24

Digital sociology in the field of devices

Adrian Mackenzie, Richard Mills, Stuart Sharples, Matthew Fuller and Andrew Goffey

Introduction

Digital sociology focuses on culture as it plays out in the vast, expanding, power-laden and complex media environments of the last few decades. In these media environments, feedback loops running between devices and social practices constantly re-define culture. Culture – that which is lived in places such as cities, cafes, airports, streets, shops, museums, parks, clinics, offices or living rooms – is rapidly rewoven by transient device-specific play of signals passing through news, entertainment, advertising and social networking platforms such as Twitter, Instagram, LiveJournal, YouTube, LinkedIn or Weibo. Digital sociology addresses the problem of how to make sense of the signals generated, captured, organised, shared, and constantly sorted in form of hyperlinks, messages, transactions, text, and images flowing across media platforms by people living with and through digital devices (Lupton 2012). Digital social researchers seek to learn about the coherence, modes of thought and value, practices, materials and forms of contemporary experience and social action as they are drawn into what recent observers have called 'a massive, culturally saturated feedback loop' (Schutt and O'Neil 2013: 5) interlacing what people do and what they experience. None of the devices, practices and subjects of this form of the social are coherent, well-understood or stable. Indeed, these feedback loops predicate constant processing, adjustment, realignment, transformation, variation and mutation in social worlds. Amidst this targeting of transformation, much hinges on interactions in what we might call, following Ruppert, Law and Savage (2013), the *field of devices*. The field of devices is a complex weave of technical elements, more or less connected to each other through relations of contact and contest, convergence and divergence, similarity, imitation and variation. It is inhabited by people who react to, who experience and are affected by durable and transient calls to order their actions with and through devices.

While digital sociology draws on well-established methods such as ethnography, rhetorical, discourse and visual analysis, in the analysis of device-specific transformations in culture new skills and digital tools, borrowed or copied from domains of statistics, software development, hacking, graphic design, audio, video and photographic recording and predictive modelling – that is, from the media-textual environments of contemporary culture themselves – must come into play, with lesser or greater relevance. Social research in such settings, as Noortje Marres notes, ‘becomes noticeably a distributed accomplishment: online platforms, users, devices and informational practices actively contribute to the performance of digital social research’ (Marres 2012: 139). Given that the sociology of contemporary culture relies on the field of devices, the critical question is: how it will participate?

Digital sociology overlaps with, borrows from and seeks to make sense of much more prestigious, well-financed and heavily equipped practices in marketing, in scientific research, in government administration and commerce that target the expanded media environments of contemporary culture by intensifying work on data (Cukier and Mayer-Schoenberger 2013). *Scale* and *pattern* are key concerns in many of these settings, and scale and pattern also allure social researchers. As danah boyd and Kate Crawford suggest, ‘big Data tempts some researchers to believe that they can see everything at a 30,000-foot view’ (boyd and Crawford 2012).[1] But the mode of ordering – ‘recurring patterns embodied within, witnessed by, generated in and reproduced as part of the ordering of human and non-human relations’ (Law 1994: 83) – of digital sociology diverge greatly from contenders and analogues such as ‘data science’ (Schutt and O’Neil 2013) or ‘predictive analytics’ (Prediction Impact Inc. 2009) or digital humanities. Aiming to pay close attention to the ways in which the data-driven description of culture has become a key economic and social concern (Savage 2009; Thrift 2006), digital sociology might take shape less as a social science of digital culture than as the study of how device-specific events pattern lives, experience, power and value on various scales. This apparently limited ambition harbours some surprising expansive possibilities. It might cultivate ‘a sociological sensibility not confined to the predominant lines of sight, the focal points of public concern’ (Back and Puwar 2012: 12) but able to develop novel empirical techniques of inquiry and evaluate the unprecedented volume of information we encounter. Its device-specific emphasis might entail novel empirical practices and conceptions of the empirical (Adkins and Lury 2009).

As we will see, particularly through the examples we draw from one moderately large but critically relevant social networking platform, GitHub.com, the process of researching device-specific events on the various scales and amidst the intricate patterns of practice in contemporary media environments is not at all straightforward. Patterns and scales of action and interaction it turns out, entwine with devices in both contemporary culture and in its analysis. The case we draw on here – Github.com – is interesting precisely because what goes on there – coding and software development amongst other things– is both typical of the vast and somewhat incoherent work done on devices as part of contemporary culture, and at the same time plays an important practical role in re-formatting that work so that it becomes more publicly visible as a call to social order. Launched in late 2007, Github at the time of writing (2014) claims to host around 13 million different code repositories made by around 6 million users who are mostly software developers (see http://www.github.com/about). While these numbers, like all social media statistics, need to be disaggregated and analysed (for instance, Github ‘users’ include automated software processes that hourly commit new code to certain repositories; or many repositories are simply copies of other repositories, etc.), they position Github as a mid-ranking social media platform in terms of size. But regardless of size, Github certainly exemplifies the profuse work done on devices that texture increasingly large parts of people’s lives. A panoply of practical implementations of algorithms (see Couldry this volume), of graphic interfaces, of power-laden protocols, standards and infrastructures (such as operating systems, databases, security systems) can be found there.

Code repositories might be seen as vectors that on varying scales comprise much of the contemporary field of devices. These vectors display field-specific characteristics – code repositories contain code written in programming languages, they are accompanied by descriptions, they relate to a finite range of domains (science, business, publishing, gaming, graphic design, etc.) – and they also affect each other, as can be seen in the many imitations, variations, copies, re-implementations and versions found there. Moreover, Github increasingly serves as a model of social action that attracts many other practices not directly related to software: how-to guides, metadata on the Tate's art collection, the White House's open data policy, legal documents, recipes, books, and blogs are just some of the diversifying use-cases now found in repositories on Github. As a recent article in *The Atlantic* suggests, Github is increasingly of interest to non-programmers because the de-centralised, distributed and trackable collaborative processes it offers for coding can be used for many kinds of work: designs, legal documents, maps, images, books, blogs or websites (Meyer 2013). In this respect, Github, we might say, is an 'indexical icon' (Lee and LiPuma 2002) of the scaling and patterning that characterise the contemporary metamorphoses of culture.

