The Data Don't Speak for Themselves: The Humanity of VOD Recommender Systems

Giorgio Avezzù, Università Cattolica del Sacro Cuore

Abstract

There is a widespread myth and rhetoric, even in academic discourse, about data and VOD recommender systems, especially with regard to the notion of automation and the innocence of this presumed automation. Behind this rhetoric lies the de-humanization of machine computation, i.e. the removal of all the processual, decisional, 'oriented' aspects informing every online recommender system. This essay focuses on content-to-content video recommendations, which are based on patterns of similarity between different contents, and it intends to show that there is nothing neutral — even in the most seemingly 'objective' form of video recommendation. The aim is to rediscover those very processual elements of the 'data supply chain' — regarding how metadata are created and collected, and how algorithms are configured — so as to make them critically observable again: the funnels, decision points, the multiple layers of human mediation and filtering, in both their relevance and sensitivity.

Rhetoric and Myths

Recommender systems are used in VOD (video-on-demand) platforms in order to help users find videos to watch, and they are considered crucial to the good functioning of such platforms. Netflix, for example, maintains that recommendations account for about 80% of all streaming hours on their platform, as opposed to the 20% taken up by contents actively searched by users.[1] Active searches and recommendations represent two alternative routes to the 'discovery' of contents: one active and informed by human agency, the other passive and machine-assisted.

In the mythically inflected scenario put forth by Netflix and other media providers, recommender systems constitute the backbone of online streaming

[1] Carlos A. Gomez-Uribe and Neil Hunt, 'The Netflix Recommender System: Algorithms, Business Value, and Innovation', *ACM Transactions on Management Information Systems*, 6.4 (2015), 1–19 (p. 5).

services. A narrative thus emerges: the need for such systems, we are told, stems from the 'increasing number of choices' that contemporary audiences face. Machines — the myth goes on to argue — are capable of navigating the wealth of potentially available items far better than their human counterparts, including so-called 'human experts'. The founding narrative of VOD platforms follows an evolutionary logic: plots are seen to have grown in number and complexity over the centuries, from the basic stories of prehistoric cave dwellers to those of our times. Fuelled by technological advances, storylines have multiplied and become more 'engaging', and, Netflix says, they are now more varied and widely distributed 'than ever before', to the point where there are just too many for us to pick: 'humans are surprisingly bad at choosing between many options.' But while human beings are likely to be overwhelmed by such abundance, a machine can easily choose for them. Moreover, recommender systems are seen as intrinsically 'democratic', because they allow direct access to a 'long tail' of contents, and especially because they do so in an 'automatic' and 'machinic' way:

> Recommender systems can democratize access to long-tail products, services, and information, because machines have a much better ability to learn from vastly bigger data pools than expert humans, thus can make useful predictions for areas in which human capacity simply is not adequate to have enough experience to generalize usefully at the tail.[2]

I believe — and am not alone[3] — that a widespread myth (and attendant ideology) can be traced, even in academic discourse, where arguments are made about data and content recommendation, especially with regard to the notion of *automation*. Take, for example, a recent article by Lev Manovich on the importance of data analytics in the contemporary mediascape, dominated by Big Data and data companies.[4] The word 'automation' and its derivatives are used 34 times just in this one essay. On top of that, they are even misused: the over 76,000 genre categories of Netflix's recommendations system are *not* created through computational analysis of media content, as Manovich seems to believe, but by *human* 'taggers', using a 36-page training manual and a tagging system conceived by other, equally human analysts. These employees are tasked to describe films and series, down to the most minute narrative details, including, for example, the amount of gore or romance, plot conclusiveness, the 'social acceptability' of the protagonists and so forth.[5] In fact, complete automation in the analysis of contents is far from being a reality.

To be sure, automated analysis has its uses, and can be especially suited to

[2] Ivi, pp. 1–2, p. 16.
[3] See footnote 12.
[4] Lev Manovich, '100 Billion Data Rows per Second: Media Analytics in the Early 21st Century', *International Journal of Communication*, 12 (2018), 473–88.
[5] See Alexis C. Madrigal, 'How Netflix Reverse Engineered Hollywood', *The Atlantic*, 2 January 2014, <https://www.theatlantic.com/technology/archive/2014/01/how-netflix-reverse-engineered-hollywood/282679/> [accessed 25 July 2017].

certain tasks. It could be applied very effectively, for example, to extract data on the colours used in certain films and then to group those same films in clusters, according to their palettes, or to gather information on other elements such as cutting rates, motion or sound features — companies such as Vionlabs also try to correlate data about action, lightning, colour and sound (e.g. the amount of dialogue and its volume) with the 'feeling' of a movie, the way it affects the spectator. A semantic engine could even be able to identify the subject of a script, and what the main themes are. No doubt, these and other similar applications are bound to galvanize those in favour of applying quantitative analysis to film. And again, there is no debating that the possibilities offered by computational stylometry can be very interesting: a statistical analysis of the various types of camera shots — the kind of things Barry Salt used to like[6] — based on automatically generated data is a compelling prospect, and not at all impossible to imagine even today. Most likely, however, machines would struggle with other aspects of film analysis, especially those not as easily related to identifiable discrete units.

