

Speculative Diagrams: experiments in mapping Youtube.

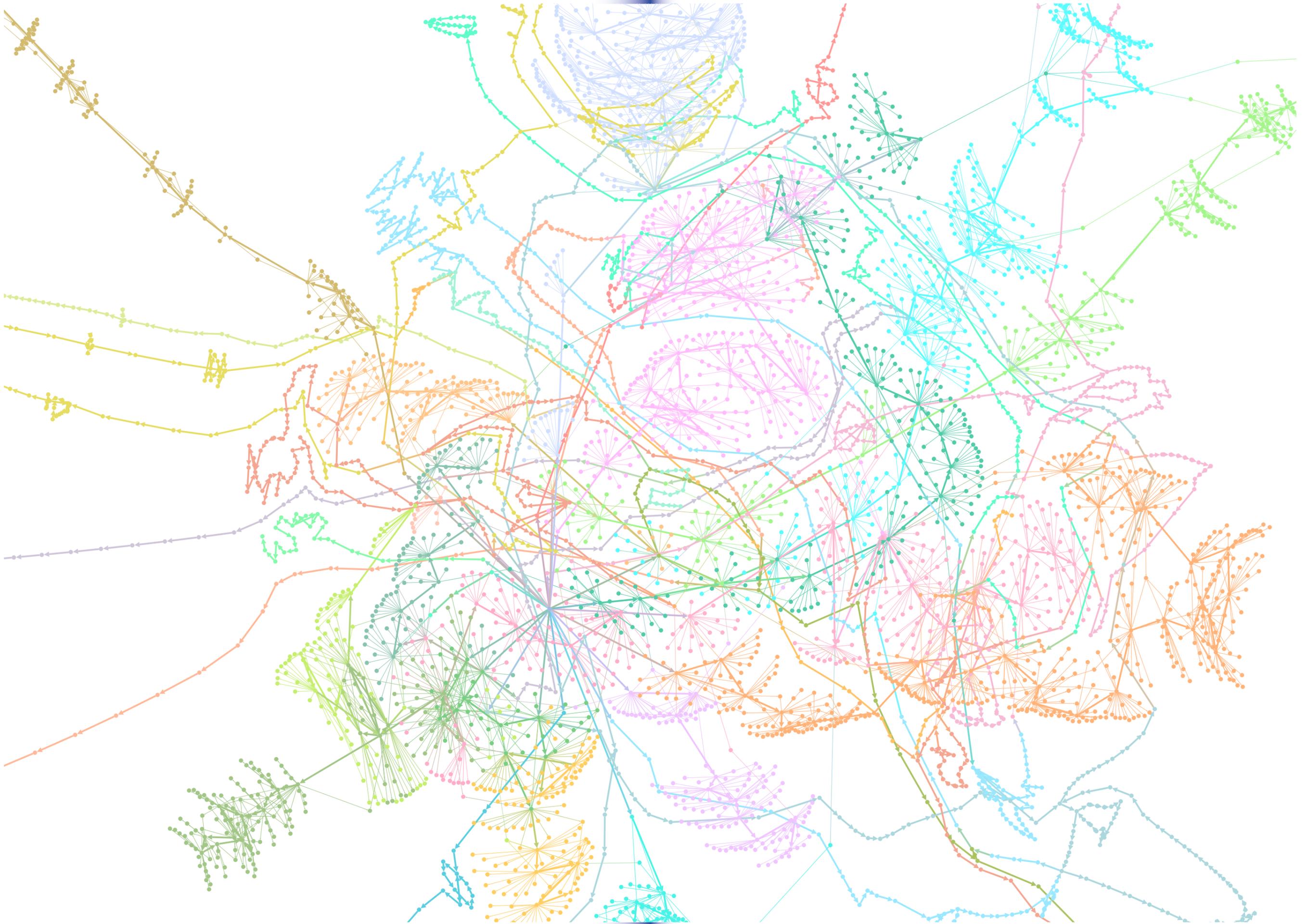
Betti Marenko¹, David Benqué²

¹ Central Saint Martins Uni-
versity of the Arts Lon-
don, United Kingdom
b.marenko@csm.arts.ac.uk

² School of Communica-
tion, Royal College of Art,
London, United Kingdom

Abstract: The notion that the future has a ‘shape’ is a deep-rooted construct, the cornerstone of how chance is mediated, for example through the ‘distributions’ of probability theory. Algorithmic prediction, via machine learning, builds on these shapes and amplifies their complexity and authority. While the problematic effects of this predictive regime and the preemptive politics it supports are objects of concern for scholars and practitioners across the humanities, social sciences, art, and philosophy, design is surprisingly disengaged from this conversation. Instead, it is either concerned with data visualization, often without questioning its positivist ontology, or with ‘seamless’ non-interfaces which effectively seek to remove choice. Against this backdrop, our proposal brings together design theory and design practice to interrogate current modes of algorithmic prediction and the construction of subjectivity enabled by ‘choice design’. Our designed artifacts are *diagrams* to think through practice about the shape(s) of the possible. Rather than designing predictable futures, we aim to use diagram-making to expose and reframe choice design. These design artifacts - initial and ongoing experiments in mapping YouTube recommendations - are a series of computational diagrams that weave together the tools of computational prediction, critical design practice, and theory.

Keywords: algorithmic prediction;
diagrams; recommendation; critical
data visualization;



Introduction

Our collaboration between design theory and practice (Marenko and Benque, 2018) investigates/ complicates/ criticises the practice and politics of algorithmic prediction: future-building via machine learning. It sets out to position diagramming as a computational counter-practice, to reclaim the possible from algorithmic capture. We use *diagrams* - speculative visualization mapping the unformed and the unstable as they feed into changeable situations (O'Sullivan, 2016) - to think through, manipulate, and theorise prediction from a design standpoint. We use diagram-making as a practice that affords ways of probing the possible as it emerges through modes of prediction and speculating on a future taken as fluid, rather a solid material. In this paper we move from this starting point into our first practical case study: a mapping of recommendations on the Youtube platform.

Any interface to a catalogue of media is, explicitly or not, an ordering of choices. While this is inescapable—there is no way to present a list without some kind of ordering—it is also problematic. While user-friendly interfaces appear to be about offering choices, their underlying goal is to direct viewers “toward content more likely to keep them engaged and subscribed” (Arnold, 2016, 99). Systems that aim to maximise user “engagement” (eyeballs capture for advertising) encourage to the continuous foregrounding of “fresh” (Covington et al, 2016) and “viral” content (Jiang et al, 2014)—this has been shown to have hugely detrimental effects, from the traumatising of children (Bridle, 2017) to the spreading of white nationalist ideology (Lewis, 2018) . As “anticipatory design” (Shapiro, 2015) shows, the idea of choice design may result in the removal of options on the pretext of making decisions on users’ behalf for—allegedly—their own benefit. Behind a rhetoric of convenience, the notion of relieving users from “decision fatigue” masks a political project. Our concern is not to challenge the need for an ordering system—with over 400 hours uploaded every minute and 1.8 billion logged-in monthly users on YouTube it is hard to argue that some interface is required to parse this content. Instead we want to critically examine the politics embedded in the design of systems that predict “what to watch next” and use this critique to underpin our design interventions.