Scales, scaling and fields of devices

One problem for digital sociology is the question of where to stand in looking at such complex processes.[2] In some respects, platforms like Github could be seen as what Andrew Barry termed 'technological zones of circulation.' These are

> spaces formed when technical devices, practices, artefacts and experimental materials are made more or less comparable or connectable. They therefore link together different sites of scientific and technical practice. Such zones take different forms. The points of access to the zones may be more or less clearly marked, with more or less well-defined and functioning gateways
>
> *(Barry 2001: 202–3).*

Github is a kind of gateway to the linkages between devices. It offers analytic and empirical descriptions and data streams as well as search functions that might help us engage with different scales and patterns of connection between technical configurations, platforms and practices. Wrangling these affordances in sociologically inventive ways entails much engagement with the practices and device specificities of the Github platform itself. As we craft ways of dealing with different scales, with wholes, with events and device-specific relations in Github, the allure of total or exhaustive description promised by the platform itself need to be reckoned with.

Given the heavily accumulative dynamics of the field of devices, there is a temptation for digital sociologists in their own empirical work to emulate the descriptive practices – the analytics – that accompany the scaling-up of digital media platforms. If digital sociology was simply a theoretically sophisticated version of media analytics or business intelligence, it would miss the transformations in culture attested by the availability of this data. The real analytical promise of large-scale data for digital sociology instead concerns the possibility of studying how re-scaling processes actually cut across pre-formatted notions of the individual or culture as a whole. The smooth conceptual poles of individual and whole culture constrain analytical purchase on open-ended, heterogeneous device-specific contestation of contemporary cultural spaces. They are attached to more static concepts of agency and structure that have long been disputed (we are not re-visiting those debates here; see (Latour *et al.* 2012) for a broader engagement with this problem, which arguably still thwarts 'social physics'). For instance, while 'users' or consumers remain the focus of much market research and advertising, their demographic attributes of race, age, ethnicity, income and educational level are perhaps becoming less important as 'postdemographic' concerns of profiling in terms of groups, tastes, interests, device usage, etc., take hold (see Rogers 2013: 153–4). Furthermore, the feedback loops between social lives running through the field of devices mean that cultural processes change scale in many different ways. Like all fields, the field of digital devices is tensioned by many different vectoral relations (see Martin and Merriman, this volume for a relevant description of Pierre Bourdieu's field theory). As Evelyn Ruppert, John Law and Mike Savage write, 'fields of devices [are] relational spaces where some devices survive and dominate in particular locations while others are eclipsed, at least for the moment' (Ruppert, Law and Savage 2013: 13).

We can glimpse some of these relational transients in Github as a platform. On the one hand, like many social media platforms, Github.com broadcasts data about what people are doing on the platform. At the time of writing (August 2014), around 220 million events are available on the so-called Github.com timeline since early 2012. The timeline is a comprehensive time-stamped series of user-generated events.[3] Many of these events are highly ephemeral. They have little afterlife. Someone creates a repository and puts something there, and then never returns. Millions of such events occur. On the other hand, the Github timeline data does not necessarily include events generated by Github itself as it develops the platform. Platform-level events – changes in architecture, modifications to interfaces, shifts in underlying design or management practice – are much harder to see from the data. Changes in the platform affect what people do. In the flow of events marked on the Github.com timeline, there are some surprising features. For instance, the 18 months of Github event data graphed in Figure 24.1 shows growth.

We might not be surprised to see a steadily increasing count of events on a social media platform. (Github is, after all, a social media platform for coding). While growth curves are common to many digital media platforms, here we see something anomalous. The number of events per week in April 2012 and especially September 2012 exceed the number of events per week at the end of 2013. Two peaks appear earlier than they should. These peaks reflect something about the way Github as a platform stored data rather than a dramatic change in what people were doing on Github. These kinds of features in the data suggest that even platforms in the field of devices themselves suffer from device-specific forces.

Even if we manage to filter these platform-generated effects or archival anomalies (statisticians and scientists often have to clean data before they analyse it), the plot in Figure 24.1

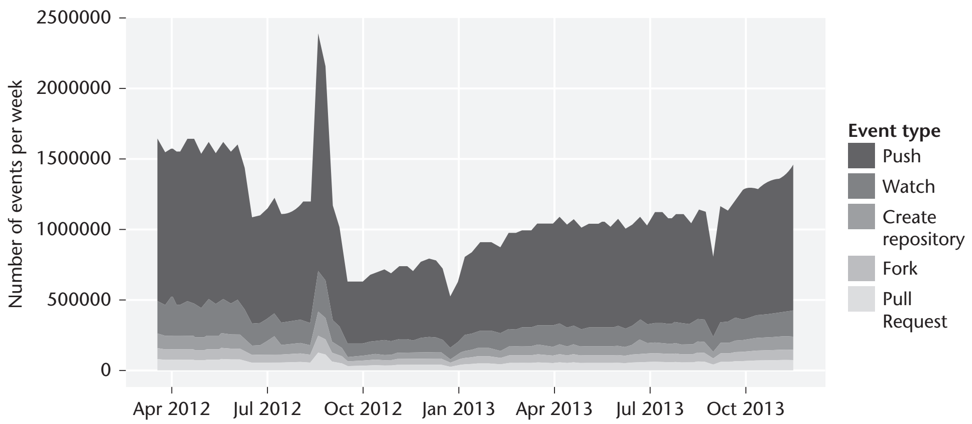

Figure 24.1 Events on the Github.com timeline

suggests other problems in analysing the field of devices. Events are not atomic social actions. 'Push' events, for instance, occur when software developers move code from a local repository to a Github repository. While there are some atomic events on the Github's timeline (for instance, a WatchEvent suggests that a person is interested in a particular code repository), other events such as PullRequest or PushEvents may wrap around complicated sequences and content 'payloads'. The payload complexity of events poses analytical problems. It steers analytical work into the depths of highly localised practices and highlights the importance of mapping local patches of action rather than general aggregates.

Yet the primary form of device-specific order in the vast textual fields of the Github data is the formatted events. Github as a device formats data on the field of devices in ways that encourage a focus on individuals and individual repositories. It affords little scope for examination of transverse flows or cross-cutting vectors in the field. The all-important APIs, the more or less real time data sources that GitHub exposes, effectively provide data about individual repositories and individual users or individual organisations. This formatting of data is typical of contemporary feedback loops: while device-specific data is readable by many, its formatting affords certain interested uses. Indeed, Github.com itself presents and encourages the production of various forms of visualisation and tabulation of what goes on in code repositories (and as mentioned above, the fact that we can access an archived version of the Github event timeline attests to this). GitHub ran a data competition in 2012 in which data analysts sought to do something with the timeline data (Doll, 2012). But what is most available from that data is the formatted events (as we saw in Figure 24.1) that more or less reinforce Github's conception of itself as first of all a *hub* and secondly as a *social* media platform. This means that we can easily, for instance, examine a most important or well-known repository on Github such as the Linux kernel, a much-vaunted, commercially, culturally and technically vital software device (https://www.github.com/torvalds/linux). Relatively quickly, individual developer contributions can be analysed, and we could begin to characterise the composition of the group of people who keep this important software object working and up-to-date. And indeed, this visualisation work is already supplied by Github.com itself (as the screenshot in Figure 24.2 shows).

