1 'Metacommunities of practice in the code-sharing commons' XXX – removed for anonymity –. Other researchers on this project were x,x,x,x.

References

Arvidsson A (2011) General Sentiment: How value and affect converge in the information economy. *The Sociological Review*, 59, 39–59, Available from: http://onlinelibrary.wiley.com.ezproxy.lancs.ac.uk/doi/10.1111/j.1467-954X.2012.02052.x/abstract (accessed 22 November 2013).

Badiou A (2008) *Number and Numbers*. Cambridge, Polity Press.

Barry A (2005) Pharmaceutical Matters: The Invention of Informed Materials. *Theory, Culture and Society*, 22(1), 51–69.

Barry A and Thrift N (2007) Gabriel Tarde: Imitation, invention and economy. *Economy and Society*, 36(4), 509–525.

Borch C (2005) Urban Imitations: Tarde's Sociology Revisited. *Theory Culture Society*, 22(3), 81–100.

Bowker GC, Baker K, Millerand F, et al. (2010) Toward information infrastructure studies: Ways of knowing in a networked environment. In: *International handbook of internet research*, Springer, pp. 97–117, Available from: http:

//link.springer.com/chapter/10.1007/978-1-4020-9789-8_5 (accessed 9 January 2015).

Bucher T (2013) Objects of Intense Feeling: The Case of the Twitter API : Computational Culture. *Computational Culture*, (3), Available from: http://computationalculture.net/article/objects-of-intense-feeling-the-case-of-the-twitter-api (accessed 16 March 2016).

Burrows R (2012) Living with the h-index? Metric assemblages in the contemporary academy. *The Sociological Review*, 60(2), 355–372, Available from: http://onlinelibrary.wiley.com/doi/10.1111/j.1467-954X.2012.02077.x/full (accessed 17 February 2016).

Bynens M (2016) Mathiasbynens/dotfiles. GitHub. Available from: https://github.com/mathiasbynens/dotfiles (accessed 15 March 2016).

Callon M and Muniesa F (2005) Peripheral Vision: Economic Markets as Calculative Collective Devices. *Organization Studies*, 26(8), 1229–1250, Available from: http://oss.sagepub.com/cgi/content/abstract/26/8/1229.

Callon M, Courtial JP and Laville F (1991) Co-word analysis as a tool for describing the network of interactions between basic and technological research: The case of polymer chemsitry. *Scientometrics*, 22(1), 155–205, Available from: http://link.springer.com.ezproxy.lancs.ac.uk/article/10.1007/BF02019280 (accessed 3 September 2013).

Campbell-Kelly M (2003a) *From Airline Reservations to Sonic the Hedgehog: A History of the Software Industry*. Cambridge, MA, MIT Press.

Campbell-Kelly M (2003b) *The history of mathematical tables*. Oxford University Press.

Chun W (2011) *Programmed visions: Software and memory*. The MIT Press.

Coleman G (2012) *Coding freedom: The ethics and aesthetics of hacking*. Princeton N.J., Princeton University Press.

Dabbish L, Stuart C, Tsay J, et al. (2012) Social coding in GitHub: Transparency and collaboration in an open software repository. In: *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work*, ACM, pp. 1277–1286, Available from: http://dl.acm.org/citation.cfm?id=2145396 (accessed 23 May 2014).

Desrosières A (2011) Words and numbers: For a sociology of the statistical argument. In: *The Mutual Construction of Statistics and Society*, pp. 41–63.

Diggle PJ (2015) Statistics: A data science for the 21st century. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 178(4), 793–813, Available from: http://onlinelibrary.wiley.com/doi/10.1111/rssa.12132/abstract (accessed 2 December 2015).

Eclipse Foundation (2016) Eclipse - The Eclipse Foundation open source community website. Available from: https://www.eclipse.org/ (accessed 3 March

2016).

Ensmenger N (2010) *The Computer Boys Take Over: Computers, Programmers, and the Politics of Technical Expertise*. MIT Press, Available from: http://www.jstor.org.ezproxy.lancs.ac.uk/stable/j.ctt5hhjdh (accessed 6 March 2016).

Espeland WN and Stevens ML (2008) A Sociology of Quantification. *European Journal of Sociology / Archives Européennes de Sociologie*, 49(03), 401–436.

Fischer D (2013) Open source contributions by location. Available from: http://davidfischer.github.io/gdc2/#languages/All (accessed 15 March 2016).

Foreman-Mackay D (2014) The Open Source Report Card. Available from: https://web.archive.org/web/20140214105201/http://osrc.dfm.io/ (accessed 11 March 2016).

Gage D (2015) GitHub Raises 250 Million at 2 Billion Valuation. Wall street journal. Available from: http://www.wsj.com/articles/github-raises-250-million-at-2-billion-valuation-1438206 (accessed 26 February 2016).

Gerlitz C and Lury C (2014) Social media and self-evaluating assemblages: On numbers, orderings and values. *Distinktion: Journal of Social Theory*, 15(2), 174–188, Available from: http://dx.doi.org/10.1080/1600910X.2014.920267 (accessed 25 January 2016).