The designed artefacts proposed here – experiments in mapping YouTube recommendations, titled *Architectures of Choice Vol.1 Youtube* – are a programme of work toward a visual understanding of ‘choice design’. Their aim is twofold: to show how architectures of choice are being designed in our everyday, impacting on the construction of algorithmic identities (Cheney-Lippold, 2011); and to draw attention to the complexities of (and negotiations needed for) studying them through a critical practice. These experiments are framed by three theoretical perspectives: the idea of algorithmic governmentality to foreground the politics implicit in choice design (Rouvroy, 2016); ideas concerning maps and diagrams as a way to extract methodological friction from existent mapping systems (Deleuze and Guattari, 1988); a critique of design’s lack of engagement with these issues through an assessment of the positivist ontology of data visualization (Drucker, 2011).

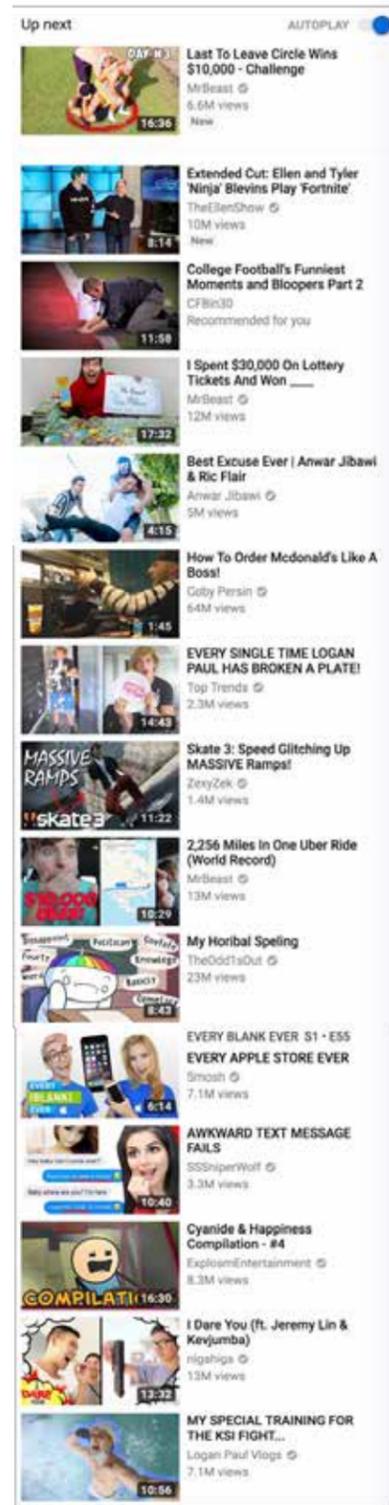


Figure 1. Youtube “Up Next” recommendations for the video “Saying Logan Paul 100,000 Times” by MrBeast. Screen capture by the authors on 19.10.2018, url: https://www.youtube.com/watch?v=_FX6rml2Yjs

1. Algorithmic governmentality: the calculative ontology of prediction

Every time a user is offered a series of recommendations generated by a digital platform - Google, Facebook, Amazon or YouTube - they are being worked on by a technology (ranking algorithms) that operates by analysing data on past behaviours to forecast future behaviours (Covington et al, 2016). However, predictive analytics is more than this. It *orients* potential tendencies. What media theorist Mark Hansen describes as “the calculative ontology of prediction” (Hansen, 2014, 38) – the impact of technomedia on the texture of human experience and on any subsequent behaviour – is a systemic way of designing tendencies and propensities at a precognitive level, in ways that are not fully accessible to conscious cognition and perceptual awareness. Our forthcoming behaviours are predicted not simply before they happen, but before we might even think of making them happen. Thus, the evaluation of how systems of choice are designed concerns also their role in building identities - what philosopher Antoinette Rouvroy calls regimes of algorithmic governmentality (Rouvroy, 2016). This is how recommendations work: when Youtube or Netflix suggest what to watch next, they not only build user profiles, but generate social identities drawn on a multitude of fragmented points. In other words, *prediction* turns into *prescription*. Classifiers do not extract categories from a neutral and objective standpoint, but actively *create* them (Cheney-Lippold, 2011). The granularity of captured user data is used to recursively construe (and obscure) a live-feed of our every single subject position. By simultaneously reflecting back and building up, this process is a continuous modulation of the subject into recursive discrete data sets, endlessly distributed on a multitude of access points - what Gilles Deleuze in his *Postscript on Control Societies* calls the ‘dividual’ (Deleuze, 1995, 180).

2. Mapping YouTube

The intricate architecture of YouTube recommendations, and the challenges posed by its sheer volume of content, are manipulated by engineering decisions concerning which type of content is promoted on the platform (Covington et al, 2016). This is a notoriously opaque process (partly because of non-disclosure agreements) for the engineers themselves who have only limited understanding of how the system predicts - let alone the public (Burrell, 2016). The secrecy surrounding YouTube recommendation system, and its lack of accountability, has prompted several initiatives to ‘map’ it from the outside: reverse engineering strategies (Albright, 2017), ‘manual’ techniques (Lewis, 2018), and most notably the *AlgoTransparency* project designed to expose the bias in recommendations after the 2016 US presidential election (Chaslot et al, 2016, Lewis and McCormick, 2018).

These examples seek to gather data/evidence in order to hypothesise on how the system might be producing detrimental effects. All three use network graphs, which model data as nodes (e.g. channels or individual videos) and relations (e.g. recommendations). While networks are well suited to mapping the the complexity of such as vast recommendation system - arguably the only available way to model such data - their claim to providing any sense of ‘transparency’ does not scale well at all. The hand crafted graph of a small number of actors used by Lewis (2018) shows a complex web of co-appearances in videos and remains relatively readable. When attempting to map Youtube in its totality however, *AlgoTransparency’s* Youtube Map becomes nothing more than a “hairball” or “spaghetti bowl” - hardly revelatory about the processes behind them (Bounegru et al., 2017), in fact counter-producing a sort of visibility-rich opacity. *Seeing*, in this case, is not necessarily knowing (Ananny and Crawford, 2016).