It would even be possible through careful reconstruction to graph the changes in the network of relations that comprise the Linux kernel development teams, and perhaps to see how patterns of work on Linux kernel have changed over time as its economic and technical significance has extended through the popularity of the Android phones that rely on Linux and through the many corporate and industry users/producers invested in Linux. Issues, conversations, team and organisational changes could all be mapped. We could look further at the thousands of 'forks' and 'mirrors' (copies) of Linux to be found on Github, and characterise in fine-grained ways how processes of imitation generate flows of readers/writers, how meanings and practices are stabilised through repetition, and how invention might occur at intersections or overlaps between fields.

This leaves digital sociology in an interesting position. On the one hand, the Github data allows us to locate an important device – Linux – in the field of devices. The way certain devices attract work, imitations, and variations can be seen from the event data. And the incorporation of Linux into other devices – for instance, the Android platform – can also be gauged. These ongoing alignments and associations are important features of the field of devices. Yet other great patterns of practice on other scales do not immediately surface in the Github data. For instance, we know that particular code constructs are imitated or reinvented in thousands of different projects. We know too that same kinds of software device are re-implemented or imitated across different languages, or for different

Figure 24.2 Top contributors to Linux on Github

platforms, often with only small variations. These common patterns striate the field of devices, but barely appear in the Github-formatted data. Those cross-cutting patterns span devices and weave them together in surprising conjunctions: the same code constructs might be found in projects focusing on an ecommerce platform, ecological models of rainforest species diversity, Mozilla Foundation's Firefox browser, the Linux kernel, a library to manipulate financial data, or a modified version of the Android operating system for smart phones. In short, despite the domain-spanning dynamics of the field of devices, the way in which code migrates across social fields cannot be readily analysed. In exploring the field of devices, digital sociology finds highly diverse domains of social action and transaction, linked by the common practices, formats and materials (code patterns, repository mechanisms).[4]

Although it might seem a purely technical or practical issue, the problem of traversing such data flows lies at the heart of digital sociology as well as many parts of contemporary culture. One response to the problem of traversing data flows is to make devices that capture, reduce or visualise data.

The scaling up of the information infrastructures needed to deal with diverse data flows, for instance, is a salient concern of the software projects on Github.com. Many Github repositories explicitly address problems of scale and data. For instance, over 1500 repositories relate to 'MapReduce,' a parallel-processing algorithm first developed at Google to improve web search engine response times (https://github.com/search?q=map+reduce); 3700 repositories engage with Hadoop, a widely used Java-language implementation of the MapReduce algorithm (https://github.com/search?q=hadoop;). This scaling attests to auratic effects propagating in the field of devices as patterns of imitation. That is, the infrastructures that increasingly bring people into various forms of relationality are difficult to concretely manage. One can gaze on somewhat sublime images of earth-spanning information flows or visualisations of networks of hundreds of millions of Facebook users, but it is the work on devices such as Hadoop that shows us how people concretely grapple with the auratic promise of data flows.

A second response, common in the field of devices, is to make judgements about what is happening to data. Why does Google make available via its 'BigQuery' service a massive public archive of all the data produced by Github since early 2012? While social researchers might find it enormously useful to have an aggregated, hourly-updated timeline of all Github actions, it is very unlikely that some group at Google marketing or engineering concerned with data architectures for cloud computing has social researchers particularly in mind. Much more likely, they seek to attract the attention of the millions of software developers who use Github.com for coding work. Trying to see themselves and others in this high volume datastream, software developers and programmers familiarise themselves with Google's BigQuery architecture, and perhaps use it more thereafter. On the one hand, work on the Github timeline data demonstrates the power of Google Inc.'s cloud computing services. On the other hand, it promotes those services to software developers by inviting them to explore an important aspect of their own work – coding – as a data flow, and to produce second order judgements on it, including many largely aesthetic judgements of taste (see Martin and Merriman, this volume). But the developers' interest in doing this presupposes that they have an interest or investment in making judgements about code, or finding patterns in coding work. For social researchers too working on Github.com, the Google BigQuery datasets enable a widely differing scale of exploration of practice in the field of devices, and constantly increase the risk of being drawn away from the singularity and variability of practices to produce large scale tabulations of results.

Against the flow: Anti-patterns in digital sociology

A final response to data flows – and this is one that we have pursued – asks: what work needs to be done against the data flows and against its inevitably device-specific formatting of data in order to apprehend forces tensioning the field of devices? One possibility here is to examine how people act in the field of devices to make sense of patterns in data. An interest in patterns lies at the heart of digital data flows. The problem of finding patterns in data is a chronic concern in science, business and government data practices. And certainly patterns and the seeing of patterns are the central pre-occupation in many contemporary sciences, in financial markets, in biomedicine and in business analytics (see for instance, the fields of 'pattern recognition' and 'machine learning', Hastie, Tibshirani and Friedman 2009). But a prior and crucial question here is how to think about the value of pattern, or pattern finding as aesthetics of the social. As we have seen, in sociological thought more generally, pattern is a long-standing concern. John Law speaks of 'recurring patterns of the social' Nick Couldry of 'patterns of flow', and Andrew Abbott suggests that 'if most things that could happen don't happen, then we are far better off trying first to find local patterns in data and only then looking for regularities among those patterns' (Abbott, 2001: 241). As Mike Savage suggests, we need to understand 'how pattern is derived and produced in social inscription devices' (Savage 2009: 171), whether these devices are objects of analysis or part of our own methods. A broader philosophical re-conceptualisation of patterns runs through some social and cultural theory (for instance, in recent work influenced by A.N. Whitehead, who writes 'beyond all questions of quantity, there lie questions of pattern' (Whitehead 1958: 195).

