

Giulia Taurino; Marta Boni

Maps, Distant Reading and the Internet Movie Database: New Approaches for the Analysis of Large-Scale Datasets in Television Studies

2018-12-31

<https://doi.org/10.25969/mediarep/14752>

Veröffentlichungsversion / published version

Zeitschriftenartikel / journal article

Empfohlene Zitierung / Suggested Citation:

Taurino, Giulia; Boni, Marta: Maps, Distant Reading and the Internet Movie Database: New Approaches for the Analysis of Large-Scale Datasets in Television Studies. In: *VIEW Journal of European Television History and Culture*, Jg. 7 (2018-12-31), Nr. 14, S. 24–37. DOI: <https://doi.org/10.25969/mediarep/14752>.

Erstmalig hier erschienen / Initial publication here:

<https://doi.org/10.18146/2213-0969.2018.jethc151>

Nutzungsbedingungen:

Dieser Text wird unter einer Creative Commons - Namensnennung - Weitergabe unter gleichen Bedingungen 4.0 Lizenz zur Verfügung gestellt. Nähere Auskünfte zu dieser Lizenz finden Sie hier:

<http://creativecommons.org/licenses/by-sa/4.0>

Terms of use:

This document is made available under a creative commons - Attribution - Share Alike 4.0 License. For more information see:

<http://creativecommons.org/licenses/by-sa/4.0>

MAPS, DISTANT READING AND THE INTERNET MOVIE DATABASE

NEW APPROACHES FOR THE ANALYSIS OF LARGE-SCALE DATASETS IN TELEVISION STUDIES

Giulia Taurino

University of Bologna

Dipartimento delle Arti visive, performative e mediali

Via Barberia, 4

40123 Bologna

Italy

giulia.taurino2@unibo.it

Marta Boni

University of Montreal

Département d'histoire de l'art et d'études cinématographiques

3150, rue Jean-Brillant H3T 1N8, Montréal (Québec)

Canada

marta.boni@umontreal.ca

Abstract: The presence of large-scale data sets, made available thanks to information technology, fostered in the past few years a new scholarly interest for the use of computational methods to extract, visualize and observe data in the Humanities. Scholars from various disciplines work on new models of analysis to detect and understand major patterns in cultural production, circulation and reception, following the lead, among others, of Lev Manovich's *cultural analytics*. The aim is to use existing raw information in order to develop new questions and offer more answers about today's digital landscape. Starting from these premises, and witnessing the current digitisation of television production, distribution, and reception, in this paper we ask what digital approaches based on big data can bring to the study of television series and their movements in the global mediascape.

Keywords: Television Studies, Cultural Analytics, Digital Humanities, IMDb, Mapping

1 Introduction

To analyze television series means to engage with long-term narratives, complex stories, transforming forms and formats, and with a growing amount of titles. Due to recent changes in the mediascape brought by the introduction of streaming platforms, the analysis of television series is in need of new methods, in order to tackle the global circulation

of content across cultural industries, as well as its social outcomes in different local realities. Being large, evolutive narratives able to transform according to audiences' reactions and markets' needs, TV series are more similar to complex systems than to linear experiences. Moreover, their quantity in the mediasphere is constantly growing, resulting in the emergence of alternative paths of circulation, which differ from those established by the classical broadcast model¹ and ultimately make it harder for analysts to keep track of all of them. These characteristics of the contemporary televisual culture, considered as a global phenomenon, encouraged some researchers to embrace perspectives based on quantity, thus going beyond close reading, traditionally used in Film and Television Studies.

It is worth noting that the presence of large-scale data sets, made available thanks to information technology, has fostered in the past few years a new scholarly interest for the use of computational methods to extract, visualize and observe data. Scholars from various areas are working on new models of analysis to detect and understand major patterns in cultural production and circulation, following the lead, among others, of Lev Manovich's *cultural analytics*², within the broader interdisciplinary field of Digital Humanities³. Computational and digital tools are used to extract information from existing data, starting from the hypothesis that forms of visualization are able to bring new questions to light, and offer more complete answers about today's socio-cultural landscape. Starting from these premises, and witnessing the current digitisation of television production, distribution, and reception, we first asked how digital approaches based on big data can benefit the study of television series. Second, we asked which datasets and tools can be used for analysing more specifically the presence of television content on a global scale.

In the following paragraphs, we will outline the epistemological implications of applying distant reading, an approach originally proposed in Literary Studies⁴, to the study of television series. Furthermore, we will consider what type of knowledge we can obtain when using digital tools to visualise and display quantitative information, following the pioneering researches in the academic field of the Digital Humanities. Our hypothesis is that cartographic visualizations made with open source GIS tools might be able to unlock and expose some unseen patterns in television production, distribution and reception. Maps, we argue, are able to offer a perspective that focuses not only on space, but also on the relationships entailed within it.

By pointing at one of the largest databases of audiovisual content, the Internet Movie Database, we will review previous uses of data extracted from the IMDb platform found in academic research. Having considered other data visualization projects in the field of Television Studies, we will bring our attention on a specific case study, the project "Les séries télévisées à l'ère des cultures numériques: production, diffusion, réception et circulation en réseau", carried out by a research group at the University of Montreal under the supervision of Professor Marta Boni. The project focuses on two interconnected aspects of television series: their formal elements, as described in the database's categories, and their socio-cultural uses. We will show how a quantitative analysis of macroscopic trends and features, in complement to close reading, can facilitate the study of television series as complex objects, in a fruitful exchange between Digital Humanities and Geography.

2 State of the Field

In the context of peak television⁵, we find ourselves facing an unexpected amount of content which makes scholarly research on televisual narratives challenging in many ways, from data collection to actual analysis. Furthermore, all

1 See Ramon Lobato, *Netflix Nations : the geography of digital distribution*, New York University Press, 2019.

2 See Lev Manovich, "Cultural Analytics, Social Computing and Digital Humanities", in *The Datafied Society: Studying Culture Through Data*, edited by Mirko Tobias Schäfer and Karin Van Es, Amsterdam University Press, 2017, pp. 55-68.