3. Designed artefacts: speculative diagrams for mapping YouTube recommendations

We approach the space of Youtube mapping by building on, and being mindful of, these insights. The prototypes we show and discuss here use automated techniques to collect, store and visualise recommendations as network data. They must be taken as explorative diagrams that attempt to expose the logics of algorithmic capture. It must be clarified that the aim of our interventions is not to claim solutions to the shortcomings outlined above. Rather, the aim is to open up paths of enquiry that confront the predictive apparatus and to keep on evidencing the challenges embedded in the tools through which this confrontation happens.

Collection

Data are collected with a headless browser (Selenium) to simulate user interaction with web-pages and 'click' through recommendations. This allows for quick prototyping using the browser's developer tools to identify parts of the page we are interested in collecting or interacting with. Starting from the YouTube homepage, with its list of new videos and topics, we follow links to video pages and recursively list recommendations. Three different 'probe' designs balance the breadth and depth of the mapping in different ways: *ripple.py* [fig. 2.a.] attempts to follow all recommendations for a given number of recursions. This exhaustive approach was very time consuming, and the resulting visualisations (see below) became unreadable after the first recursion. A drastically simpler approach is taken in *simple_digger.py* [fig.2.b.] which selects only one video at random in the recommendations list and repeats the process for a given number of times. This approach was also unsatisfactory, as it presented single paths which fail to capture the interconnected nature of the recommendations made evident by the previous experiments. *Digger.py* [fig.2.c.] is a compromise which captures all of the recommendations at each step before choosing one at random.

The collected data are stored in a graph database (Neo4j) where videos are represented by nodes, and recommendations by edges connecting them. Each session is logged with a time stamp and indicates the type of probe used. This database can then be queried to return particular sessions or sub-graphs for visualisation.

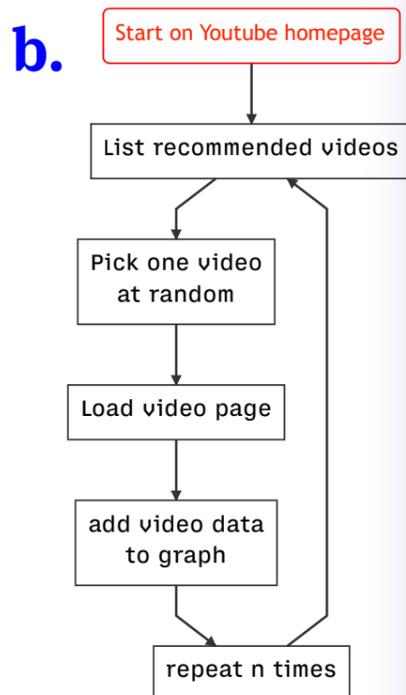
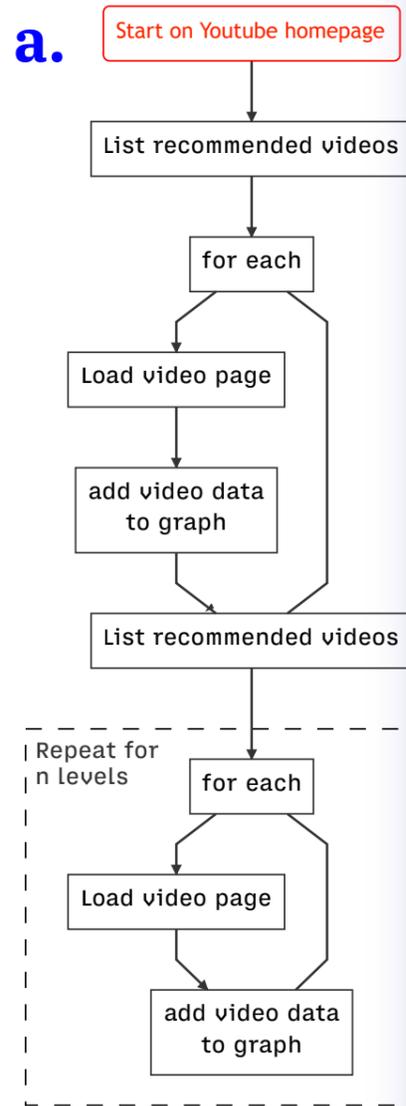
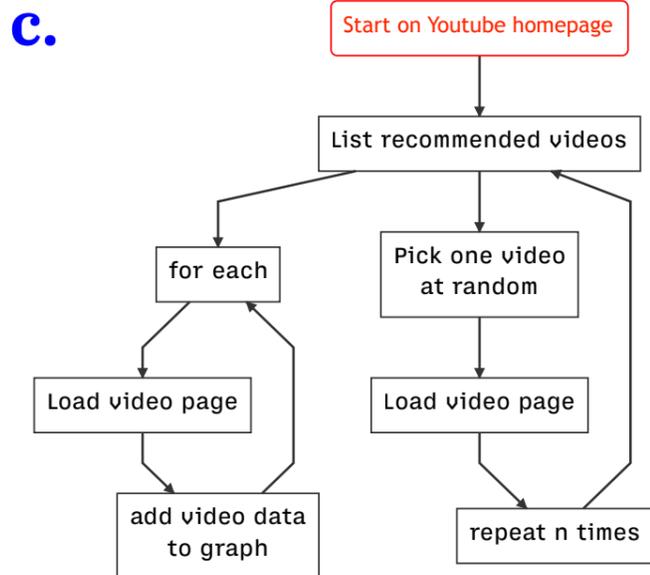