However it is conceptualised, the fact remains that concrete work on patterns largely takes a quite limited number of forms. This limitation in forms suggests that the device-specific formatting of data is hard to resist, and that the social aesthetics generated by the field of devices is a symptom of this. People craft many data visualisations. But the principal visible forms of pattern are rather limited. They include plots that show lines, curves, peaks and clusters of points, networks, trees, and maps. Drawing on the spectrum of plots, graphs and diagrams developed in the last few centuries (see http://www.datavis.ca/milestones/ for a catalogue; Edward Tufte's work (Tufte 2001) is a standard reference for quantitative digital data), contemporary visual displays of pattern abound in 'predictive analytics and the move back to visualization in social statistics, the new cartography and associated Web 2.0 innovations, [and] visual montages designed to represent amalgams of "variables"' (Burrows 2012: 585). These patterns can take the form of plots (bar graphs, scatter plots, lines and curves drawn through clouds of data points, network diagrams), maps on many different scales (especially maps that superimpose different geo-located datasets), information visualisations (typically combining data graphics, text and typographic design elements), as well as tables and textual-graphic forms (word clouds), with varying degrees of animation and with many different forms of scale. Practices of data visualisation are routinised through the proliferation of certain visual forms (the network diagram, heat maps, bubble plots and chloropleth maps are widely found) in different places. They have field-specific attributes. Although network visualisations, tag clouds, stream graphs and the like have abounded in data visualisation on the web, especially with the growth of graphics libraries and packages such as Hadley Wickham's *ggplot2* (Wickham 2009), Mike Bostock's *d3.js*, or widely-used scientific plotting packages such as *matplotlib* (software projects all currently hosted on Github), we have little sense yet of the visual culture of these devices and their visual forms. While devices of various kinds may have been involved in producing them (for instance, many smoothing algorithms used to draw lines through points effectively fit a series of local linear models – splines – in building a smooth curve), the models themselves are not interpreted as such but act in the world more like things than thoughts. If pattern matching and pattern recognition are becoming mundane parts of contemporary culture in many different ways, it is partly because these forms of judgement or perception are endemic in the field of devices. (For instance, in the face recognition logic now built into many digital cameras, or the much-discussed recommendation systems typical of online commerce). They are in any case widely distributed through various social fields where visual devices associated with displays, gauges, metrics, dashboards, graphs, and visualisations form part and parcel of social life, whether in the graphic displays that users of supply chain and inventory management systems or financial traders gaze at (Knorr Cetina and Bruegger 2004), or in the many news-related visualisations produced by data journalists for news sites such as *The Guardian* datablog (n.d.) or the *New York Times* (2012).

Some commentators suggest that the production and use of such visualizations is a key concern for contemporary sociology: 'the discipline . . . will have to take visualization methodologies far more seriously than we have hitherto' writes Roger Burrows (Burrows 2012: 585) because of the ways that they are being used and could be used to understand 'particular patterns of association that exist between persons, objects, symbols, technologies and so on' (585). The renewed emphasis on visualization in digital sociology differs somewhat from adjacent efforts such as computational social science (Giles 2012; Housley *et al.* 2013) where visualisation is usually close evaluation of statistical or predictive models. Sociological work on the transformations of data visualisation is still rather scarce. Scientific visualisation offers some leads here (Latour *et al.* 1990; Myers 2008) alongside work on visualisation in finance (Pryke 2010), but the visual culture of data as it moves out of scientific publication has received little attention. There is much scope for investigation of the seeing in data visualisation as forms of visual culture in which, as Gillian Rose writes, 'different ways of seeing are bound up into different, more-or-less conscious, more-or-less elaborate, more-or-less consistent practices' (Rose 2012: 549). This is a challenge to methodological practices precisely because the visual forms attest to a shift away from some traditional sociological concerns with abstractions, models and structures as deep explanations of social processes, and a lighter, perhaps more responsive descriptive attunement to patterns, groupings and flows.[5] What do they make visible? Both in the visual culture of data, and in its own visualisation of digital data, digital sociology faces the problem of describing how patterns are produced at the intersection of various concretisations and abstractions as reactions to certain aspects of experience. Digital sociology, we suggest, might take data graphics and the many judgements and discussions of data graphics seriously as a form of judgement endemic to the field of devices.

Patterns arise in very different ways. Many data visualisations seek to render perceptible something that occurs on spatio-temporal scales that are difficult to directly see, but they often struggle to distinguish something amidst the generic schematic formatting of the data. For instance, Figure 24.3 seeks to convey something of the patterns associated with different scales of activity in Github repositories by counting events that appear in the Github timeline over an interval of two years (2012–2013). The general pattern shown here is the somewhat ubiquitous 'power law' distribution of events, a distribution that often shows up in social media data. At the left hand end, the high point refers to the millions of Github repositories consisting of one or two events. At the low end on the right, where the curve approaches the x-axis, a small number of repositories receive many thousands of events. The power law distribution of events in social media often vexes data analysis and data visualisation. Many social media datasets yield heavy-tailed distributions when graphed. This common scaling of