What is more, such a level of automated analysis does not appear to be even remotely as widespread and fundamental for the running of VOD platforms and their recommender systems today as some enthusiastic commentators seem to believe. Nor, for that matter, is Amazon Prime Video using face-recognition algorithms yet, as some seem to imply:[7] the 'X-Ray' feature, despite the technologism of its name, uses IMDb data, and relies on *human* work (not just human review or curation) to describe the characteristics of each scene as it is streamed: music, trivia, filming location and names of the actors present in the frame. This explains why, in *Forrest Gump* (Robert Zemeckis, 1994), X-Ray designates the titular character as Tom Hanks even in those sequences where the main character is a child, played by an obviously different actor: a tag has been applied to the character, and the match is not the outcome of face recognition.

Coupled with the myth of automation, and equally widespread, is the myth of the *innocence* of this presumed automation: 'I believe', Manovich says, 'that computing and data analysis technologies are *neutral*. They don't come with some built-in social and economic ideologies and effects.'[8] One senses in these words a blind faith in the self-evidence of data, the conviction that automated systems and algorithms will be capable of delivering (finally) unequivocal interpretations, more so than any human analytic framework. Another respected media guru, Chris Anderson, expresses a similar sentiment as he celebrates Big Data in his *The End of Theory*:

[6] See Barry Salt, *Film Style and Technology: History and Analysis* (London: Starword, 1983).
[7] See the explanation provided by software engineer Christopher Brian at <https://www.quora.com/How-does-Amazon-IMDB's-X-Ray-work> [accessed 25 July 2017].
[8] Manovich [emphasis in the original]. These sentences were included in the article's 'Fall 2015 – Spring 2016' draft version, which was available on the author's Academia.edu page and on his personal website (<http://manovich.net/index.php/projects/media-analytics> [accessed 25 July 2017]). However, they are no longer present in the revised version of the article. Manovich's 'belief' still seems to inform the text, though (see p. 482).

> With enough data, the numbers speak for themselves. […] We can stop looking for models. We can analyze the data without hypotheses about what it might show. We can throw the numbers into the biggest computing clusters the world has ever seen and let statistical algorithms find patterns […]. Science can advance even without coherent models, unified theories, or really any mechanistic explanation at all.[9]

Anderson's words demonstrate the strength of what I see as the contemporary rhetoric of data, data driven recommendations and algorithmic systems, as well as the anti-humanistic diffidence that appears to be widespread in current computational reasoning. What interests me is the scientistic inflection that transpires in these accounts, the stress on the necessity and perfection of automation, on its neutral and democratic character. The few excerpts I quote here are representative of a much larger discursive trend, which denies the presence of any ideological, theoretical or otherwise *oriented* aspect in the configuration of data-driven systems, be they used for the analysis and interpretation of contents and tastes, or to provide recommendations.

In their very wording, statements like Manovich's cannot but remind the film theorist of controversies, dating back to the 1960s and '70s, about the neutrality of the cinematic apparatus, such as those inspired by Jean-Patrick Lebel's dismissive claim that 'the camera […] is an instrument which is ideologically neutral inasmuch as it is an instrument, an apparatus, a machine. It rests on a scientific basis and it is not constructed according to an ideology of representation.'[10] Among those who joined the debate in response to Lebel were Marcelin Pleynet and Jean-Louis Baudry, who intended to demonstrate precisely the opposite, exposing the ideological underpinnings of the cinematic apparatus, informing the 'scientific basis' of the *dispositif*.

The analogy I suggest here between those arguments and mine, in relation to data and the non-neutrality of recommender systems, is less far-fetched than it might seem. After all, the role these systems play allows us to see them acting very much in the way of strategic apparatuses, translating specific ideas about cinema (and its spectators) into 'conditions of recommendability'. That is to say, if we paraphrase Foucault's definition of the *episteme*, that they turn those ideas and assumptions into the pre-conditions 'which permit of separating out from among all the [recommendations] which are possible those that will be acceptable'[11] — a point to which I return later in this article.

[9] Chris Anderson, 'The End of Theory: The Data Deluge Makes the Scientific Method Obsolete', *Wired Magazine*, 16.7 (2008), <https://www.wired.com/2008/06/pb-theory/> [accessed 25 July 2017].

[10] Jean-Patrick Lebel, 'Cinéma et idéologie', *La Nouvelle Critique*, 34 (1971), p. 72.

[11] See Michel Foucault, *Power/Knowledge: Selected Interviews and Other Writings 1972–1977*, ed. by Colin Gordon (London: Harvester, 1980), p. 187. It originally reads '[…] all the statements […]'.