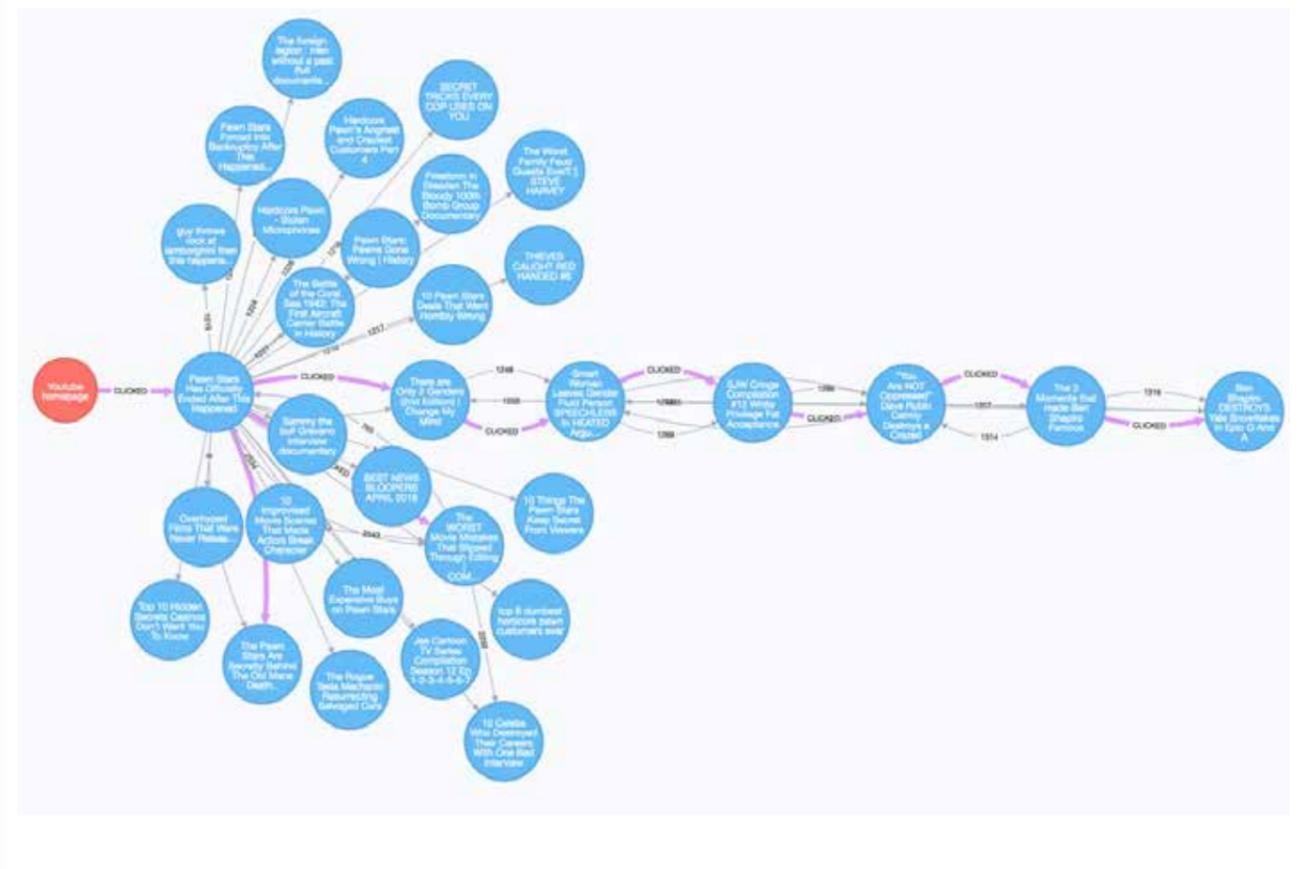
Figure 2a, b, c. Three probe designs:
 a. ripple.py
 b. simple_digger.py
 c. digger.py

Visualisation

We then query the graph database to retrieve the data from probe sessions and visualise them. Python was used to experiment with popular network visualisation packages networkx and matplotlib [fig. 5]. We also relied on the built in Neo4j browser during testing, and it provided one of the more flexible ways to explore the data, a mixture of automated layouts and manual manipulation [fig. 3]. We tested generating flowcharts for Mermaid.js [fig.4] but these quickly reached the limits of the library resulting in crashes and unreadable outputs. Finally, we tested Gephi, an open source graph analysis and visualisation software popular amongst Digital Humanities practitioners. [fig. 6].

None of these tools were satisfactory. Beyond the structural aspect of simply having too many data points to show, simple things like the lack of text-wrapping functionality to display video titles on multiple lines of text rendered our outputs unreadable very quickly. We addressed this by zooming in on the granularity of individual paths/traces and to design our own visualisation using D3.js. [fig 7 & 8]. Importantly, practical concerns for readability highlighted the performative nature of data as material and the challenges of data visualization - discussed in the next section.

Figure 3. Visualisation in the Neo4j Browser interface.