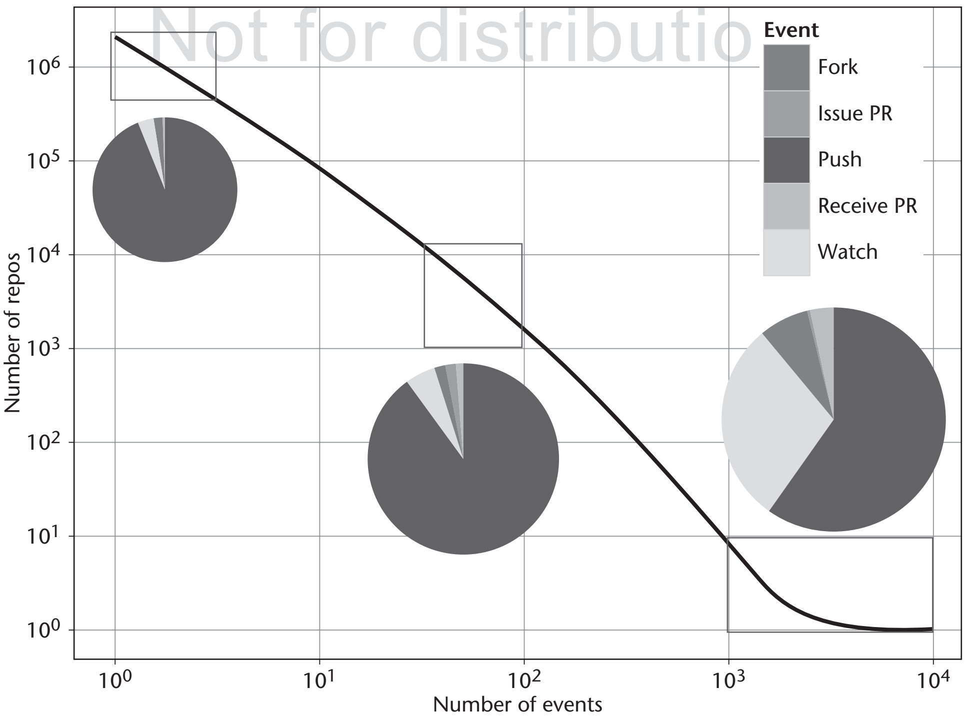

Figure 24.3 Patterns of repository events on Github

events across the 'many-some-rare' scales (Conte *et al.* 2012: 334) requires site-specific work. The visualisation of repository event counts begins to do this by showing something of the different composition of the repositories on the different scales. The many small repositories mainly consist of the a few Push events. Mid-scale repositories show the presence of more social events such as Watch and Fork. The rarer very large repositories attract many more social events – Fork, Watch and PullRequest. But if this patterning across many-some-rare is so common as a reaction to something in the contemporary field of devices ('In recent years, due to ubiquitous computerization, networking and obsessive data collection, reports of heavy-tailed distributions have almost become a routine' (Muchnik *et al.* 2013: 1), what does it say about the global organization of the field of devices?

On the one hand, many data visualisations, whether in the form of networks, scatter plots or line graphs today present the power-law or scale-free pattern of digital media. If is often said that patterns are supplanting causes as modes of explanation in many places, and the growth of data visualisations might be understood in these terms. These descriptive visualisations might prompt some causal interpretation in their viewers but they are not premised on any such abstraction. Might they be seen perhaps more generally as an integral part of the cultural saturated feedback loop running through the field of devices? As technological concretisation binds practices, habits, emotions, and interactions through digital devices and infrastructures and devices, the derivation of patterns increasingly depends on abstractive devices that classify, cluster, calculate and predict events precisely in order to shape them. Predictive analytics, as demonstrated in Google Research's work on how users' searches foreshadow airline ticket bookings or car sales (Varian and Choi, 2009), derives patterns from data using a much more technical armature of machine learning techniques. This modelling is an increasingly dense force affecting the feedback loop between people and digital infrastructures (Pariser 2011). Algorithmic classificatory techniques such as k-means clustering, nearest neighbour classification, linear regression, logistic regression, principal component analysis, neural networks, decision trees, random forests, and support vector machines are rapidly becoming an integral part of every level and niche of digital assemblages, ranging from playful mundane devices such as 'kittydar', a neural network to detect cat photos (https://github.com/harthur/kittydar) through to thousands of projects implementing 'face detectors' or 'motion detectors' for smart phones, web browsers, and for different operating systems. Even a single technique like the popular random forest classifier (Breiman 2001) can be found hundreds of thousands of times in Gitbhub.com repositories, and tens of thousands of times in popular programming languages such as R (R Development Core Team n.d.). Again, the proliferation of these classificatory or pattern finding techniques is perhaps much less visible, and the ways in which they imprint or weave through flows of meaning and things is harder to analyse. They are somewhat withdrawn elements in the feedback loops of cultural space. These predictive models and classifiers sometimes operationally shape the experience and action (as in Netflix or Amazon recommendation systems, or in the classifiers that detect and classify body gestures in the Microsoft Kinect game controller), and sometimes they are analytic tools used by people working on platforms trying to make sense of emerging or divergent patterns in practice. As always, the feedback loops between knowing and acting are hard to disentangle precisely because they are becoming more tightly coupled. If decision trees were an analytic technique developed by statisticians in the late 1970s trying to make sense of air pollution measurements in Los Angeles (Breiman *et al.* 1984), in the Kinect game controller they become predictive devices that intensify the immediacy of computer game play. Online learning – the constant updating of predictive models in response to the flow of current events – is increasingly common in social network media and online transactions.

In the field of devices, patterns are generated, acted upon, altered and re-imprinted. Pattern recognition is no longer a practice conducted at leisure by expert interpreters or elite analysts. Patterns are operational components of device-specific zones of culture. The transformation of abstract analytic devices of many different kinds (linear regression models, clustering algorithms, Bayesian models, etc.) into things that either circulate much more widely in the world in gadgets and devices or into meta-things such as search engines that modulate flows on a large scale is a central component of the flow of texts, meanings, audiences, viewers, visitors, spectators, readers and players in many settings. These abstractions exercise an auratic effect akin to the data infrastructures we were discussing above. Journalists, social researchers and commentators on digital technology tend to attribute great potency to algorithms in general. The proliferation of algorithms is hard to deny, but digital sociology might play an important role in describing what happens as these algorithms shift shape and move into different settings ranging from pay day loans to computer game play, from web search engines or ecommerce recommendations to face recognition in digital cameras. The aura of algorithms as epistemic prime movers akin to steam power or electricity covers over their diverse provenance (they do not come just from computer science but also from statistics, psychology, cognitive science, ecology, archaeology or geology) and their diversity in practice. Perhaps more significantly the patterns that these algorithms produce or derive from data are neither obviously legible in descriptive devices, nor is their relation to existing structures, groupings, or classifications direct.

For digital sociology, recognising the effects of this pattern-making is a significant challenge. The imprinting of flows of meaning, media and practices is sometimes legible (for instance, in the 'induced viralities' seen in various social media platforms that identify trends using pattern recognition techniques and then shape flow of messages or network connections accordingly), but not always. Another difficulty is much more challenging. While descriptive devices can relatively easily slip into the analytical toolkit of social researchers – the spread of tag-clouds or Wordle graphics would be a typical instance of this – the appropriation and re-purposing of machine learning, pattern recognition or data mining approaches is more problematic. Many of these devices rely on formidable mathematical apparatus, ranging across linear algebra, probability theory, function analysis and numerical optimisation. The diverse provenance of the techniques means that, although they operate abstractly (that is, with little regard for the concrete specificities of a given situation), they handle notions of group, classification, difference and similarity heterogeneously. Nearly all of them bring to bear powerful scaling processes that reduce the high dimensions and volume of data to legible forms of variation and pattern.