3 A first theorisation of the field of *Digital Humanities* was given in 2012 in Anne Burdick, Johanna Drucker, Peter Lunenfeld, Todd Presner, Jeffrey Schnapp, *Digital_Humanities*, The MIT Press, 2012.

4 See Franco Moretti, *Graphs, Maps, Trees: Abstract Models for a Literary History*, Verso, 2005.

5 Michael Schneider, "There is Simply Too Much Television". *TV Insider*. 2015. Accessed December 28, 2017. - <http://www.tvinsider.com/article/33397/fx-boss-there-is-simply-too-much-television/>

this content comes together with information about formal and textual elements of television series, as well as “social data” concerning their uses. Despite the unprecedented amount of content produced, the risk is to focus on the same case studies that emerge more prominently in the industry due to their success and popularity. Starting from a similar remark on the very limited amount of books that are studied in academia, compared to those mass produced every year since the beginning of print, in 1999, the scholar Franco Moretti proposed an approach based on distant reading, “where ~distance~ is [...] a specific form of knowledge: fewer elements, hence a sharper sense of their overall interconnection. Shapes, relations, structures. Forms. Models.”⁶

Let's note that Moretti's distant reading and the necessity of substituting it to a close reading approach moves from a basic assumption: the whole is *not* equal to the sum of its parts. If it was, we would then imply that culture is simply an assemblage of individual texts and, in doing so, we would underestimate the importance of the connections between them. Following Franco Moretti's approach on the evolution of the novel, we became interested in developing a quantitative and morphogenetic approach for tracking tendencies and variations in TV programs, globally⁷. Our work notably considers contemporary television series as complex cultural objects, made of several episodes and seasons, and constantly inscribed in transnational dynamics. While movies can be conceived as discrete objects, the level of interconnectedness of television serial narratives, derived from mechanisms of deep intertextuality and practices of transmedia storytelling⁸, makes it difficult to isolate them in singular units. By definition, television series need to be studied as serial objects.

In this sense, textual analysis might be able to tackle phenomena of repetition and variation⁹ taking place across several episodes. However, when it comes to long-running shows, the impossibility of studying *all* of them at once through a close reading approach is evident. The challenges of seriality also lie in the necessity of studying a series' *formula*¹⁰ or *matrix*¹¹ or the kernel from which the identity of the product is industrially created and reproduced. To do this, it is essential to observe television series from a distance, going beyond the traditional notion of semiotic text.

Instead of the notion of “text”, we thus suggest to adopt the theoretical framework outlined by Guglielmo Pescatore and Veronica Innocenti, who consider television series as *narrative ecosystems*¹². In their perspective, television series are complex organisms seeking their survival across several seasons, in relation to internal or external, contextual influences, being caused by production and distribution strategies or by reception dynamics. A TV series is to be observed as the agglomeration and constant reconfiguration of all these elements. In this view, analysis should be carried out by focusing on the intertwining between industrial producers of meaning and grassroots instances, and by considering the interplay between audiovisual texts, practices and social institutions. While this can still be done by performing analysis of single episodes and single series, within the digital context of streaming platforms quantitative methods might be more effective.

While in other scholarly areas quantitative studies are deeply rooted, in Television Studies, approaches based on quantity have started to unfold only in more recent years. Television scholars such as Jeremy Butler¹³ in 2014 or J.P. Kelley¹⁴ in 2017, or, later on, Taylor Arnold (Statistics) and Lauren Tilton (Digital Humanities) within the project

6 Franco Moretti, *Graphs, Maps, Trees: Abstract Models for a Literary History*, Verso, 2005, *op. cit.* p. 1.

7 For a previous discussion of the project, see also Marta Boni, “Working with maps: systems, networks, series”, in *A History of Cinema Without Names II. A Research Project*, *FilmForum/2015*, edited by Diego Cavallotti, Federico Giordano, Leonardo Quaresima, Mimesis, 2017, pp. 69-76.

8 See Henry Jenkins, “Transmedia Storytelling 101”, March 22, 2007, Web. - Retrieved from: http://henryjenkins.org/2007/03/transmedia_storytelling_101.html.

9 Umberto Eco, “Tipologia della ripetizione”, in F. Casetti, *L'immagine al plurale. Serialità e ripetizione nel cinema e nella televisione*, Marsilio, 1984.

10 Cawelti J. G. ([1969] 2006) “The concept of formula in the study of popular literature”, *The Journal of Popular Culture*, vol. III, n. 3, pp. 381-90

11 Guillaume Soulez, “La double répétition. Structure et matrice des séries télévisées”, *MAP* n°3, 2011.

12 For a better description of the notion *narrative ecosystem*, see Pescatore & Innocenti 2012; 2017; Bioni & Innocenti 2013; Pescatore, Innocenti, Brembilla 2014.

13 Jeremy Butler, “Statistical Analysis of Television Style: What Can Numbers Tell Us About TV Editing?”, *Cinema Journal*, 54, 1, 2014, pp. 25-44.

14 John Paul Kelley, “Quantitative studies in television: analysing and visualising 80,000 frames in 20 seconds”, *CST Online*, 2017 - <http://cstonline.net/quantitative-studies-in-television-analysing-and-visualising-80000-frames-in-20-seconds-by-jp-kelly/>

“Distant Viewing TV”¹⁵, developed various tools to study formal elements of television shows - namely, shot length and rhythm - through algorithms and quantitative measurements. These works thrive in the wake of what we could call a “metric turn” emerged, since the 2000, in Film Studies¹⁶. Another examples of how to use digital tools to produce alternative forms of knowledge are found in the experimental field of “deformative criticism”, which looks at a series as an organic aesthetic superimposition of several shots.

These works consider television series as the sum of their stylistic elements. However, other methods can be adopted for understanding not only data, but also the metadata associated to television shows. A case we should mention is that of Abigail and Benjamin De Kosnik’s project *alpha 60*, which aims at mapping the distribution of TV shows’ torrents across the world¹⁷. Here, cartography is used to highlight logics of access and the impact of fandom in the informal distribution of a series, also putting forward the hypothesis that “piracy is the future of television”¹⁸. However, even by simply observing official distribution, interesting evidences might emerge. And even without going as far as the project *alpha 60* went, while looking for information about illegal streaming, we already find interesting data within public, open access databases.