Reification v. Processuality

For now, I begin by positing that behind the myth of algorithmic recommendations, and the rhetoric of automation, lies the *reification* of machine computation, a phenomenon which in turn relies on the removal from sight of all its *processual* aspects. The de-humanization of technology, along with the rhetorical suppression of its decisional, operational and relational aspects, accounts in my view for the anti-humanistic and post-theoretical views that I outlined above, of both data and data-based recommender systems. Indeed, a much more critical approach to these issues is needed — such as the one articulated by David Berry in *Critical Theory and the Digital*,[12] which I share, and which I endeavour to apply here, albeit with a more limited scope.

Refocusing our attention on the processual elements of the supply-chain of data can easily pave the way for a full evaluation of all sorts of human, theoretical and ideological aspects. Here, however, I prefer to postpone that much needed evaluation, and engage instead in a preliminary survey, so to speak. My intention is to rediscover those very processual elements, so to make them *critically observable* again: the funnels, decision points, the multiple layers of human mediation and filtering, in both their relevance and sensitivity.

It should be noted, in fact, that the algorithms behind every online recommender system — in their initial setup, during their actual operation, and in the results they generate — must, in order to function, unavoidably contaminate their machinic perfection with factors that are, strictly speaking, *human*. These factors should be identified and acknowledged as such — that is to say, as elements that are neither automatic nor machinic (including, for example, rules, editorial filters, strategic decisions, logical assumptions and operations). Those elements, in turn, can be used to investigate deeper layers of meaning. Hidden as they are,

[12] See David Berry, *Critical Theory and the Digital* (New York: Bloomsbury, 2014), p. 10. Closer to our subject are Patrick Vonderau, 'The Politics of Content Aggregation', *Television & New Media*, 16.8 (2015), 717–33; Ramon Lobato, 'The Politics of Digital Distribution: Exclusionary Structures in Online Cinema', *Studies in Australasian Cinema*, 3.2 (2009), 167–78; Ted Striphas, 'Algorithmic Culture', *European Journal of Cultural Studies*, 18.4-5 (2015), 395–412; Blake Hallinan and Ted Striphas, 'Recommended for You: The Netflix Prize and the Production of Algorithmic Culture, *New Media & Society*, 18.1 (2016), 117–37. I fully endorse the critical perspective on big data, computational methods, algorithms and digital humanities exemplified by articles and volumes such as Rob Kitchin, *The Data Revolution: Big Data, Open Data, Data Infrastructures and Their Consequence*s (Los Angeles: SAGE, 2014); Stephen Ramsay, *Reading Machines: Toward an Algorithmic Criticism* (Urbana: University of Illinois Press, 2011); *'Raw Data' Is an Oxymoron*, ed. by Lisa Gitelman (Cambridge, MA: The MIT Press, 2013); Tarleton Gillespie, 'The Relevance of Algorithms', in *Media Technologies: Essays on Communication, Materiality, and Society*, ed. by Tarleton Gillespie, Pablo J. Boczkowski and Kristen A. Foot (Cambridge, MA: The MIT Press, 2014), pp. 167–93; Danah Boyd and Kate Crawford, 'Critical Questions for Big Data: Provocations for a Cultural Technological, and Scholarly Phenomenon', *Information, Communication & Society*, 15.5 (2012), 662–79; and also in most of the articles published in *The Datafied Society: Studying Culture through Data*, ed. by Mirko Tobias Schäfer and Karin van Es (Amsterdam: Amsterdam University Press, 2017).

those 'human biases' grant us a precious insight into what ideas and assumptions are at play in determining what contents VOD platforms offer, and how they conceive of their users — assumptions that are usually unspoken, unconscious and taken for granted. I refer, for example, to those underlying the decision of what factors ought to be counted when determining the degree of similarity (and dissimilarity) between different films, or between different user profiles.

All of this applies both to *collaborative filtering* and *content-based* algorithms. The first type uses *play data* to recommend content and build clusters of users based on their presumed tastes. In doing so, these algorithms could be seen to normalize and simplify different psycho-social profiles; also, they could equate patterns of viewing based on superficial similarities, without taking into account the possibility that identical behaviours might be the outcome of very different rationales. Yet, a discussion of the limitations of collaborative algorithms is beyond the scope of this article.

Rather, I want to focus on the second type, the so-called *content-based* algorithms, which rely on *metadata*, that is to say data that describe contents, their characteristics and features, and *not* on play data generated by the users' viewing patterns. These algorithms may look more objective than the others, but they are *not*. Among content-based algorithms we have, for example, LSAs, which are semantic algorithms capable to infer (among other things) that if two different films cast a certain actor, then the directors of the two films in question are correlated, even when their other films do *not* cast that same actor.[13] More significant for my present argument, however, is another content-based algorithm called kNN. The kNN content algorithm uses information about films in order to assess the presumed *similarity* between them. In order to do so, the algorithm puts all the films it needs to assess on a map, or, rather, a two-dimensional translation of high-dimensional values and relations (fig. 1). Films are displayed as points (or vectors), and the nearer a point is to another, the highest the similarity (NN stands for 'nearest neighbours', and k is the number of items considered).