Gillespie T (2010) The politics of 'platforms'. *New Media & Society*, 12(3), 347–364, Available from: http://nms.sagepub.com/cgi/doi/10.1177/1461444809342738 (accessed 5 September 2014).

Git (2014) Git. Available from: http://git-scm.com/ (accessed 25 February 2014).

Gitelman L (2013) *'Raw Data' is an Oxymoron*. Cambridge, Massachusetts ; London, England, MIT Press.

Github (2015) About · GitHub. Available from: https://web.archive.org/web/20151216055610/https://github.com/about (accessed 16 March 2016).

Github (2014a) About · GitHub. Available from: http://web.archive.org/web/20140605120710/https://github.com/about (accessed 16 March 2016).

Github (2013) Data Challenge II Results. GitHub. Available from: https://github.com/blog/1544-data-challenge-ii-results (accessed 15 March 2016).

Github (2016) Github Events Application Programmer Interface. Available from: https://api.github.com/events (accessed 15 March 2016).

Github (2012) GitHub · Social Coding. Available from: http://wayback.archive.org/web/20121127113950/https://github.com/ (accessed 15 March 2016).

Github (2014b) Results of the GitHub Investigation. GitHub. Available from: https://github.com/blog/1823-results-of-the-github-investigation (accessed 16

March 2016).

Google (2016) Github timeline data on Google BigQuery. Available from: https://bigquery.cloud.google.com/table/githubarchive:github.timeline (accessed 29 February 2016).

Grigorik I (2012) GitHub Archive. Available from: http://www.githubarchive.org/ (accessed 31 March 2014).

Guyer JI (2014) Percentages and perchance: Archaic forms in the twenty-first century. *Distinktion: Journal of Social Theory*, 15(2), 155–173, Available from: http://dx.doi.org/10.1080/1600910X.2014.920268 (accessed 25 January 2016).

Guzman E, Azócar D and Li Y (2014) Sentiment Analysis of Commit Comments in GitHub: An Empirical Study. In: *Proceedings of the 11th Working Conference on Mining Software Repositories*, MSR 2014, New York, NY, USA, ACM, pp. 352–355, Available from: http://doi.acm.org/10.1145/2597073.2597118 (accessed 25 February 2016).

Hacking I (1990) *The taming of chance*. Cambridge University Press.

Halpern O (2015) *Beautiful Data*. Durham, N.C, Duke University Press.

Hardy Q (2012) Dreams of 'Open' Everything. Bits blog. Available from: http://bits.blogs.nytimes.com/2012/12/28/github-has-big-dreams-for-open-source-software-and-more/ (accessed 27 March 2014).

Jenkins H (2004) The cultural logic of media convergence. *International journal of cultural studies*, 7(1), 33–43, Available from: http://ics.sagepub.com/content/7/1/33.short (accessed 9 December 2013).

Kelty C (2008) *Two Bits: The Cultural Significance of Free Software*. Duke University Press.

Kitchin R (2014) Big Data, new epistemologies and paradigm shifts. *Big Data & Society*, 1(1), 2053951714528481, Available from: http://bds.sagepub.com/content/1/1/2053951714528481 (accessed 19 February 2015).

Kuhn TS (1996) *The structure of scientific revolutions*. Chicago, IL, University of Chicago Press.

Lampland M (2010) False numbers as formalizing practices. *Social Studies of Science*, 40(3), 377–404.

Latour B (1996) *Aramis, or the love of technology*. Cambridge, MA & London, Harvard University Press.

Latour B (1987) *Science in action : How to follow scientists and engineers through society*. Cambridge, Mass., Harvard University Press.

Latour B and Lepinay V (2008) *L'Economie*.

Latour B, Jensen P, Venturini T, et al. (2012) The Whole is Always Smaller than its Parts. How Digital Navigation May Modify Social Theory. *British*

Journal of Sociology, 63(4), 590–615.

Lazzarato M (2002) *Puissances de l'invention : La psychologie âeconomique de Gabriel Tarde contre l'âeconomie politique.* Paris, Empãecheurs de penser en rond.

Leydesdorff L (1998) Theories of citation? *Scientometrics*, 43(1), 5–25, Available from: http://www.springerlink.com/index/6W021473716M78X8.pdf (accessed 8 October 2012).

Leydesdorff L, Carley S and Rafols I (2013) Global maps of science based on the new Web-of-Science categories. *Scientometrics*, 94(2), 589–593, Available from: http://link.springer.com/article/10.1007/s11192-012-0784-8 (accessed 8 March 2016).

Mackenzie A (2016) Combinatorial infrastructures. In: Harvey P, Jensen CB, and Atsuro M (eds), *Infrastructures and Social Complexity*, London & New York, Routledge.

Mackenzie A (1997) Undecidability: The history and time of the Universal Turing Machine. *Configurations*, 3, 359–379.

Mackenzie A, Fuller M, Goffey A, et al. (2016) Code repositories as expressions of urban life. In: Kitchin R (ed.), *Code and the City*, London, Routledge.