Databases like the IMDb, for example, offer numerous possibilities for studying films and television series as parts of cultural industries and economic systems. A first attempt to use the IMDb as a reference database to approach cultural networks was made as part of the project *Culturegraphy*¹⁹, where Kim Albrecht, Marian Dörk and Boris Müller analyse the IMDb by connectedness and community, in order to “visualize[s] the exchange of cultural information over time. Treating cultural works as nodes and influences as directed edges, the visualization of these cultural networks can provide new insights into the rich interconnections of cultural development such as that seen in movie references.”²⁰ In the model proposed by Albrecht *et al.*, a network is conceived based on the level of interconnectedness between audiovisual products like movies.

This study initiated a more solid path towards the use of IMDb for understanding cultural networked dynamics in Film and Television Studies. Among other examples, a research presented by Livio Bioglio and Ruggero G. Pensa, titled “Is This Movie a Milestone? Identification of the Most Influential Movies in the History of Cinema”²¹, shows an alternative way to measure the impact of a film over an extended period of time without relying on box-office revenues. In particular, they worked on “a subset of the IMDb citation network consisting of around 65,000 international movies [...]”. As they explain, “for each movie we also collect its year of release, genres and countries of production, and we analyze such features for finding trends and patterns in the film industry.”²²

These projects contributed to establish a “metric” and “geographic” paradigm where we aim to locate our research, by proposing a work on IMDb and other datasets in integration with cartographic tools. Mapping sample corpora of television content can favour the analysis not only of production and distribution, but also of the frequency or density of narrative forms and practices. This is particularly useful for studying non-linear television, where a widespread tendency towards digitisation results in an increased production of original content. Online platforms also cause a proliferation of paratextual content, both official and fan-made. Indeed, when dealing with such a big volume of data and such heterogeneous objects, the main problem is defining a corpus. This can be done via two intertwined actions. First, we have to select the corpus, either manually or via digital tools. Second, we need to take the tabular form of databases and give it an infographic form, in order to display information in a more intuitive way and highlight links between objects that would otherwise appear as separate.

15 Taylor Arnold and Lauren Craig Tilton. “Distant Viewing : Analyzing Large Visual Corpora.”, Digital Scholarship in the Humanities, Vol. 0, No. 0, 2019 - <https://distantviewing.org/>.

16 See for example the software Cinemetrics: <http://www.cinemetrics.lv/index.php>.

17 <https://benjamin.dekosnik.com/futures/alpha60/>

18 http://convergenceculture.org/research/c3-piracy_future_television-full.pdf

19 Kim Albrecht; Marian Dörk; Boris Müller, “Culturegraphy”, in *Leonardo*, 50, 2, 2017, pp. 209-210.

20 *Ivi*, *op. cit.* p. 209.

21 Livio Bioglio and Ruggero G. Pensa, “Is This Movie a Milestone? Identification of the Most Influential Movies in the History of Cinema”, in *Complex Networks & Their Applications VI. COMPLEX NETWORKS 2017. Studies in Computational Intelligence*, edited by Chantal Cherifi, Hocine Cherifi, Márton Karsai, Mirco Musolesi, Springer, 689, 2018, pp 921-934.

22 *Ivi*, *op. cit.* 921.

3 From Data to Knowledge and Back

When we look at the manipulation of data in Media Studies, that of Television Studies appears as a “dense” field that can justify an approach based on big data, given the amount of content produced and its resilience in surviving across different media. As we mentioned, a macroscopic approach to the study of televisual culture(s) and the narratives they produce should rely on a preliminary step: establishing a consistent and coherent corpus. While, on the one hand, linguistic corpora in computational research can be defined and delimited, on the other hand, when entering the domain of cultural analysis, corpora tend to be indefinite, open and, by their own overall adaptive and constantly self-redefining nature, incomplete. A *cultural-type survey*²³ therefore poses fundamental problems related not only to the size of the corpus, but also to its density, complexity and level of interconnectedness. Hence, the first operation consists in selecting the corpus or dataset from which we want to extract knowledge.

Rather than being seen as given raw material, potentially able to generate immediate answers, data should be seen as *capta*²⁴: its selection depends on research settings, aims and scopes. When using a quantitative approach for interpreting large-scale datasets, instead of developing descriptions of single elements, even before data analysis, the focus should be on practices of data cleaning and filtering. Brachman and Anand²⁵ notably suggest that creating knowledge with databases is a process made up of data discovery, data cleaning and data analysis. Data analysis thus comes at the very end of the process, and only when combined with data discovery and cleaning it can lead to the acquisition of meaningful outputs and results. Stressing on this point, in the article *From Data Mining to Knowledge Discovery in Databases*, Fayyad, Piatetsky-Shapiro and Smyth identify five stages for “extracting useful information (knowledge)”²⁶. They list them as it follows: i) data selection; ii) data (pre-)processing; iii) data transformation; iv) data mining; v) data interpretation/evaluation. Additionally, they define *Knowledge Discovery in Databases* (KDD) as “the nontrivial process of identifying valid, novel, potentially useful, and ultimately understandable patterns in data.”²⁷.

This framework shows a directional flow from data to knowledge as one of the leading dynamics of producing knowledge. However, we should note that data discovery also requires the researchers to master their background knowledge in the field that sources the data²⁸. In this sense, for the analysis of television series, it is important to stress out that the analyst is asked to manage a multi-layered knowledge, by being aware of the *textual*, *contextual* and also *pragmatic* dimensions of the life of the complex narrative ecosystem that is a series. In this sense, computational methods, based on a constructivist approach, result in being very similar to what is envisaged in the domain of Cultural Studies, where any text or practice is examined as intrinsically related to its context of production and circulation.