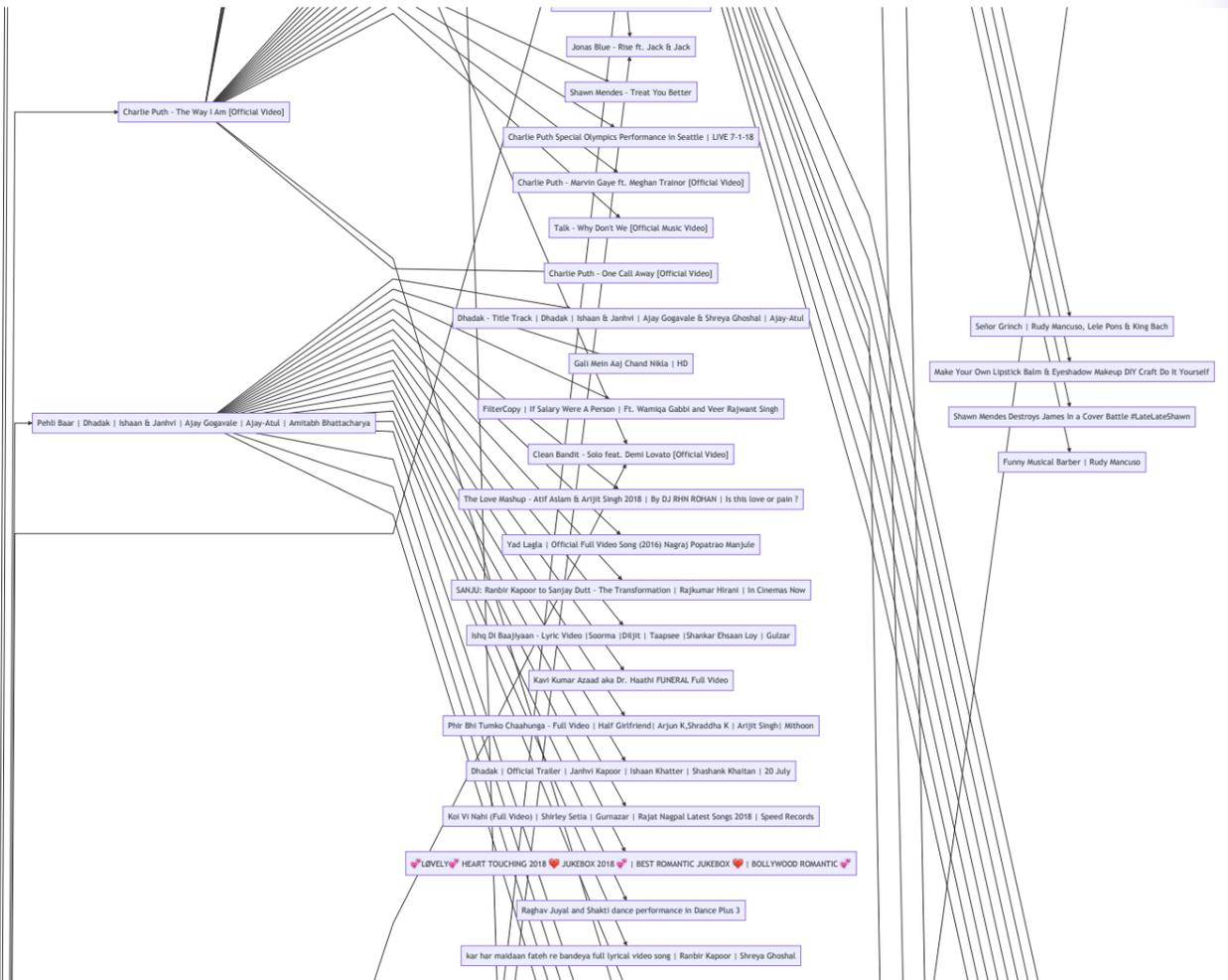


Figure 4. Visualisation using Mermaid.js

4. The positivist ontology of data visualization

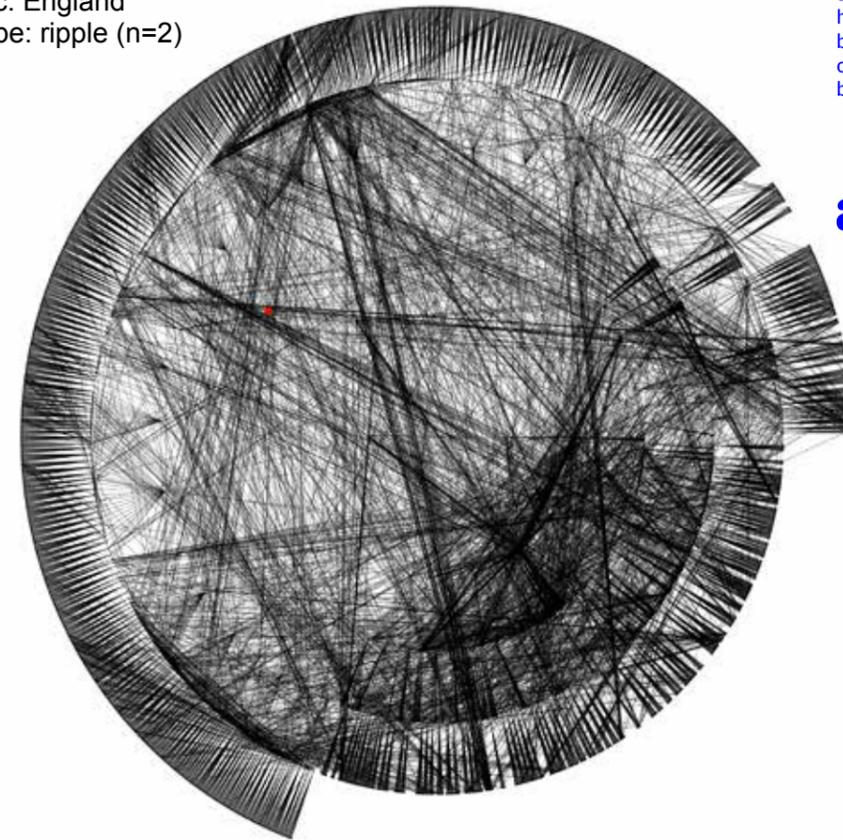
This project is caught in a paradox - which is also at the core of the digital humanities (Kitchin, 2014): how are we to critique the epistemologies of data and algorithms, while at the same time relying on them for our research? Here we bring the theoretical distinction between maps and tracing to bear on the specifics of data visualisation.

Mapping concerns imagining, and experimenting with, forms of world-building only partially embodied in, and expressed by, their given visual representation (Deleuze and Guattari, 1988). As a tool that unfolds potential (Corner, 1999), mapping emphasizes a productive future-facing capacity to visualize simultaneously what is and what is not there yet. Conversely, tracing belongs to a logic of reproduction. It copies from the same model, it is always identical to itself and therefore lacks genuine openness to change. This distinction is helpful in articulating the pitfalls of using data visualisation to expose/critique algorithmic systems, as most data visualisations claim to be tracings—neutral and objective depictions of reality.

Data visualisation has been critiqued as a process where decisions are made reflecting “assumptions, unstated (often unacknowledged) ideological perspectives and subjective judgments” (Boehnert, 2016). The choices embedded in data visualization concern not only what to include, and what to leave out, but also claims of objectivity and neutrality made by what is effectively the result of (ideological) se-

2018-07-12 22:20:20

isp: Virgin Media
loc: England
type: ripple (n=2)



a.

b.

2018-07-13 14:29:51

isp: Virgin Media
loc: England
type: ripple (n=1 + 1)

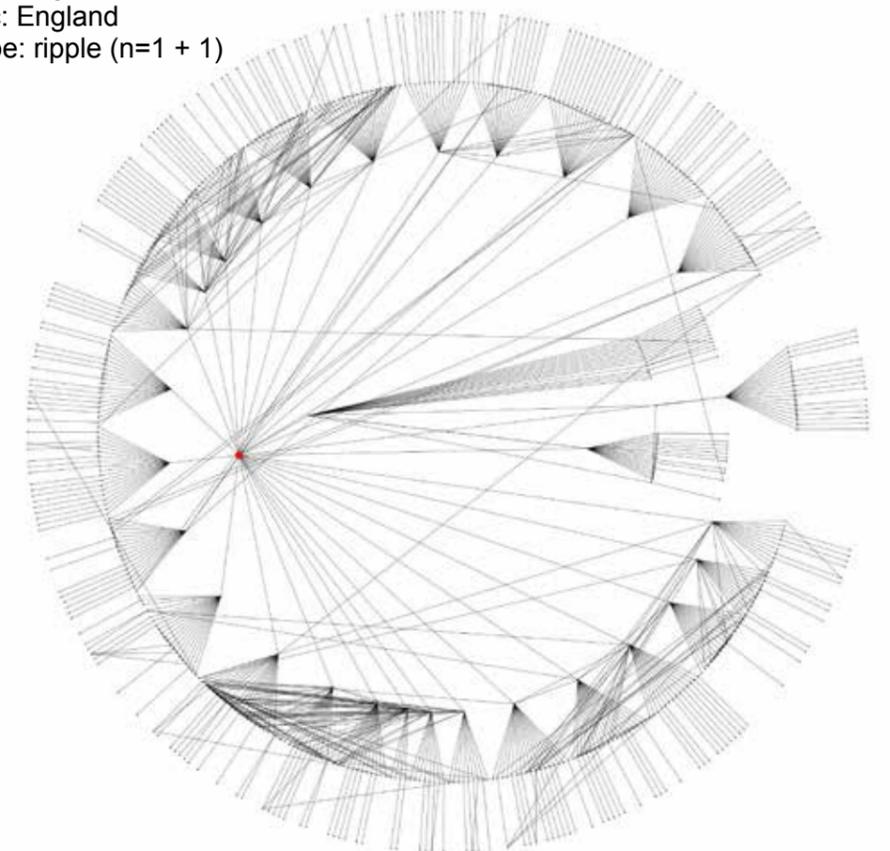


Figure 5. Visualisation of two ripple.py sessions using Python packages networkx and matplotlib. a. shows two levels of recommendations from the Youtube homepage. b. shows the first level of recommendations followed by one random pick.