Conclusion

While Bourdieu's comments on devices are not abundant, he wrote about traffic lights:

> The social world is full of calls to order which function as such only for individuals who are predisposed to notice them, and which, as a red light causes braking, trigger deep-rooted bodily dispositions without passing through consciousness and calculation
>
> *(Bourdieu 2000: 176).*

Contemporary culture is deeply textured by device-specific calls to order. Reactions to these signals, for those who notice them, are often deep-rooted and bodily enacted. The repositories of Github.com, which we have only lightly explored here, illustrate something of the variety of signals flashing in contemporary social worlds. It points to forms of reactions that go well beyond braking or accelerating, yet remain for all that somewhat non-conscious or largely affective. The experience of spiralling in and out of device-specific feedback loops generates many forms of reaction and reactivity.

In the field of devices, relations between devices connect, disconnect, attract, configure, imitate, intensify and re-distribute signals for ordered social action. The case of Github shows too that attempts to configure social patterns of action have animated the growth of large-scale digital infrastructures and social media. The algorithmic elements of search engines such as Google Web Search's *PageRank* algorithm or Facebook's *Social Graph* are two better known instances of the ways in which the detection of social patterns and flows of meanings, texts and readers has been pivotal in the growth of digital culture. Devices devoted to pattern recognition or data visualisations themselves flash signals amidst a field of devices. In analysing the vectoral components of the field of devices or tracing the reactions to that field, digital sociology is not doing anything radically different to people who inhabit this field. Savage suggests that:

> a core concern [for social research] might be to scrutinise how pattern is derived and produced in social inscription devices, as a means of considering the robustness of such derivations, what may be left out or made invisible from them, and so forth. We need to develop an account which seeks to criticize notions of the descriptive insofar as this involves the simple generation of categories and groups, and instead focus on the fluid and intensive generation of potential
>
> *(Savage 2009: 171).*

Patterns in cultural life today derive from social inscription devices assembled in wide-ranging feedback loops. Feedback runs between recording what people do, visualising or graphing what they do, finding/generating patterns in the recording, and shaping what they encounter next. If digital sociology attempts to describe what is happening in contemporary textual and media environments, it needs to map the paths of these feedback loops running across publics, infrastructures, techniques, textual and media forms, and diverse, expanding practices. At the same time, patterns themselves are increasingly heavily analysed and modelled inside the culturally saturated feedback loop between people and social inscription devices. Digital social research is not alone in its interest in these processes. Reputational, attentional and sentiment economies (Arvidsson 2011) directly act on that patterning.

Identifying events that animate this patterning and scaling is a key concern for digital sociology. The 'massive, culturally-saturated feedback loop' arises from device -saturated social action. As the case of Github shows, the process of making, configuring, arranging and aligning devices is itself a highly dynamic field of linkages, associations, and imitations, where social action is often concerned with problems of pattern and of scale. The 'simple generation of categories and groups' that Savage refers to can certainly be found there, replicating and propagating at scale.

We have suggested that social researchers in the field of devices react, like other participants, to calls to order in their own practice. Digital sociology and certainly social research more generally is not immune from the auratic imperatives to methodologically emulate and align themselves with the infrastructures and practices of the field of devices (such as 'big data'). It grapples with the reflexive-recursive problem of its own implication in methods, techniques and infrastructures for deriving pattern. Digital social researchers find ourselves lost in the labyrinth of technical possibilities opening up around platforms, tools, visual forms and data flows. The experience of being somewhat caught up in the entanglements of pattern and scale might for digital sociology be a necessary step towards sensing the fluid generative potentials in the field of devices. These entanglements between scale and pattern, and especially between the different ways in which pattern might be found could offer a way for sociology to deviate from the scaled-up homogeneity and uniformity of predictive analytics with their highly constrained commitment to increasing advertising revenue or sales.

Notes

1 We largely leave aside any further discussion, analysis or critique of these developments. They are extensive and multi-faceted, and can be seen at work in the digital humanities (Galloway 2014), in cultural analytics (Manovich 2011), in social physics and computational social science (Pentland 2014). The constant updating of events, the relatively frequent advent of new flows of data, the teeming and burgeoning ways of inhabiting the reefs of digital infrastructures, many of which are developed and publically available, confront digital sociology with challenges and alluring possibilities. Unlike the major social media platforms or even the legion of startup companies in London, New York, San Francisco, Shanghai, Berlin or Amsterdam who offer ways of packaging, summarising, monitoring or shaping flows of data in networks, digital sociologists do not have hundreds of developers to wrangle data, dozens of computers and disk-drives arrayed in racks to expedite the process of searching or exploring the data. Given the billions of events welling-up in diverse APIs, what can digital sociology do? Should it sample and filter according to the criteria of adequacy and representativeness? The expansive forms of textual environment we have just been describing are writ large at the moment under the broad banner of 'big data'. The proliferating discussions of 'big data' need to be analysed in their own right in terms of how they intensify desires to connect information flows, previously disparate infrastructures and systems (energy, telecommunications, entertainment, transport, retail and manufacturing), and how they actually reorganise work, domestic lives, forms of sociality and value in the name of flows of data. 'Big data' is certainly part of the feedback loop or accumulation strategy in which social practices recorded as data become the basis of new textualities that seek to enrol further readers or writers, to align reading, viewing and buying, writing and working, as well as other forms of value. The physical, life and environmental sciences offer both a lead and something distracting here. The term 'data' carries with it an aura of scientificity, objectivity or neutrality that digital sociology is still wrestling with. 'Big data' is an expansive grouping and its membership continues to grow: house prices, clicks on hyperlinks, vehicle detection loops on roads, mobile phone call details, satellite photos of crops, electronic payment, stellar images from orbiting telescopes, transactional data such as credit card authorisation or supermarket checkout scanners – these are just examples in a list that keeps growing. The listing of data sources is an interminable feature of most talk about big data (and digital sociology shares this habit). Scientific data, however, has a particular resonance and perhaps anchors some forms of referentiality in data talk. The standard reference to allude to transformations in scientific data is *The Fourth Paradigm: Data-Intensive Scientific Discovery* (Hey, Tansley and Tolle 2009), a book published by Microsoft Research Press (a publisher that probably lies quite close to the source of much business data practice). This book furnishes vignettes of a range of scientific enterprises ranging across physical, earth, environmental and life sciences in which flows of data have transformed knowledge-making practices. The data intensive sciences authorise data practices in specific ways. The auratic power of scientific instruments such as DNA sequences or infrared satellite photos differ, as Mike Savage observes, from the 'mundane descriptions,... ordinary transactions, websites, Tesco loyalty cards, CCTV cameras in your local shopping centre, etc., that are the stuff of the new social' (Savage, 2009: 171). It may be that this auratic/mundane difference, important though it is in differentiating certain practices, also usefully links different domains of the social. Auratic scientific instrument data, with its referential links to the diversity of life, the fate of the planet or the conundrums of missing matter in the universe, rivets data to things. The scientific examples allow data more generally in all its administrative, transactional or media-derived forms to carry universalising epistemic value. It suggests that the birth of stars in remote galaxies can be analysed in similar terms to the birth of stars in the media environments of Xfactor or reality TV shows. In this respect, the popularity of the term 'data *science*' suggests that the referential power of science matters to business, commerce, industry and government as they seek to commodify, extract or regulate contemporary cultural spaces.