Marczak B and Weaver N (2015–2015-04-10T05:55:18+00:00) China's Great Cannon. The citizen lab. Available from: https://citizenlab.org/2015/04/chinas-great-cannon/ (accessed 16 March 2016).

Marres N, Gerlitz C and others (2015) Interface Methods: Renegotiating relations between digital social research, STS and sociology. *Sociological Review*, Available from: http://research.gold.ac.uk/11343/ (accessed 30 July 2015).

Melnik S, Gubarev A, Long JJ, et al. (2010) Dremel: Interactive Analysis of Web-Scale Datasets. In: *Proc. of the 36th Int'l Conf on Very Large Data Bases*, pp. 330–339, Available from: http://www.vldb2010.org/accept.htm.

Metacommunities (2016) Metacommunities/metacommunities. GitHub. Available from: https://github.com/metacommunities/metacommunities (accessed 18 December 2014).

Meyer R (2013) Github, Object of Nerd Love, Makes Play for Non-Programmers. The atlantic. Available from: http://www.theatlantic.com/technology/archive/2013/08/github-object-of-nerd-love-makes-play-for-non-programmers/278971/ (accessed 13 February 2014).

Miller B (2009) What Does it Mean that PRIMES is in P? Popularization and Distortion Revisited. *Social Studies of Science*, 39(2), 257–288.

Mol A (2003) *The Body Multiple: Ontology in Medical Practice.* Durham, N.C, Duke University Press.

Nafus D (2014) Stuck data, dead data, and disloyal data: The stops and starts in making numbers into social practices. *Distinktion: Journal of Social The-*

ory, 15(2), 208–222, Available from: http://dx.doi.org/10.1080/1600910X.2014.920266 (accessed 25 January 2016).

Pentland A (2014) *Social Physics: How Good Ideas Spread—The Lessons from a New Science*. New York, Penguin Press HC, The.

Porter TM (2008) LOCATING THE DOMAIN OF CALCULATION. *Journal of Cultural Economy*, 1(1), 39–50, Available from: http://www.tandfonline.com/doi/abs/10.1080/17530350801913627.

Porter TM (1996) *Trust in numbers: The pursuit of objectivity in science and public life*. Princeton Univ Pr.

Riles A (2001) *The network inside out*. University of Michigan Press, Available from: http://books.google.co.uk/books?hl=en&lr=&id=iOu_LP1w2LEC&oi=fnd&pg=PR9&dq=annelise+riles+networks+2004&ots=SKkvqn85hP&sig=cCS3Jakbv4tK24ttTEzy2wlq7iE (accessed 9 June 2014).

Rogers R (2004) Information Politics on the Web. Available from: http://mitpress.mit.edu/catalog/item/default.asp?ttype=2&tid=10329 (accessed 17 July 2012).

Roussell D (2015) Octoboard. Github activity dashboard. Available from: https://web.archive.org/web/20150801193208/http://octoboard.com/ (accessed 11 March 2016).

Ruppert E, Law J and Savage M (2013) Reassembling Social Science Methods: The Challenge of Digital Devices. *Theory, Culture & Society*, 0263276413484941, Available from: http://tcs.sagepub.com/content/early/2013/05/13/0263276413484941 (accessed 20 November 2013).

Russell MA (2013) *Mining the Social Web: Data Mining Facebook, Twitter, LinkedIn, Google+, GitHub, and More*. ' O'Reilly Media, Inc.', Available from: http://books.google.co.uk/books?hl=en&lr=&id=_VkrAQAAQBAJ&oi=fnd&pg=PR4&dq=github&ots=JqiqtzTxmK&sig=sfea4ce1ue2XYt_dERD41VpSTS4 (accessed 23 May 2014).

Suchman L (2012) Configuration. In: Lury C and Wakeford N (eds), *Devices and the Happening of the Social*, Routledge, pp. 48–60.

Takhteyev Y and Hilts A (2010) *Investigating the geography of open source software through GitHub*. Available from: http://takhteyev.org/papers/Takhteyev-Hilts-2010.pdf (accessed 23 May 2014).

Tarde G de (1895) *La logique sociale*. Paris, F. Alcan.

Tarde G de (1902) *Psychologie économique*. Paris, F. Alcan.

Tarde G (1895) *Les lois de l'imitation: Étude sociologique*. 2nd ed. Available from: http://openlibrary.org/b/OL23311522M/lois_de_l%27imitation.

Thung F, Bissyandé TF, Lo D, et al. (2013) Network structure of social coding in GitHub. In: *Software Maintenance and Reengineering (CSMR), 2013 17th*

European Conference on, IEEE, pp. 323–326, Available from: http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=6498480 (accessed 23 May 2014).

Verran H (2001) *Science and An African Logic*. Chicago, London, The University of Chicago Press.

Verran H (2012) The changing lives of measures and values: From centre stage in the fading ‘disciplinary’ society to pervasive background instrument in the emergent ‘control’ society. *The Sociological Review*, 59, 60–72.