lections. “Most information visualizations are acts of interpretation masquerading as presentation (...) they are images that act as if they are just showing us what is, but in actuality, they are arguments made in graphical form” (Drucker, 2014). Boehnert uses the terms “digital positivism” (over-simplistic take on how data mediates knowledges), “datawash” (data visualisation that obfuscates knowledge on controversial issues) and “darkdata”(what is not measured or datified may be more relevant than what is) - to zoom in on the political nature of data visualization practices and tools. Data visualization carries value judgments that are patently obfuscated by the blanket of digital positivism (Boehnert, 2018). Modes of “knowledge visualisation” can be used instead, which declare their ongoing situatedness and partiality and build on these to create awareness and stimulate critical insights. Drucker’s distinction between data “which are assumed to be given” and capta which are “actively taken” is a key demarcation here (Drucker, 2011). While algorithmic transparency - “designed to give audiences an appreciation of programming decisions and consequences” (Ananny and Crawford, 2016, 5) might be seen by some as a goal, it is also fraught with limitations. What matters, rather are “meanings achieved through *relations*, not revelations” (ibid.).

Our last experiment [new fig.] shows how foregrounding the relations between a collection of paths or *traces* through the recommendation system might offer new understanding. By acknowledging the impossibility of a full map (problems of scalability), this experiment does not just suggest that knowledge about the system can only be grasped at a micro-scale. Importantly, it relies on traces to build new understanding and a mode of knowledge-making that is situated, incomplete and speculative.

Figure 6 ripple.py session visualised with Gephi.

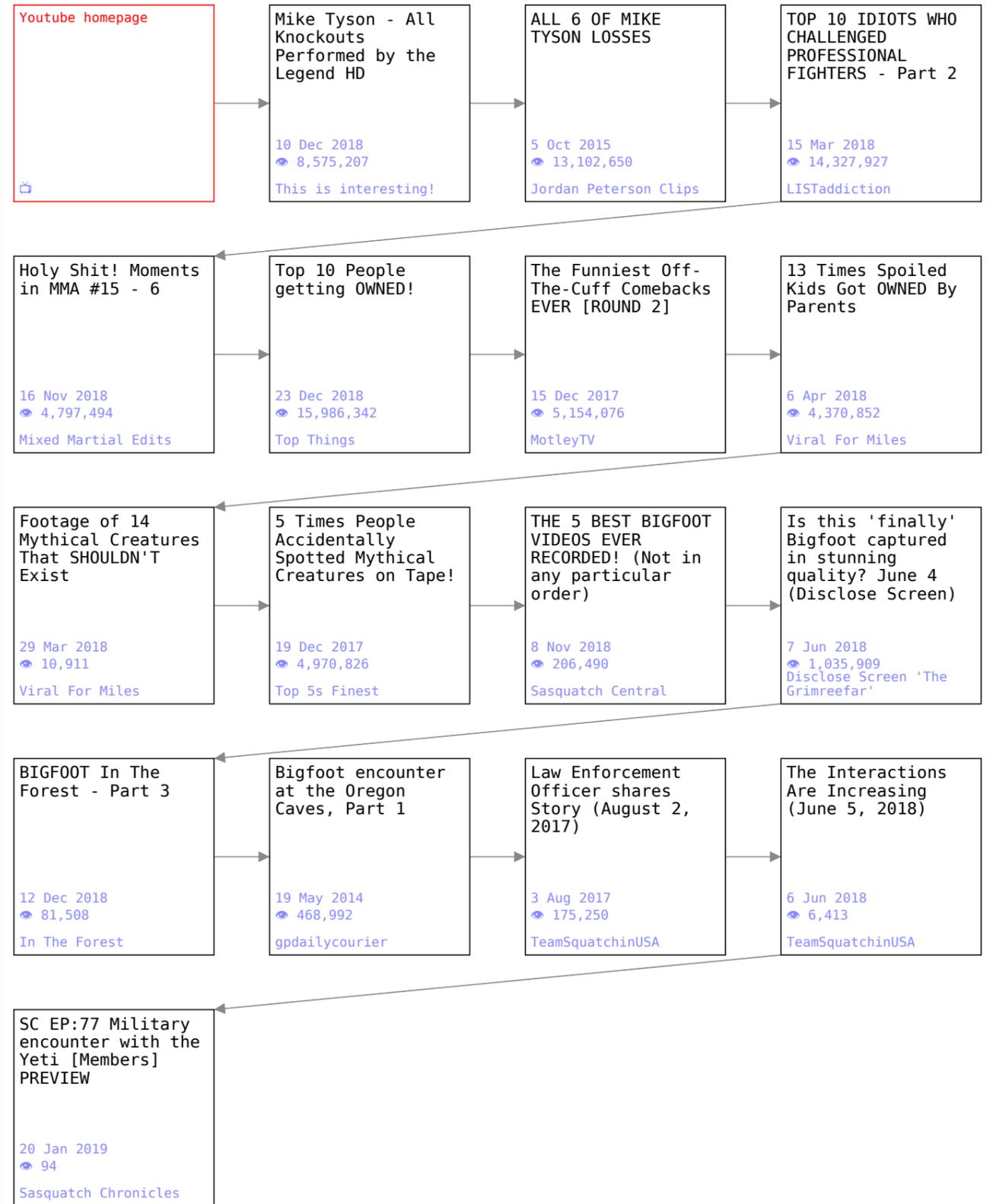
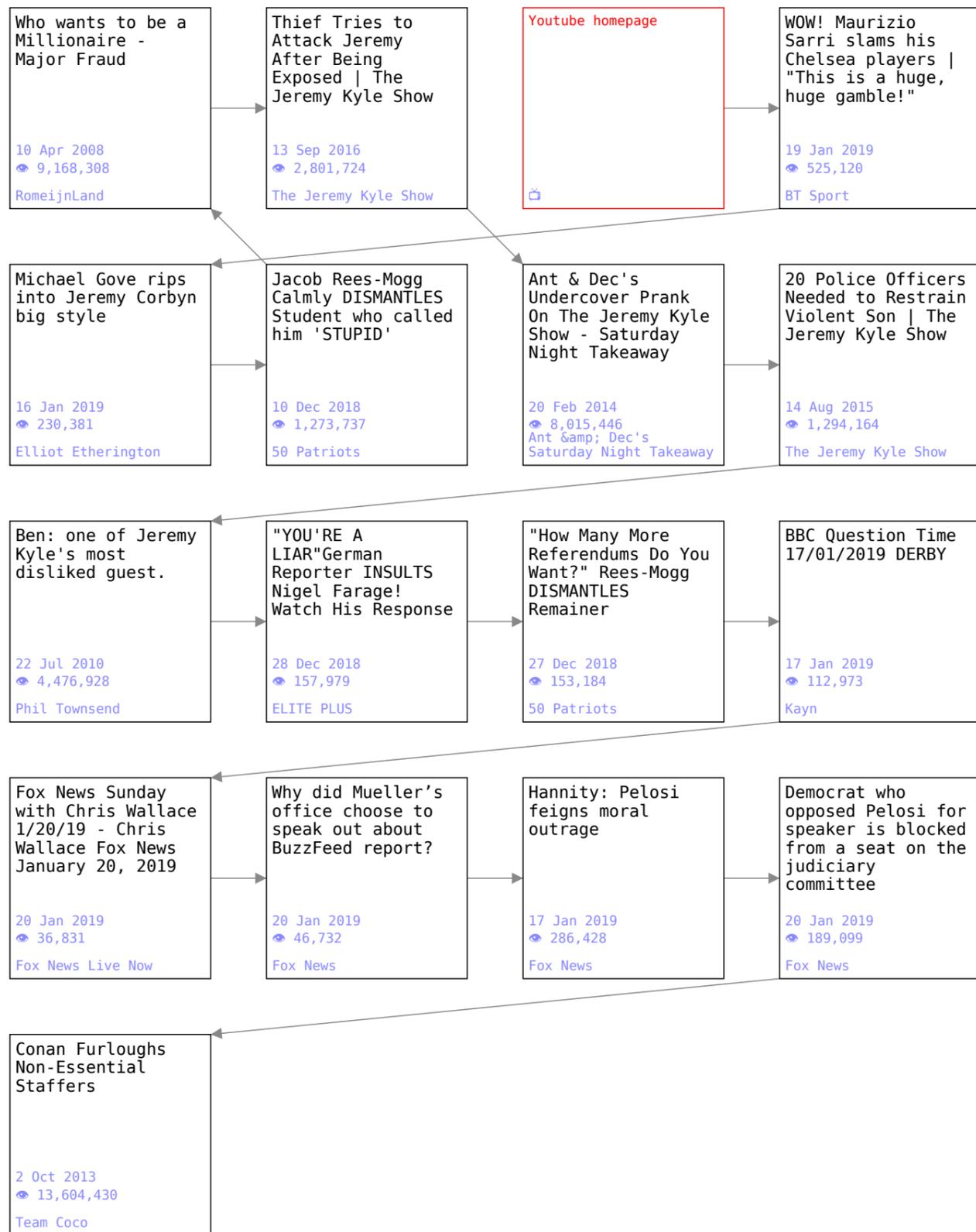