This loop movement from data to knowledge and from knowledge to data found a more explicit theorization in Lev Manovich’s concept of *cultural analytics*, described as “the analysis of massive cultural data sets and flows using

23 Anna Maria Lorusso, *Cultural Semiotics: For a Cultural Perspective in Semiotics*, Palgrave Macmillan, 2014.

24 “*Capta* is “taken” actively while *data* is assumed to be a “given” able to be recorded and observed. From this distinction, a world of differences arises. Humanistic inquiry acknowledges the situated, partial, and constitutive character of knowledge production, the recognition that knowledge is constructed, *taken*, not simply given as a natural representation of pre-existing fact.” Johanna Drucker, “Humanities Approaches to Graphical Display”, *Digital Humanities Quarterly*, 5.1, 2011, par. 3 - <http://www.digitalhumanities.org/dhq/vol/5/1/000091/000091.html>

25 Ronald J. Brachman and Tej Anand, “The Process of Knowledge Discovery in Databases: A Human Centered Approach”, in *A KDDM*, AAAI/MIT Press, 1994, pp. 37-58.

26 Usama Fayyad, Gregory Piatetsky-Shapiro, and Padhraic Smyth, “From Data Mining to Knowledge Discovery in Databases”, in *Ai Magazine*, 17, 3, 1996, pp. 37-54, *op. cit.* p. 37.

27 *Ivi*, *op. cit.* p. 49.

28 In a constructivist approach, data are *capta*: “*Capta* is ‘taken’ actively while *data* is assumed to be a ‘given’ able to be recorded and observed. From this distinction, a world of differences arises. Humanistic inquiry acknowledges the situated, partial, and constitutive character of knowledge production, the recognition that knowledge is constructed, *taken*, not simply given as a natural representation of pre-existing fact”, Johanna Drucker, “Humanities Approaches to Graphical Display”, in *Digital Humanities Quarterly*, 5, 1, 2011 - <http://www.digitalhumanities.org/dhq/vol/5/1/000091/000091.html>

computational and visualization techniques”²⁹. Manovich paved the way for further discussion on innovative ways of pursuing scholarly research in the Digital Age. In his book “Bit by Bit: Social Research in the Digital Age”, Matthew J. Salganik³⁰ notably gives an overview on the best practices of “research design” for analysing social practices, by exploring the positive outcomes of running digital experiments, in addition to simple observation and more analog approaches. Digital or hybrid approaches, he argues, allows for mass collaboration between projects, while also minimizing the expenses of doing research thanks to free-to-access digital infrastructures, among other reasons.

When compared to time and cost-consuming field-work, digital methods indeed show great advantages. A similar perspective on using computational methods for social research is taken by Noortje Marres³¹, who advocates for a new “science of society” or else a digital sociology that introduces new methods for expanding social research using data-oriented approaches. Even more recently, Richard Rogers³² went beyond the purely theoretical debate, by addressing digital methods through the lens of their actual applications and by considering data sources such as Google or social media platforms like Twitter. Even though datasets from Google or Twitter are often messy and contains missing or noisy values, they can still be powerful means for querying and mapping the web in search for information about human societies and cultures.

One of the major challenges of dealing with data and digital tools is in fact finding a good database. When working with public databases, we have to acknowledge that data extraction, collection and visualization tell us sometimes more about the databases themselves than about the actual objects they refer to. Having a variety of dirty data, we need to filter them and eventually create a sub-corpus that is unbiased and can be used for analysis. Only by being aware of the biases and errors of our database, we are able to develop a reliable analysis that takes on the various, intertwining levels of the life of an audiovisual content, including the storing of its metadata. A preliminary, manual filtering of the data thus avoids the risks of working with data that might be “broken”³³ or compromised.

To give a more specific example, nowadays, information concerning television series is constantly collected, stored and sometimes made freely accessible in the form of lists and databases, such as the Internet Movie Database. Known as one of the oldest movie catalogs available on the World Wide Web since 1993, the IMDb is descriptive of a complex and multidimensional system that includes granular data constantly updated in a seemingly real-time database, along with multiple networks of relations established at different stages. Nevertheless, far from being complete, the IMDb should be integrated, or merged - to use a more technical term -, with other databases and, depending on the focus of the analysis - production, distribution, reception level -, joined with other data sources. Additionally, we need to filter it and work on smaller samples of data.

From recent academic publications, we note the proliferation of quantitative analysis applications on the IMDb, with methods ranging from data mining and statistics used to make predictions, up to sentiment analysis and social network analysis for understanding patterns of algorithmic recommendation. Various approaches have been developed, with a very technical aim and scope, either following a more “hard sciences” approach or stemming from a humanist perspective. However, most research projects tend to move away from the primarily humanistic intent we want to adopt and they are likely to underestimate the importance of merging IMDb data with other datasets or filtering them manually. While they do create knowledge from data, this knowledge is often not backed up by a broad understanding of contextual dynamics of media ecology, which leads to unreliable results.

29 Lev Manovich, “Cultural Analytics, Social Computing and Digital Humanities”, in *The Datafied Society: Studying Culture Through Data*, edited by Mirko Tobias Schäfer and Karin Van Es, Amsterdam University Press, 2017, pp. 55-68, *op. cit.* p. 55.

30 Matthew J. Salganik. *Bit by Bit: Social Research in the Digital Age*. Princeton University Press, 2017.

31 Noortje Marres, *Digital Sociology: The reinvention of social research*, Polity Press, 2017.

32 Richard Rogers, *Doing Digital Methods*, Sage, 2019.

33 Sarah Pink, Minna Ruckenstein, Robert Willim, and Melisa Duque. “Broken Data: Conceptualising Data in an Emerging World.” in *Big Data & Society* 2018.

It is clear that a solely statistical approach appears too deterministic to be applied to mutable and abstract objects such as narratives are, and it ultimately seems unable to grasp the complexity of the audiovisual industry as a whole cultural and economic system. Computational methods and visualization techniques might be more effective tools for understanding large-scale patterns and generating knowledge from data, if we look at the historical and geographical evolution of audiovisual narratives and the network of relationships they entail. Overall, what we want to underline is that this preliminary process of database selection requires scholars to be able to discern information that is well-grounded and to recognise incomplete or incorrect data within a knowledge-to-data process. Knowledge-to-data and data-to-knowledge therefore go in synergy, since no meaningful visualization is possible without a proper selection of the data.