2 The anthropologist Anna Tsing writes: 'scale is the spatial dimensionality necessary for a particular kind of view, whether up close or from a distance, microscopic or planetary. I argue that scale is not just a neutral frame for viewing the world; scale must be brought into being: proposed, practiced and evaded, as well as taken for granted' (Tsing, 2005: 58). *Scale* refers to the relative dimensions of enlargement or reduction in a map, picture or model, as well as the marks or degrees used to measure intervals (as in the scale on the axes of a scatter plot), but it implicitly positions observers as well. *Scale,* however, is a verb as well as a noun. In digital culture, re-scaling or re-dimensioning is common. The *scaling-up* of databases, of transactions, of geographies (e.g. Amazon's data-centres divide the globe into eight regions), and capacities in many settings testifies to one aspect of this re-scaling.

3 While the textual environment for online code repositories such as Github includes the many forms of text and graphic visible on Github webpages, it also includes the flows of data that these platforms generate for use by others. As millions of people interact even with mid-level platforms such as Github, their actions generate large volumes of data that can be streamed as time-stamped events. So, like many other social media platforms, Github publishes all of these individual time-stamped events to anyone who wants to use them. (While some users may pay to have their repositories and working practices remain invisible, very many do not.) The events are available more or less as they happen (through the Github API – application programmer interface) or in bulk through various archives (https://www.githubarchive.org/; a mirror of the data is also published by Google as a demonstration of their 'BigQuery' cloud computing service). The data derived from what people do on social media is mixed in form. It includes when things were done (date-time), various free textual forms (descriptions, tags, titles, etc.), structured text (links or URLs of associated webpages; names of associated organisations and groups; names of user), categorical attributes (on Github these are limited: the event type and the programming languages used), and often a mathematically encoded summary of the result of actions that change the contents of the repository (in the case of Github, the *hash digest* of any new content or change to existing contents of the repository). Github events are categorised according to 18 different event types (PushEvent, CommitEvent, AddUserEvent, DeleteEvent). These event types are organic to Github.com, but as variables in any data analysis they are defined by the platform designers rather than by any questions that social researchers might bring to bear on what people do on Github. 'Source-defined variables' are a central concomitant of data analysis practice in digital sociology. We might say that in the new social fields, data has an 'organic' aura: it is generated and collected by virtue of the existence of the infrastructures and platforms that are part and parcel of the social field rather than from instruments or measuring devices introduced by market or social researchers. But even this relationship is becoming increasingly complicated by virtue of the intricate and shifting relations between 'organic data' and 'paid for data' (Google Inc. 2009). Even if this data is 'observational' rather than experimental, the fact that it is generated intrinsically as part of a social field has powerful referential attractions, and attracts much work.

4 In much social research, problems of scale largely related to scarcity. Where data was abundant, scale problems were handled by setting limits on data through sampling strategies, research methods and research designs that allowed social researchers to be more or less confident that their research covers the social field of interest (for instance, making sure that differences in age, gender, ethnicity, sexuality, nationality, education-level, or income are represented in the data; many of the chapters in a typical introductory statistics textbook for social science address these problems of scarcity and representativeness). Selecting and sampling strategies seem to work differently in digital sociology as it encounters expanding textual environments where data scarcity is rarely an issue. The much discussed problem is how to cope with the vast amount of material. What is worse, textual environments such as social media are explicitly expansive.

5 This contrast has been extensively debated in sociology. (See Andrew Abbot's discussion of patterns versus causes; the 'empirical crisis in sociology' literature; as well as the explicit focus on digital devices in recent sociology (Abbott 2001; Ruppert 2013; Savage and Burrows 2007) and we will not rehearse these debates in great detail here. They have been debated fairly widely elsewhere (see Burrows 2012 for an overview).

References

Abbott, A. (2001) *Time Matters: On Theory and Method*, Chicago: University of Chicago Press.

Adkins, L. and C. Lury (2009) 'Introduction: What Is the Empirical?' *European Journal of Social Theory*, 12(1): 5–20.

Arvidsson, A. (2011) 'General Sentiment: How Value and Affect Converge in the Information Economy', *The Sociological Review*, 59: 39–59.

Back, L. and N. Puwar (2012) *Live Methods*, Oxford: Wiley-Blackwell.

Barry, A. (2001) *Political Machines: Governing a Technological Society*, London: Athlone.

Bourdieu, P. (2000) *Pascalian Meditations*, Stanford: Stanford University Press.

Boyd, d. and K. Crawford (2012) 'Critical Questions for Big Data', *Information, Communication and Society*, 15(5): 662–79.

Breiman, L. (2001) 'Random Forests', *Machine Learning*, 45(1): 5–32.

Breiman, L., J. Friedman, R. Olshen and C.J. Stone (1984) *CART: Classification and Regression Trees*, Belmont, CA: Wadsworth, 156.

Burrows, R. (2012) 'Digitalization, Visualization and the "Descriptive Turn"', in Heywood, I., Sandywell, B. (eds), *The Handbook of Visual Culture*, London: Berg, 572–88.

Conte R., N. Gilbert, G. Bonelli, C. Cioffi-Revilla, G. Deffuant, J. Kertesz, V. Loreto, S. Moat, J.P. Nadal, A. Sanchez, A. Nowak, A. Flache, M. San Miguel, and D. Helbing (2012) 'Manifesto of Computational Social Science', *European Physical Journal-Special Topics*, 214(1): 325–46.

Couldry, N. (2000) *Inside Culture: Reimagining the Method of Cultural Studies*, London: SAGE.

Cukier, K.N. and V. Mayer-Schoenberger (2013) 'The Rise of Big Data. How It's Changing the Way We Think About the World', *Foreign Affairs*, Available online at http://www.foreignaffairs.com/articles/139104/kenneth-neil-cukier-and-viktor-mayer-schoenberger/the-rise-of-big-data (accessed 10 February 2014).