Figure 7 simple_digger.py session visualised with D3.js.

Figure 8 simple_digger.py session visualised with D3.js.



5. Reflections via mapping and diagramming

The mapping experiments presented above are diagrams of how choice design in YouTube recommendations constructs algorithmic subjects. As digital media theorist Luciana Parisi argues, this subject is “neither solely an enslaved component of machines not its deluded interactive user. Instead the subject is being reconfigured from the standpoint of a learning machine” (Parisi, 2019). Put differently, the subject emerges from forms of predictive patterning which retrofits choice (and thinking itself) as determined by its searching into the unknown. Thus, any attempt at mapping choice ought to acknowledge these mechanisms. For this reason we consciously repurpose computational tools as “cognitive prostheses” (Srnicek, 2012) to provide purchase on otherwise ungraspable systems. Such acts of building mediators between the impenetrable noise of big data and their cognitive tractability is “one of the most important areas where political art could be situated today” (Srnicek, 2012, 4), and - we add - a key area of design intervention.

Moreover, faced with vast, unintelligible systems designed to capture attention and orient choice for profit, the “inability to cognitively map the gears and contours of the world system is as debilitating for political action as being unable mentally to map a city would prove for a city dweller” (Toscano and Kinkle, 2015). How is any form of reclaiming to take place if the terrain is unknown? What practical counter-strategies are possible?

Giorgio Agamben’s notion of apparatus is illuminating here as “anything that has in some way the capacity to capture, orient, determine, intercept, model, control, or secure the gestures, behaviours, opinions, or discourses of living beings” (Agamben, 2009, 14) with a resulting dispersion and granularization of the self (Deleuze’s *dividual*, again). Only the restitution to common use of what has been captured can succeed as counter-apparatus interventions.

This is what our proposals attempt to do. They use diagrams to map the contours of the capture apparatus and crucially to offer some raw glimpses of their inside. They use diagrams as tools that by displaying how choice is designed as predictable patterning aim to be both *practical* and *speculative*. They are practical because they visualize data-driven choice design by framing it through the challenges of data positivism. They are speculative because they offer traces through which some partial and situated knowledge can be constructed, staying clear of interpretation and representation. Against the positivist claims to complete objectivity made by data visualisation and smuggled into so many of its tools, and unlike tracings (always reproductions with a likeness to their subject), traces offer opportunities for what historian Carlo Ginzburg calls ‘conjectural knowledge’ (Ginzburg 1980) - a mode of knowledge that is qualitative, contingent and incomplete, akin to divination (chance-led inquiry into the unknown) or the conjectures of the tracker (sensing an animal/event that cannot be directly experienced).

Conclusion

We think that our programme of work is important for design research for key reasons. Diagrams precede the instalment of a technology. For a technology to be possible, in other words, its materiality/machinery/toolkit have to be 'selected' by a diagram. By extracting the diagram expressed by current technologies of choice design we might develop counter diagrams that situate and dislodge its mechanisms of preferences and recommendations. To do so, an understanding of diagramming is essential especially for designers who do not design the technologies that they use.