Furthermore, as Anne Burdick, Johanna Drucker, Peter Lunenfeld, Todd Presner and Jeffrey Schnapp³⁴ stated, when they reported the current state-of-art of the Digital Humanities, the use of visualization tools is capital for overcoming the barriers between a qualitative and a quantitative approach. The notion of *knowledge design*³⁵ is particularly pertinent here: according to Jeffrey Schnapp, once knowledge is discovered and extracted, in order to become understandable, it has to undergo a structural design, which allows access to the *information architecture*³⁶ of the corpus or data set. Examples of how audiovisual data and information are stored, queried, processed and, ultimately, made available online can be found in many fields, and forms of infographic representations vary depending on each specific research. What we propose here is to use maps as forms of knowledge design.

4 Maps

According to Fernand Braudel³⁷, geography, intended as the analysis and repertoire of overlaying gestures over time, can be useful for understanding macroscopic, slow and long term movements in a certain civilisation. Braudel notably invites us to work on the *longue durée* and follow the slow, macrostructural phenomena that only geography can display. Instead of embracing a certain historical approach based on big, significant events and characters, the role of the historian should consist in seeking repetitions and regularities on a large scale through an unbiased perspective. Braudel's geography of the Mediterranean builds upon the notions of structure and conjuncture, of immobility and movement, of slowness and excess of velocity. The digital context encourages us to rediscover a similar model of *total history*, echoed in 1997³⁸ by Barbara Klinger, who introduced this concept in Film Studies. This "total approach" might help us in examining all the aspects of a series, from an aesthetic, economic, historical and cultural standpoint. More specifically, we ask: "can we uncover everything that has been said about a series?"³⁹

Such a perspective fruitfully dialogues with Arjun Appadurai's concept of *total trajectory*⁴⁰ of a medium, or at least it highlights a similar approach that can lead to new interpretations in terms of overlapping layers and dynamics of evolution or progressive deterioration happening over the years across several territories. While, as in the case of other narrative forms, it is possible to emphasize changes within a more vast history - that of serial productions, or

34 Anne Burdick, Johanna Drucker, Peter Lunenfeld, Todd Presner, Jeffrey Schnapp, *Digital Humanities*, The MIT Press, 2012.

35 See Jeffrey Schnapp "Knowledge Design", *Herrenhausen Lectures*, Volkswagen Foundation, 2014.

36 Peter Morville and Louis Rosenfeld, *Information Architecture for the World Wide Web, 3rd Edition. Designing Large-Scale Web Sites*, O'Reilly Media, 2008.

37 See Fernand Braudel, *On History*, The University of Chicago Press, 1982.

38 "Rather more in the spirit of Fernand Braudel's concept of *histoire totale*, we can acknowledge both the unattainability of such a history and the benefits of its pursuit. In a total history, the analyst studies complex interactive environments or levels of society involved in the production of a particular event, effecting a historical synthesis, an integrated picture of synchronic as well as diachronic change." - Barbara Klinger, "Film history terminable and interminable: recovering the past in reception studies", in *Screen*, 38, 2, 1997, pp. 107-128, *op. cit.* pp. 108-109.

39 See also Marta Boni, "Anonymity, Complexity and Distant Reading. Toward a Total History for Audiovisual Studies", in *A History of Cinema Without Names. A Research Project. FilmForum/2015, XXII International, Film Studies Conference. University of Udine*, edited by Diego Cavallotti, Federico Giordano, Leonardo Quaresima, Mimesis, 2016, pp. 129-135.

40 Arjun Appadurai, *The Social Life of Things: Commodities in Cultural Perspective*, Cambridge University, 1986.

better, *cultural series*⁴¹-, every single series appears to have a duration in time and space, punctuated with limited, microscopic serial occurrences, made up of different stages and mutations. Geography will be our framework for reasoning both in terms of macroscopic, global systems and microscopic, localized events.

In order to overcome the limits of a database's tabular form, we first suggest the use of open source mapping tools like Esri, developed by ArcGIS. Thanks to GIS tools, we can adopt cartography to track homogeneity and heterogeneity, repetitions and variations, in the evolution of television seriality, both from a diachronic and from a synchronic perspective. Maps are in fact types of graphs with a topological component represented by geographical nodes often connected to each other by links (e.g. routing schemes), which allows us to overcome the structure of online databases, making it closer to the concept of network. While the entire corpus of IMDb, or other audiovisual content databases, can be read and analysed as composed by a network of hyperlinks, at the same time it can be visualised more intuitively on a map.

"What maps reflect, above all, are the viewpoints and worldview of their users (and forms of rationality in particular), as well as the use of graphical mediation to tame the invisible."⁴² Hence, the project of studying television's technology and cultural forms through geography is a statement of relativity. Instead of believing in big data perhaps too optimistically, by using maps we rather choose to understand phenomena as locally situated and temporally bounded, and, in a way, *designed* by the researchers themselves. What is certain (and this represents an interesting difference from traditional methods) is the following: spatial visualizations help us see television series starting from patterns, frequency, density, all elements that tabular and linear methods for building and displaying information do not allow.

Our quantitative method is based on several, interconnected steps that take into account not only preliminary data cleaning processes, but also the final steps involved in making data accessible through effective visualization. After having cleaned data extracted using IMDb's public API, we need to transform it in a form able to show the localization and spatialization of content. On IMDb, information is originally presented and gathered in a tabular form, which limits the possible queries and hypothesis, and does not display the object's complexity in its entirety. If we had to rely on a tabular form, we would risk to find only what we are looking for, without being exposed to new, previously unseen results. This is why we need a more effective visualization, a map. And since a series is not only made up of its text, we should additionally consider transmedial ramifications spreading from it: web pages, games and other satellite platforms that thrive in a scenario where media convergence facilitates a constant mutation of production, distribution and reception practices. Once we have visualized the tabular form of our IMDb corpus on a map, in a following step, we then proceeded to the manual gathering, from several sources, of official paratextual material that surround a show, which we also added on a cartographic visualization.