The resulting assessments are then used to compile rows of content-to-content recommendations on VOD platforms. On Netflix, kNN is used, for example, to fill the 'Because You Watched' row of videos: Netflix IT experts refer to it as a 'video-video similarity', or 'sims' algorithm. In this context, the term 'row' refers to an array of contents, presented to the user as a scrollable list of movie images or posters, displayed in a horizontal line. It is also worth noting that this type of content-to-content recommendation is rarely experienced by the users in its purest form, that is, without the interference of any custom filter. An exception to this is offered by Infinity, a VOD platform owned by the Italian media company Mediaset. At the time of writing (summer 2017), a content page on Infinity

[13] For the sake of simplicity, all the scenarios I discuss here refer exclusively to films, even though VOD platforms obviously offer a wider variety of audiovisual content, and despite the fact that the relevance of my observations can be extended to other types of content.

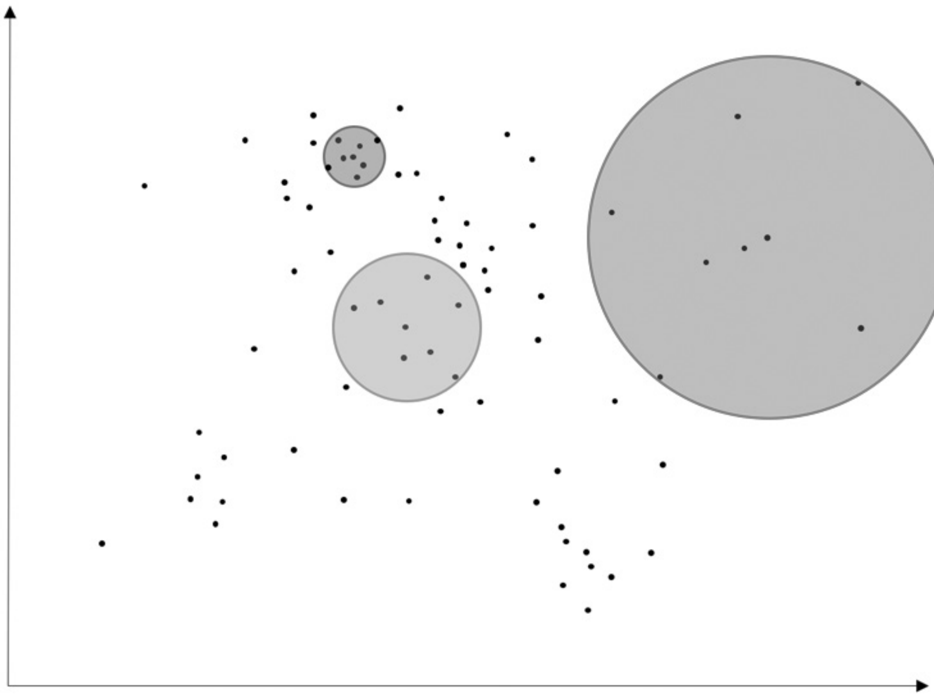

Fig. 1. A hypothetical kNN algorithm map of film similarity. Each circle includes the seven most similar films to another one (k=7).

displays what appears to be a content-to-content row in its pure form, i.e. an unfiltered list that always contains the same matches for a given film, regardless of the user. As far as OTT (Over The Top) services go, moreover, Infinity is still under development, and is thus particularly well suited to prove my point — that is, its *humanity* is more apparent than in other, more 'polished' video streaming platforms.

It should be noted, however, that I do not intend to present Infinity as a fully-fledged case study, also because VOD services are constantly evolving: rules, criteria, the variables used to direct recommendations, all these critical elements are always in flux, as are the catalogues legally available to each platform. Such volatility makes it extremely difficult to produce a snapshot of VOD services at any given point in time, and any such example would be at risk of becoming irrelevant from one week to the next.

Notes on the Data Supply Chain

Rather than focusing on a specific case study, in this article I attempt to sketch an outline of the process leading to content-to-content recommendations in *its* *most general terms*. In defining such terms, however, I also draw from my personal experience in the Strategic Marketing Department at Mediaset, which, under the direction of Federico di Chio, has been engaged since 2015 in a vast project to create an archive of metadata for their catalogue of audiovisual content: a project intended, among other things and in the near future, also to improve recommender systems on the VOD platforms owned by the company itself.

Based on this experience, it seems to me that three main phases can be distinguished in an ideal model of the data supply chain behind kNN and similar algorithms: (a) data collection, (b) algorithm configuration, (c) business rules configuration.

(a) Data collection. First, the media company has to decide *which* and *how many* films have to be described. This decision obviously depends on the portfolio of streaming rights the company has, or has acquired, yet the final list may well include 'external' items. Indeed, films regarded as 'classic' or 'relevant' may also be tagged, to serve as points of reference: the pool of potential titles, in this case, includes films whose rights may be acquired in the future, and even films that are never going to enter the portfolio: on Netflix, for instance, external items are tagged in order to recommend 'similar' titles in the available catalogue in response to the users who search for them.