Doll, B. (2012) *Data at Github*, GitHub, Available online at https://github.com/blog/1112-data-at-github (accessed 3 March 2014).

Galloway, A.R. (2014) 'The Cybernetic Hypothesis', *differences*, 25(1): 107–131.

Giles, J. (2012) 'Computational Social Science: Making the Links', *Nature*, 488(7412): 448–50.

Hastie, T., R. Tibshirani and J.H. Friedman (2009) *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, New York: Springer.

Hey, T., S. Tansley, and K. Tolle (2009) 'The Fourth Paradigm: Data-Intensive Scientific Discovery', *Microsoft Research*.

Housley, W., M. Williams, M. Williams, and A. Edwards (2013) 'Computational Social Science: Research Strategies, Design and Methods Introduction' *International Journal of Social Research Methodology*, 16(3): 173–5.

Knorr Cetina, K.K. and U. Bruegger (2004) 'Traders' Engagement with Markets: A Postsocial Relationship', in Ash A. and Thrift N., *The Blackwell Cultural Economy Reader*, Oxford: Blackwell.

Latour, B., M. Lynch, and S. Woolgar (1990) 'Drawing Things Together', in M. Lynch and S. Woolgar (eds), *Representation in Scientific Practice*, Cambridge, MA: MIT Press.

Latour, B., P. Jensen, T. Venturini, S. Grauwin, and D. Boullier (2012) 'The Whole is Always Smaller than its Parts. How Digital Navigation May Modify Social Theory', *British Journal of Sociology*, 63(4): 590–615.

Law, J. (1994) *Organizing Modernity*, Oxford: Blackwell.

Lee, B. and E. LiPuma (2002) 'Cultures of Circulation: The Imaginations of Modernity', *Public Culture*, 14(1): 191–213.

Lupton, D. (2012) *Digital Sociology: An Introduction*, Sydney: University of Sydney, Available online at http://prijipati.library.usyd.edu.au/handle/2123/8621 (accessed 14 November 2013).

Manovich, L. (2011) 'Trending: The Promises and the Challenges of Big Social Data', *Debates in the Digital Humanities*, Available online at http://manovich.net/index.php/projects/trending-the-promises-and-the-challenges-of-big-social-data (accessed 15 April 2013).

Marres, N. (2012) 'The Redistribution of Methods: On Intervention in Digital Social Research, Broadly Conceived', *The Sociological Review*, 60(1): 139–65.

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

Muchnik, L., S. Pei, L.C. Parra, S.D.S. Reis, J.S. Andrade Jr., S. Havlin, and H.A. Makse (2013) 'Origins of Power-Law Degree Distribution in the Heterogeneity of Human Activity in Social Networks', *Scientific Reports*, 3, Available online at http://www.nature.com.ezproxy.lancs.ac.uk/srep/2013/130507/srep01783/full/srep01783.html (accessed 1 March 2014).

Myers, N. (2008) 'Molecular Embodiments and the Body-Work of Modeling in Protein Crystallography', *Social Studies of Science*, 38(2): 163–99.

Pariser, E. (2011) *The Filter Bubble: What the Internet is Hiding from You*, London: Penguin.

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

Pryke, M. (2010) 'Money's Eyes: The Visual Preparation of Financial Markets', *Economy and Society*, 39(4): 427–59.

Rogers, R. (2013) *Digital Methods*, Cambridge, Massachusetts, London: The MIT Press.

Rose, G. (2012) 'The Question of Method: Practice, Reflexivity and Critique in Visual Culture Studies', in I. Heywood and B. Sandywell (eds) *The Handbook of Visual Culture*, London: Berg.

Ruppert, E. (2013) 'Rethinking Empirical Social Sciences', *Dialogues in Human Geography*, 3(3): 268–73.

Ruppert, E., J. Law and M. Savage (2013) 'Reassembling Social Science Methods: The Challenge of Digital Devices' *Theory, Culture and Society*, 30(4): 22–46.

Savage, M. (2009) 'Contemporary Sociology and the Challenge of Descriptive Assemblage', *European Journal of Social Theory*, 12(1): 155–74.

Savage, M. and R. Burrows (2007) 'The Coming Crisis of Empirical Sociology', *Sociology*, 41(5): 885–99.

Schutt, R. and O'Neil, C. (2013) *Doing Data Science*, Sebastopol, CA: O'Reilly & Associates Inc.

Thrift, N. (2006) 'Re-Inventing Invention: New Tendencies in Capitalist Commodification', *Economy and Society*, 35(2): 279–306.

Tsing, A.L. (2005) *Friction?: An Ethnography of Global Connection*, Princeton, NJ: Princeton University Press.

Tufte, E. (2001) *The Visual Display of Quantitative Informations*, second ed., Cheshire, CT.: Graphics Press.

Varian, H. and H. Choi (2009) 'Official Google Research Blog: Predicting the Present with Google Trends', Available online at http://googleresearch.blogspot.com/2009/04/predicting-present-with-google-trends.html (accessed 23 May 2011).

Whitehead, A.N. (1958) *Modes of Thought; Six Lectures Delivered in Wellesley College, Massachusetts, and Two Lectures in the University of Chicago*, New York: Capricorn Books.

Wickham, H. (2009) *ggplot2: Elegant Graphics for Data Analysis*, New York: Springer. Available online at http://ggplot2.org/.

Online Sources

Google Inc. (2009) Back to Basics: Direct, Referral or Organic – Definitions Straight from the Source – Analytics Blog. *Google Analytics Blog*, Available online at http://analytics.blogspot.co.uk/2009/08/back-to-basics-direct-referral-or.html (accessed 7 February 2014).

The Guardian (n.d.) Data Journalism and Data Visualization from the Datablog. Available online at http://www.theguardian.com/news/datablog (accessed 30 January 2014).

New York Times (2012) *2012: The Year in Graphics*. Available online at http://www.nytimes.com/interactive/2012/12/30/multimedia/2012-the-year-in-graphics.html (accessed 30 January 2014).

Prediction Impact Inc. (2009) Predictive Analytics World Conference: Agenda. Available online at http://www.predictiveanalyticsworld.com/sanfrancisco/2009/agenda.php#usergroup (accessed 24 May 2011).

R Development Core Team (n.d.) The R Project for Statistical Computing. Available online at http://www.r-project.org/ (accessed 11 June 2010).