Finally, we considered forms of social discourse expressed on online networks and arenas such as Twitter. We were notably able to visualize how (and where) television series, thanks to their *spreadability*⁴³, become less and less "discrete". This proves again that they cannot be described as a unique experience, offered by a specific medium. On the contrary, they appear as complex organisms that develop strategies to survive across several seasons, in relation to the surrounding context and viewers' uses. Social media are great sources of data and they can be used for identifying social discourses around television series, both as mirrors of the society and as separate-yet-connected spaces where communities expand their possibilities of conversation. As Abigail de Kosnik⁴⁴ underlines, the mass of content produced by *The X-Files*' fans significantly overcomes the official texts. In this sense, a series can be seen as a mass, made up of heterogeneous material. This mass goes through mutations and processes of constant re-elaboration resulting from the interactions with audiences and public uses. In the digital context, such a mass becomes even more visible, widespread and chaotic. In our point of

41 See André Gaudreault, *Film and Attraction. From Kinematography to Cinema*, University of Chicago Press, 2011.

42 Christian Jacob, "When Maps Reflect", in Jacques Lévy, *A Cartographic Turn*, EPFL-Routledge, 2015, *op. cit.* p. 102.

43 Henry Jenkins, *Spreadable Media: Creating Value and Meaning in a Networked Culture*, NYU Press, 2013.

44 Abigail de Kosnik, *Rogue Archives: Digital Cultural Memory and Media Fandom*, MIT Press, 2016.

5 First Field of Application: A Cartography of Structural Mutations

Using GPS coordinates linked to titles, each series is then transformed into a set of dots on a geographic information system thanks to the software ArcGIS (Esri). We can then see which countries TV series are produced in, or even which ones are the most common locations, or the link between the actual locations and the diegetic places they take place in. It is also possible to work on the intersection of different data, such as year and genre, or country of origin and genre. Through geographical maps or diagrams, such data tells us where to look at, so that we can focus on specific information, or else analyse the convergence and divergence of serial forms over time. Questions can be asked, for example, about the average of episodes per season in a certain country, or about different stages in the adoption of social networks as continuations of the serial universe.



9

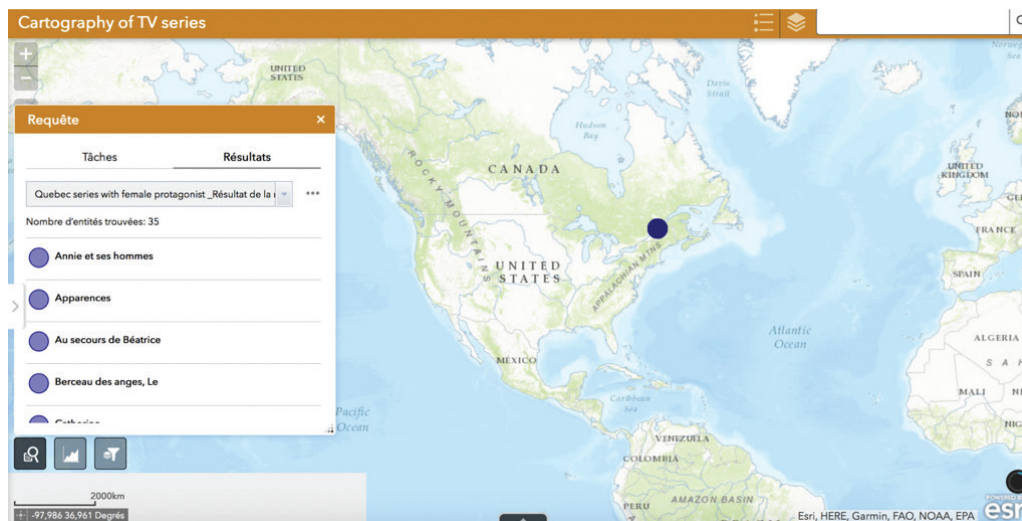


Figure 2. Map and list displaying manually-gathered information on French-Canadian series with a female protagonist.

This is an interesting way to show detailed, specific questions, or to display the evolution and mutation of digital paratexts for genres, countries and time-periods. By doing so, we can come to a better understanding of the “transmedia television” trend. However, we also need to consider some limitations of this approach. In fact, such a method turned out to be too time-consuming and not effective when applied on a much larger scale dataset.

Still, when working on a smaller scale and starting from a sample database, this set of analyses might be practical for linking serial texts to a certain place or moment in history. Studying television serial forms’ developpements over an extended time-frame and across different countries facilitates the appearance of salient traits and features, adaptation processes, as well as dead-ends or minor resolutions that didn’t survive history at large. Genetic links between similar forms, cultural and historical markers are seen here as causes of differentiation, linked to the transformation and slow mutation of forms over time. For example, taking TV series as our objects of study, we are able to outline the geographic distribution in specific moments in time, and therefore observe connections between locations and geopolitical issues, to ultimately reflect on cultural hegemonies or co-production networks over a certain period.

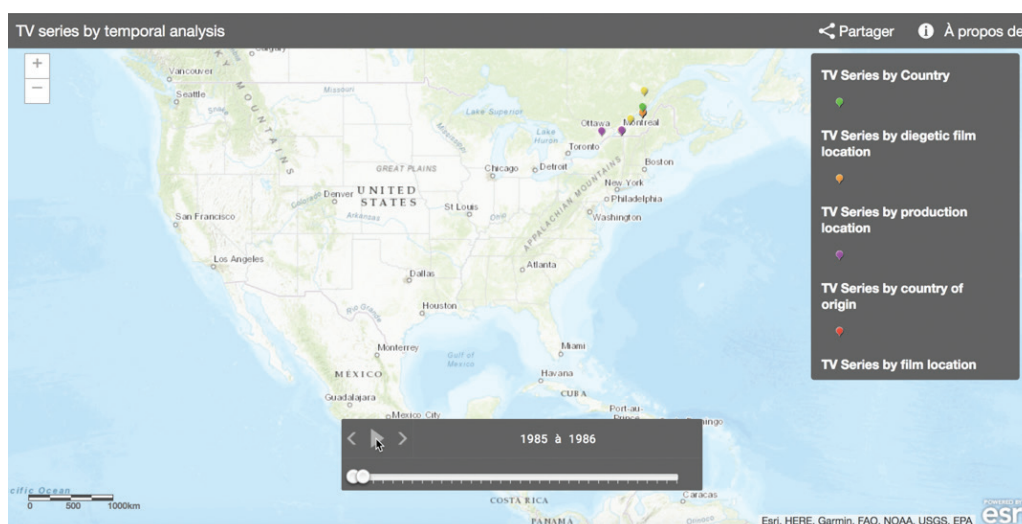


Figure 3. Esri platforms also allow animated maps, useful to display evolution over time: here, the growth of television series production from 1985 to 2017.