Having reached this point, the media company must decide *who* is to collect data about the selected films. Here, executives face a classic 'make or buy' alternative. Our hypothetical company may opt to buy metadata from an external supplier (such as Gracenote, for example), or to collect the metadata by itself, internally. Buying metadata from an external vendor is probably cheaper, but there might be limitations in the databases available for sale, both in term of granularity and extension. Our company might thus decide to circumvent the problem by collecting metadata internally. Such course of action, however, requires money, and, crucially, competences: it requires *human* experts, people able to analyse audiovisual content, and other people able to coordinate, standardize and clean the process of gathering data. Also, it requires time: the company must decide which films ought to be given priority.

Neither option, i.e. neither the proprietary collection of data nor the use of external pre-existing databases available for sale, can be considered objective, or unproblematic. And here I do not refer solely to the inevitable degree of arbitrariness involved in the tagging process — manifest, for example, in scalar variables: is the level of gore in *The Wild Bunch* (Sam Peckinpah, 1969) a 4 or a 5 on a scale of 5? The same as *Cannibal Holocaust* (Ruggero Deodato, 1980)? Even if we discount for that inbuilt arbitrariness, in fact, companies must nonetheless face — above all else — the problem of *what data* they gather, and *for what*. The owner of a VOD platform has to decide *how* to describe its films, and *which* factors can be meaningful for its goals among the measurable data that algorithms can read and compute. In other words, when a company collects certain data (and not others), it must first form some preliminary idea of *what to* *do with them*, or, at least, of *what features*, in the contents it needs to describe, are the most interesting and relevant, or *representative*.

On a basic level, then, the phrase 'data collection' is misleading. Data are neither truly 'given' nor 'collected': they are created. Data do not exist in nature, and the decision to record certain kinds of information rather than others is, in and of itself, the outcome of a generative operation, which involves interpretation and simplification — a *datification* of the available audiovisual material. The media company has to decide its own *metrics*, its tagging system. Put otherwise, it has to decide *what can be data*. It has to identify a certain numbers of variables, such as, in the case of films and just to name a few: directors, actors, production year and country, cinematography, geographical and temporal settings, genre, keywords, themes, plots, various degrees of narrative details, all of which will then be used to describe each film, and to assess potential patterns of similarity. Netflix must have considered whether the 'social acceptability' of the protagonist was relevant, as a variable, to the pursue of its goals. Having decided that it was, then, it must have created a definition, along with an entire typology of related possibilities. Similarly, a media company has to decide *what* a genre is, and what it is not, *how many* genres exist, and *how many* of them can be identified for a single film, and so on. Does 'Kung fu', for instance, count as a genre? Or is it a sub-genre, a subdivision of the wider 'Action' genre? Why not a sub-genre of the 'Martial Arts' film, then, or even of the 'Sport' film? Or, even, neither a genre nor a sub-genre, but a theme? Can a film belong at one time to the genres Kung fu, Action, Comedy, Martial Arts and Sport?

Needless to say, the resulting data architecture can be very articulated, with many different levels for each variable. Nor is there a single *correct* way to organize the descriptors. The outline of the final taxonomy will depend, among other things, on the characteristics of the catalogue, and on what the company hopes to achieve in relation, for example, to its target audience, or to any strategic 'vertical market' that may exist for some or all aspects of the data. Equally significant will be the assumptions and habits of thought, and indeed the culture and nationality of the individuals who conduct the tagging and provide the service, as well as particular market standards and 'currencies'.

(b) Algorithm configuration. The collected data are then transferred to the kNN algorithm, where they are managed using a specific interface. During the transfer, a process of extraction, transformation and loading (ETL) takes place, which is likely to result in a whole remapping of the variables used in the collection. Those, in fact, must now conform to a different logic and architecture: that of the algorithm, but also that of the content management system (CMS), which determines the layout and internal organization of the webpage on which the final recommendations will be published.

Moreover, and crucially, most of the times the algorithm only considers a *subset* of the variables included in the collection, which is selected and *configured* for use independently from the collection phase. *Someone* — not a machine — has to decide *how many* and *which* variables ought to define similarity between films, and *what their relative weight is*. A recipe of sorts has to be invented. How many actors should be considered in determining the recommendations? Two, ten? Should they all carry the same weight in the assessment, or should the main stars be given more consideration? Do 'ensemble films' — assuming, of course, that such a category has been created and previously defined — require a different approach as to the number and relative relevance of the actors considered? To take another example, does a film like *Cry-Baby* (John Waters, 1990) relate better to other films dealing with the juvenile delinquency in the 1950s, even though it parodies them, or to *Grease* (Randal Kleiser, 1978), which is closer in terms of release and resembles it in theme and genre but not in tone, or even to other parodies dealing with completely different themes? *Someone* has to decide whether a matching theme carries more weight than, say, a convergence of genre, year of production, or tone. Equally, a decision has to be made as to whether *different variables* can correlate: can a film *from* the 1950s be made to match a film *on* the 1950s?