Still, these initial and ongoing experiments are attempts at building relations, not revelations. In demonstrating the project's feasibility, they provide an understanding of the tools widely used in the prior art on network mapping. They highlight how tools such as Gephi have many limitations for our purposes, starting with built-in assumptions that make tracing, rather than mapping, their default mode of operation. Importantly, they emphasize the challenge of using data to critique data systems, thus offering, we hope, a constructive reflection to an issue that is central to doing research through design, that can be productively addressed through the theoretical framing of the predictive regime of algorithmic governmentality and the preemptive politics it enforces.

On this basis, some questions remain: how do we visually acknowledge the limitations of our system, conveying its partiality and incompleteness? How do we communicate the indeterminacies and uncertainties through meaningful practice and knowledge production? Which other 'architectures of choice' might be diagrammed with this approach?

Acknowledgements

The project *Architectures of Choice Vol.1 Youtube* was initially commissioned for *Everything Happens So Much*, an exhibition at the London College of Communication (LCC) for London Design Festival 2018. It was a collaboration between Supra Systems Studio at LCC and BBC R&D.

David Benqué's doctoral research at the Royal College of Art is supported by a scholarship from Microsoft Research Cambridge (UK).

References

- Agamben, G. (2009) *What is an apparatus?* Stanford, Stanford University Press
- Albright, J. (2017) 'FakeTube: AI-Generated News on YouTube' <https://medium.com/@d1gi/faketube-ai-generated-news-on-youtube-233ad46849f9>
- Ananny, M. and Crawford, K. (2016) Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability. *New Media & Society*. 1–17
- Arnold, S. (2016) Netflix and the Myth of Choice/Participation/Autonomy, in McDonald, K. and Smith-Rowsey, D. (eds), *The Netflix Effect: Technology and Entertainment in the 21st Century*, Bloomsbury
- Boehnert, J. (2018) *Design, Ecology, Politics: Towards the Ecocene*. London, Bloomsbury
- Boehnert, J. (2016) Data Visualisation Does Political Things. *DRS2016* <https://drs2016.squarespace.com/387>
- Bounegru, L. et al. (2017) 'Narrating Networks. Exploring the affordances of networks as storytelling devices in journalism', *Digital Journalism*, 5(6), pp. 699–730. doi: 10.1080/21670811.2016.1186497.
- Bridle, J. (2017) Something is wrong on the internet. Medium <https://medium.com/@jamesbridle/something-is-wrong-on-the-internet-c39c471271d2>
- Burrell, J. (2016) How the machine 'thinks': Understanding opacity in machine learning algorithms. *Big Data and Society*. 3,1 1-12
- Chaslot, G., Montcoudiol, A., Ledésert, S., Wielonsky, N., Bardolle, F. and Grac, M. (2016) *AlgoTransparency* [Online]. Available at <http://algotransparency.org/> (Accessed 3 October 2018).
- Cheney-Lippold, J. (2011) A New Algorithmic Identity: Soft Biopolitics and the Modulation of Control. *Theory, Culture & Society*, 28, 6. 164–18
- Corner, J. (1999) The agency of mapping: speculation, critique and invention In *Mappings*, Denis Cosgrove ed. London, Reaktion 213-252
- Covington, P., Adams, J. and Sargin, E. (2016) Deep Neural Networks for YouTube Recommendations. *Proceedings of the 10th ACM Conference on Recommender Systems*, RecSys Boston, MA, ACM, 191–198
- Deleuze, G. (1995) *Negotiations*. New York, Columbia University Press
- Deleuze, G. and Guattari, F. (1988). *A Thousand Plateaus. Capitalism and Schizophrenia*. London, Continuum
- Drucker, J. (2014) *Graphesis. Visual Forms of Knowledge Production*. Harvard University Press
- Drucker, J. (2011) 'Humanities Approaches to Graphical Display', *DHQ Digital Humanities Quarterly*, vol. 5, no. 1 [Online]. Available at <http://www.digitalhumanities.org/dhq/vol/5/1/000091/000091.html>.
- Ginzburg, C. (1980). 'Morelli, Freud and Sherlock Holmes: Clues and scientific method'. *History Workshop*, 9. Oxford, Oxford University Press. pp. 5–36
- Hansen, Mark B.N. (2014) The Operational Present of Sensibility. *The Nordic Journal of Aesthetics* 24, 47. 38-53

- Jiang, L., Miao, Y., Yang, Y., Lan, Z. and Hauptmann, A. G. (2014) 'Viral Video Style: A Closer Look at Viral Videos on YouTube', Glasgow ACM Press, 193–200
- Kitchin, R. (2014) Big Data, new epistemologies and paradigm shifts. *Big Data & Society*. 1–12 <http://journals.sagepub.com/doi/10.1177/2053951714528481>
- Lewis, R. (2018) Alternative Influence: Broadcasting the Reactionary Right on YouTube, Media Manipulation Initiative, Data & Society. <https://datasociety.net/output/alternative-influence/>.
- Lewis, P. and McCormick, E. (2018) 'How an ex-YouTube insider investigated its secret algorithm', *The Guardian*, 2nd February. <https://www.theguardian.com/technology/2018/feb/02/youtube-algorithm-election-clinton-trump-guillaume-chaslot>
- Marenko, B. and Benque, D. (2018) 'Speculative diagrams: plotting to reclaim algorithmic prediction.', in *Art, Materiality and Representation*. British Museum, London: Royal Anthropological Institute. Available at: <https://theairpump.davidbenque.com/speculative-diagrams-plotting-to-reclaim-algorithmic-prediction/>.
- O'Sullivan, S. (2016) On the diagram (and a practice of diagrammatics), in *Situational Diagram*, eds. Karin Schneider and Begum Yasar, New York: Dominique Lévy, 13-25
- Parisi, L. (2019) The alien subject of AI. *Subjectivity*. 10.1057
<https://doi.org/10.1057/s41286-018-00064-3>
- Rouvroy, A. (2016) Algorithmic Governmentality: Radicalization and Immune Strategy of Capitalism and Neoliberalism? *La Deleuziana* 3: "Life and Number", 30-36
<http://www.ladeleuziana.org/wp-content/uploads/2016/12/Rouvroy2eng.pdf>
- Shapiro, A. (2015) 'The Next Big Thing In Design? Less Choice' *Fastcodesign*. <https://www.fastcodesign.com/3045039/the-next-big-thing-in-design-fewer-choices>.
- Srnicek, N. (2012) 'Navigating Neoliberalism: Political Aesthetics in an Age of Crisis', *The Matter of Contradiction: Ungrounding the Object* https://www.academia.edu/1925994/Navigating_Neoliberalism_Political_Aesthetics_in_an_Age_of_Crisis.
- Toscano, A. and Kinkle, J. (2015) *Cartographies of the absolute*, Winchester & Washington, Zero Books