And yet, what such a temporal analysis cannot show is the actual circulation of specific series over time and its reruns. This is due to the fact that the IMDb remains, for the most part, a film-centered database, and even when it collects major information on television shows, it unfortunately appears very much biased towards a U.S. centric perspective. Given its controversial structure, the IMDb does not really allow investigation on television programs. We therefore looked for other data sources that can be found online in the form of big data, through which we can study not only series' locations, but also the metadiscourses associated to them across several countries. This strategy helped us overcoming some biases of the IMDb in favor of a more transnational view.

6 Second Field of Application: Tracking Social Discourses

A geographical approach is particularly pertinent if we think of a series in terms of spheres of discourse. By tapping into the Twitter's API through a script or through the official Twitter tool we can gather data on a specific hashtag related to a series (i.e. a series' title or a term which became common in a specific fandom), and track how it spread globally. We can then study the frequency of tweets through the open-source text analysis software Voyant Tools. Several tweets are associated to a geographical position (geolocation), which allows, again, to go back to our ArcGIS map and insert such points, as living traces of appropriation of a certain series in complement to more structural results previously generated. The Esri map also have a research tool that allows data from other sources than Twitter to appear (YouTube, Instagram, Flickr).

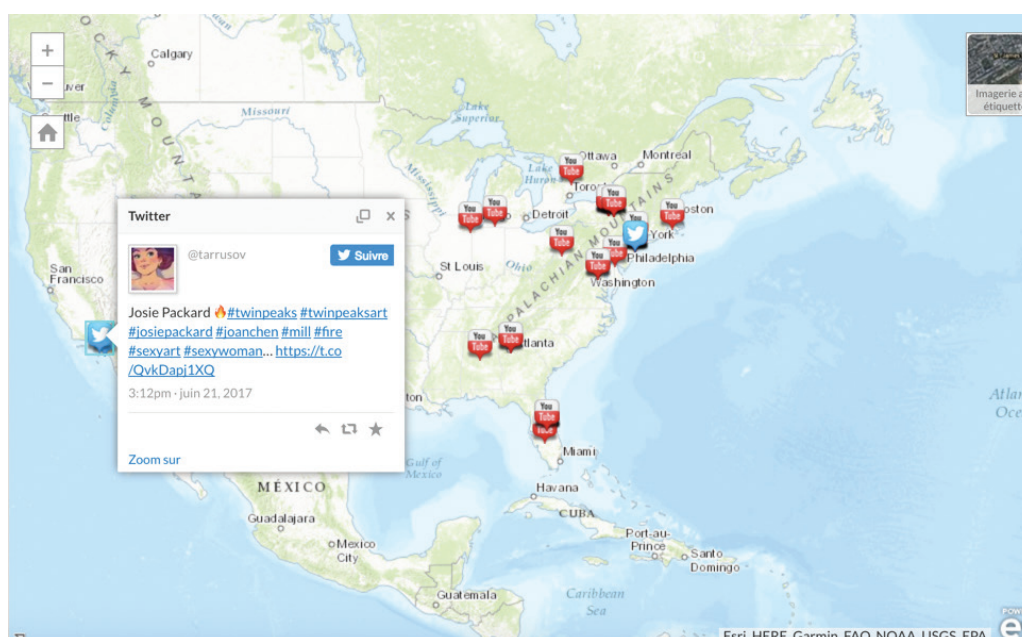


Figure 4. Mixed social network interactions for *Twin Peaks: The Return* via the Esri map (July 2017).

Furthermore, a real-time observation of the quickness and spread of tweets can be performed through software OneMillionTweets, which is able to visualize instant Twitter messages as shown in figure 5.

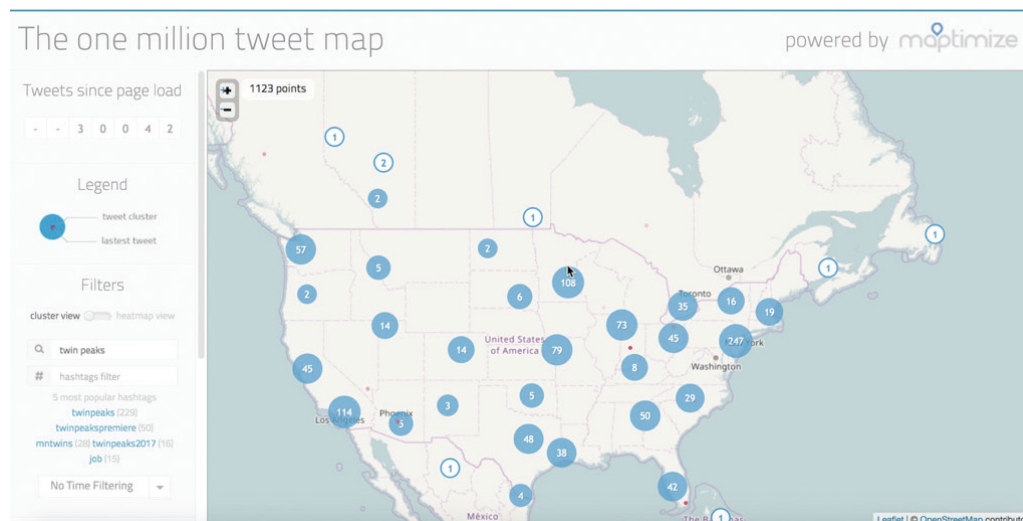


Figure 5. Tweets on the *Twin Peaks: The Return* Premiere (May 21st, 2017), via One Million Tweets.