Indeed, the configuration of the subset presents the 'human expert' with a plethora of such decisions. Is the presence of a certain actor *more or less* important than a correspondence in genre? *How much more or less* important is it? Is the year of production more or less important than the production country? Some of these decisions appear, from a conventional *cinephile* perspective, more striking than others. The presence of the same director, for example, may or may not be counted as a condition of similarity. On the current version of Infinity (summer 2017), for example, the director does not appear to have any impact in determining content-to-content similarity. If one accesses the full details page of *Full Metal Jacket* (1987), no other Stanley Kubrick film is suggested as 'similar', despite the fact that some of them are present in the catalogue. It would appear that the director, here, is not just weighted less than other variables: he is simply not considered at all — in other words, *someone* decided his role was just not that interesting, at least for their goals (and for their users). Such a decision might appear to make little sense, considering how the director is, at least to some extent, an invention of film marketing. Yet, it is a fact that any platform can pursue their legitimate editorial interests, above and beyond what we deem to be objective or even 'sensible' criteria of similarity among films. If that is the case, however, we should also acknowledge the editorial nature of the resulting recommendations: the patterns and correlations 'discovered' by the algorithm are partial and certainly not universally valid, despite any claim to the contrary.

Once selected and weighed, the chosen subset of variables is passed on to the algorithm. Again, some companies choose to develop their algorithms internally, while others acquire the code from external developers, a decision which in turn can affect who can access and control the algorithm itself once it is running. Moreover, the code itself can play a key role in determining the final recommendations. Among the factors that can influence the results we can find a 'normalization logic' that assigns different values to tags depending on their incidence. Irrespective of the weight assigned to the corresponding variable, tags with a lower incidence 'count' more, and those recurring more often 'count' less. Such a normalization logic is widely used in statistical analysis, with computer scientists arguing that it is indispensable for every 'state of the art' recommendation system. In some cases, however, its usefulness appears dubious, but we cannot explore this issue further here. What is most important to us is that such statistical logic — which, in all likelihood, heavily influences the results in many current recommender systems — can be seen to create a hierarchy even in the case of variables whose data, in the original tagging system, were 'flat', that is, not hierarchically arranged. Therefore, here too, a fairly arbitrary ordering principle is introduced, whereby the number of instances for a specific descriptor comes to be considered, roughly speaking, inversely proportional to its importance — which is a human assumption.

(c) Business rules configuration. One final crucial factor comes into play in determining which films appear in the content-to-content rows. The algorithm *rarely* works just by itself. In fact, it never does. 'Business rules' intervene to filter some titles or push up some others, or even to balance the results according to pre-established criteria, so that the platform will have a more diverse row of films, for example. 'Push' rules are used, among other things, to highlight new items in the catalogue. Conversely, 'filter' rules are used to exclude titles, often to protect younger users. On Infinity, at the time of writing, horror or erotic movies are never suggested as similar to a movie, unless the latter is itself listed as horror or erotic. But this rule can be more general, and ensure, for example, that PG (Parental Guidance Suggested) and R (Restricted) rated titles are never suggested as similar to any G (General Audiences) rated movie (a category, it should be noted, which includes films that are not necessarily meant solely for families and younger audiences), and that regardless of how close a match they may be in relation to other variables.

Moreover, filter rules can be used to exclude a portion of the catalogue from the content-to-content row (typically the oldest part, containing films made before the 1980s), or to hide films produced, say, more than ten years before the film they resemble — and that is because older films are usually considered less valuable. Such filters may well result in a catch-22: lesser-watched films (such as the older ones) are considered less valuable, and thus penalized in the suggestions rows, which makes them even-lesser-watched, and so on.

Filter rules can also be used to prevent types of films that are considered *radically* different from appearing alongside 'normal' ones. For example, rules can be put in place that allow documentary films to be listed *only if* the starting title is a documentary film too — this rule was, at some point, part of the Infinity algorithm configuration. In someone's opinion, fiction films could never be similar to documentary films: thematic similarity was not enough, apparently, to allow correlation.

Similarly, on this platform, and even today, animated films appear to correlate and be correlated only to other animated films. The outcomes of this rule are questionable: on the page for *Rango* (Gore Verbinski, 2011) the algorithm fails to recommend any non-animated westerns, or live action Johnny Depp movies. Conversely, on the page for *Batman: The Killing Joke* (Sam Liu, 2016), adapted from a graphic novel by Brian Bolland and Alan Moore and featuring a blood-splattered poster, the algorithm recommends as similar *The Ice Age* (Chris Wedge, 2002) and *Penguins of Madagascar* (Eric Darnell, Simon J. Smith, 2014). *Rango* is a PG-rated film, while *The Killing Joke* is R-rated, which also suggests the existence of a *hierarchy* between business rules: someone decided that the animation rule should be stronger than the parental rating rule. Some other filter rules really seem bizarre: the page for *The Aviator* (2004) only recommends biographical films, because Scorsese's film is considered one, and a rule apparently dictates that biographical films can only lead to other biographical films, as if they were a genre too radically different from all the others.

Business rules are everywhere in VOD platforms, limiting the discovery of films, and not just through the list of recommendations, but even through the search field: users cannot really search *what they want*, even if the film they want is *there*. Not if someone does not want them to find it, and has blacklisted the title according to some (arbitrary) criterion, such as the year of production. In those instances, the only way to reach the desired title is through the 'tag cloud', as in the case, on Infinity, of *The Firm* (Sydney Pollack, 1993), a fairly recent film that is nonetheless unreachable from the platform search engine, even using the actors' or the director's name.

As those examples clearly show, business rules are an extremely powerful tool. Some *human being* has to decide *which* rules to use, *what for*, and *how* to make them work: their context of use and their scope of action. Granted, there are technical constraints to how they can be configured and how they can operate. These constraints, however, *also* derive from some *human* beliefs about what kind of filtering and pushing makes sense or not.

Similarity

So far I have limited the scope of my discussion to the process leading to content-to-content recommendations, which, as I said, are based on patterns of similarity between different contents. Now, without delving into the philosophical origin of the concept of similarity as a whole, it should at least be noted that even the narrower notion of 'film similarity' possesses a long history of its own. Similarity has long been used to differentiate products and stabilize demand within the film industry. It is, in fact, one of the elements behind the rise of the star system, and the adoption of genres in the studio era. At its root, similarity among films can be considered as an integral element of the economics of cinema, and of film marketing in particular. It certainly was so during specific periods in the history of the medium: one needs only to think of the exploitation of stock plots and the systemic, occasionally unlikely combination of genres that marked the high concept movies of the 1980s and 1990s. Both of these strategies were meant to push the marketability of a film highlighting the similarities with other titles, the familiar and successful elements, while simultaneously reducing the risk of economic losses.

On VOD platforms, algorithms suggesting similar contents always entail a specific idea of film similarity, while also excluding *other* ideas. Recommender systems — that is to say, all the *people* involved in their designing and functioning — decide which data can adequately describe films, which data can be used to correlate 'similar' movies, and, conversely, which data can be used to separate 'dissimilar' movies. Content-based algorithms, in their purest form, imply that these criteria are *universally* true, as if similarity was independent from the spectator. In fact, there is no such thing as objective similarity. If I consider *Chinatown* (1974) as a Roman Polanski's film, I will want to see other films by the same director listed in the content-to-content row. Someone else may consider it a Jack Nicholson's film, or a Robert Towne's film, or a John Huston's film, or a New Hollywood film, or a (neo-?) noir film, or a Los Angeles film — each of these stances should affect the contents of the similarity rows, but they do not. And this problem can only be *partially* fixed using personal ratings and play data as filters, like in the 'Because You Watched' row on Netflix — however, this solution does not change the weights of the variables considered by the algorithm when assessing the similarity: it merely acts as a filter, or adds a new variable with a much heavier weight.

Moreover, recommender systems (and the people behind them) usually seem to believe that their criteria of similarity should be equally valid for *all the films* in the catalogue, regardless of their country or year of production and so on — genre or cast are weighed the same in a 1950s Hollywood movie and a contemporary Italian film. The director is weighed the same (or is not weighed at all) both in an art-house production and a blockbuster movie — which should probably not be the case: certain metadata should be more relevant for certain films than others. The notion of genre as well as the notion of director (and many others) assume different values in different eras and places: the corresponding data, though they may refer to the same variables, ought to reflect this changing relevance, and be given different values when specific combinations occur. Assuming that it makes sense at all to keep thinking in terms of data (an assumption which in itself may well be reductive: describing all such combinations in discrete terms may prove an impossible task), what this means is that certain data ought to be counted differently when they appear in certain combinations rather than others. To translate a wide catalogue into a homogeneous set of data, using the same variables to describe significantly different contents, can be misleading. If not all users give the same importance to the same variables, it is also the case that not all variables apply equally to all contents. From the point of view of its contents, a catalogue is not a homogeneous collection, and cannot therefore be described by the same parameters.

In this respect, another misleading impression conveyed by content-to-content rows relates to the density of the kNN map, which is, as I said above, the spatialized representation of the patterns of similarity within the catalogue. The output of a recommender system appears to imply that such density is generally homogeneous, and that the films populating the row always have the same degree of similarity among themselves. They do not: that map has different densities, which the linearity of the rows smooths over, effectively hiding the different degrees of similarity between the recommended contents. The recommender system does not tell you exactly how similar those contents are. A title may have more like content in the catalogue than another one, but both titles will display the same amount of similar films in the content-to-content row. The row always includes the same amount of items, regardless of how closely clustered the recommended films are on the map, which is to say without considering how similar they are in terms of the algorithm (see the different circles of fig. 1).

Besides, the relative density of the kNN map is not the only element hidden from the user. Most if not all the steps forming the supply chain of data, as I described them above, from collection to recommendation, cannot be accessed if not through the back-end of the platform. Indeed, analysing VOD platforms as black boxes can be extremely frustrating, as one tries to infer how recommendations work without knowing the weights in the algorithms, or the business rules, or which metadata are used in which part of the system. The metadata fuelling the algorithm can be different from those displayed on the page (indicating, for example, the genre of the selected film): they can belong to completely different data sets. VOD platforms look transparent but are very much *opaque*, if seen from the outside — this is exactly why their functioning seems impersonal, or automatic, and their objectivity indisputable.

Conclusions: The Conditions of Recommendability

I did not want to consider a single case study, nor to expose the flaws of a particular VOD platform in a specific moment of its history, as my point is much more general. I sought to discuss the process behind content-to-content VOD recommendation, and show that there is nothing neutral even in the most seemingly 'objective' form of film recommendation. There is no real scandal in this — there can be legitimate editorial reasons behind the criteria establishing similarity between films. What is largely groundless is the widespread anti-humanistic myth of automation and disintermediation, as well as that other, parallel myth, describing a supposedly new, machine-enabled democracy of choice. Recommender systems do not really promote discovery: on the contrary, the criteria regulating the patterns of similarity tend to *reduce* the complexity of a catalogue. Rather, it seems to me that those systems contribute, if anything, to what Cherchi Usai, talking about something else, defined as the (necessary and unavoidable, to be sure) *destruction* of cinema[14] — as they select and define criteria of relevance, and priorities, much in the same way as, for example, film historiography does. Ironically enough, the only way to make the most of the long tail of a VOD platform would be to offer *random* rather than similar recommendations. (Incidentally, algorithms are often designed to include an element of serendipity in their recommendations — which may well appear paradoxical and even contradictory, given that such serendipity is nonetheless subject to certain pre-established conditions).

There is much more than meets the eye, in the setup and operation of these systems: theory, subjectivity, unquestioned (scientistic) assumptions, judgements, values, habits. People who decide, define, describe, choose, interpret, think and believe. These systems are much more human and less automatic than what enthusiasts of computational methods claim.

It falls on us to reflect, therefore, on the reasons behind this rhetoric of transparency and disintermediation, the futurism of commentators, their tendency to glorify the 'digital sublime', and to 'advertise the future'.[15] At the same time, we must consider the reasons behind the widespread diffidence towards any manifestation of doubt: the expression of a post-theoretical, anti-humanistic attitude, marked by an unquestioning acceptance of positivist ideas, all too ready to extol 'hard' sciences as immune from partiality and impervious to any situated or oriented influence.

And yet, the perception of technological efficacy is, first and foremost, a product of discourse and culture. And as such, it can very well change. The very rhetoric of machine-generated recommendations may face a turn of tide in the near future. The notion of algorithms falling short (to put it brutally) is gaining some momentum in the culture. Spotify, always particularly proud of what it can achieve through the use of data, appeared to brag in 2016 that 50% of the content played on its platform came, instead, from 'human curated' playlists.[16] Equally, a job posting for a position as film and book editor at Apple, dated 2017, proudly notes that 'at the heart of iTunes is human curation'.

While opposing the rhetoric of automation, however, we also need to reflect more critically about which ideas of audiovisual contents form the basis of VOD platforms; which *conditions of recommendability*, as I call them, those platforms adopt and foster, and for what reason. We need to investigate where such conditions come from, and where they may be taking audiovisual consumption, production and culture. One may venture to speculate that the success of a certain film, at least in terms of its digital consumption, is (also) determined

[14] Paolo Cherchi Usai, *The Death of Cinema: History, Cultural Memory and the Digital Dark Age* (London: BFI, 2001).
[15] Vincent Mosco, *The Digital Sublime: Myth, Power, and Cyberspace* (Cambridge, MA: The MIT Press, 2004); Armand Mattelart, *Histoire de l'utopie planétaire. De la cité prophetique à la societé globale* (Paris: La Découverte, 2000), p. 362.
[16] See Reggie Ugwu, 'Inside the Playlist Factory', 13 July 2016, <https://www.buzzfeed.com/reggieugwu/the-unsung-heroes-of-the-music-streaming-boom> [accessed 25 July 2017].

by its potential for recommendation — its ‘discoverability’, or ‘streamability’. Success, in other words, could be linked to how closely a film matches *certain criteria*, according to which it is *deemed similar to others*, or *suitable for a certain audience*, i.e. for a cluster of users whose interests are, again, *deemed* to match specific aspects and contents.

Data are necessary for recommendations and correlations, but also, as is well known, for advertising and content intelligence, and, by the same token, they end up playing a role in orienting audiovisual production. From this perspective, there seems to be a clear incentive for focusing not only on the analysis of data, or on how algorithms can process them, but also on the criteria that inform their collection, criteria that establish the *possibility* of description, similarity, correlation and interpretation. It is perhaps on the sensitive operation of definition of those criteria of similarity and correlation, from what I called the conditions of recommendability, that the shape and characteristics of much future cinema will depend, and perhaps does already. This, too, encourages us to look at recommender systems as strategic apparatuses, both machinic and fatally human.