Other social network analysis and visualization tools could be used, such as Netlytic, developed by the Social Media Lab group at Ryerson University. These data mining strategies allow us to quantify and to situate on a map or on a timeline phenomena such as the series' premiere or finale, thus visualizing and localizing the effects of a single media event. In other cases, the results can prove the existence of a parallel calendar, based on fan reception and productivity⁴⁵. A series' size, mass and intensity emerge. In this case, the map is complementary to an ethnographic approach on the study of users' discourse. Cultural semiotics gives us the idea of a sphere made up of multiple elements – and developing in the geographical space: we can think of a semiosphere⁴⁶. Thanks to its gaps, intended as intervals where links can appear, a series functions as a catalyzer of discourse and constitutes, over time and across space, its own identity: by inserts, anastomoses, selections, and crossovers.

By revealing the intensity of users' interactions within a specific moment of the season, maps show relevant nodes and display the most used terms in the metalanguage to talk about a series. In this sense, our analysis of Twitter focuses more on what a series can do in the actual space than on what production dynamics are. If we consider the analysis of forms and platforms through IMDb, it is interesting to note that we have two types of information: on the one hand, factual data (filming locations), on the other, descriptors (genres). Both rely on IMDb's language, structure, and organization. The former tells us about actual places and gives us information about production; the latter, informs us on the way the IMDb structures its information and organizes its categories. They are subjective, as in any description, even when it is the result of a collaborative effort. Every analysis of such data, to insist once again on this point, tells us not only about television series, but also about the database we are using: we are going in circuit from knowledge to data and from data to knowledge design, to actual knowledge.

7 Conclusions

The decision to adopt cartography as a mode of inquiry builds upon a body of theories and studies previously conducted in the Humanities. From Bruno Latour's⁴⁷ first extensive introduction for mapping networks in human

45 Marta Boni, "Des téléspectateurs géolocalisés : traces de la réception et construction de l'espace à l'ère du numérique", in *Sens Public. Ontologie du numérique*, edited by Servanne Monjour, Matteo Treleani and Marcello Vitali-Rosati, 2017 - <http://www.sens-public.org/article1283.html>.

46 Yuri M. Lotman, "On the semiosphere". in *Sign Systems Studies*, 33.1, (1984) 2005.

47 Bruno Latour. *Reassembling the Social: An Introduction to Actor-Network-Theory*, Oxford University Press, 2005.

societies through Actor-Network Theory to Jeremy Crampton's⁴⁸ critical approach to GIS, maps entered social sciences in various ways and soon showed an interesting set of valid applications not only in academic research for creating disciplinary and formal knowledge, but also for participatory, grassroots mapping, amateur practices and experimental art. Maps can therefore live after being born in the scholarly *milieu*, ultimately entering the world of activism and geopolitics. By generating a map, we are making a broader statement than simply analysing television content and social practices. Let us first take a step back and review our cartography project keeping in mind this scope of including cartographic research into the wider space of a social discourse, thus allowing for fruitful interactions between the scholarly field and other communities active outside of the academia.

The practice of mapping served us as a starting point for considering scattered information on single televisual objects in the larger context of a global, interconnected space. For instance, on one hand, maps (forms of cartography, but also conceptual diagrams) helped us outlining the interdependency between forms and platforms and their evolution through time in relation to media mutations. On the other hand, maps were used to collect, visualize and understand ephemeral and transient information, often geolocalized, such as viewers' uses and meta-discourses arising from each television series. In this sense, maps are representative tools that also function as performative search engines, favoring spaces of production of networks and connections, but also showing patterns of emergence.

Following Jacques Monod, Edgar Morin defines emergences as "qualities and properties that appear once the organization of a living system is constituted, qualities that evidently do not exist when they are presented in isolation"⁴⁹. In fact, the rhythm, paths and trends in television shows often evolve from a pre-determined production project into a more *adaptive* creative result, which grants the survival of each series over time, as it forms, develops and travels across the globe. In our perspective, television shows were considered complex not only with regards to their narrative structure, but most importantly because of the links between heterogeneous, non-narrative elements that constitute them and connect them with a larger system.⁵⁰ In particular, we focused on transmedial strategies of relocating and repurposing content through multiple platforms and technologies. We showed that, altogether, they create a continuity among different components of serial narratives, turning them into complex systems. Uses, along with formal and industrial strategies, are necessary elements of the series. Television series spread with unprecedented momentum in the current panorama, and each series leaves traces of discourse that become a mass in the mediasphere. In our applications, we demonstrated how we can visualize and make sense of the interaction of all such elements, including narrative and non-narrative snippets of content that circulate with, across and within the replicable form of a serial object.

It should also be stressed that our data are textual, rather than audiovisual. In this sense, the analysis suggested here is more similar to a distant reading of metadiscourse concerning series than to a *distant watching* of them. Other projects, closer to art history than to literature, and more similar to what Lev Manovich develops for images in social network, including color-detection or filter-selection, could be developed. Up to this stage, our analysis is not only structural and formal, but it additionally aims at acquiring a broader understanding of the actual spatialization of TV series nowadays. This is a useful way for studying the interaction between television forms and platforms, as well as the organization of information regarding television content at large. And beyond this, it can also give insight about the use of tools coming from the field of Digital Humanities for Film and Television Studies. What we are ultimately aiming for is creating an open access map that can be navigated by scholars, television audiences and other types of users.

The final outcome of this project is an interactive, open access atlas, which allows users to browse for specific series or countries and places, in order to discover links and new clusters of content. Different platforms such as Esri's Story Map or KnightLab's TimelineJS can also help us organise in space or time our database, each one providing different yet complementary ways to explore a big set of data. By visual means, we are therefore building

48 Jeremy W. Crampton, *Mapping: A Critical Introduction to Cartography and GIS*. Wiley-Blackwell Publishers, 2010.

49 Edgar Morin, *Restricted Complexity, General Complexity*, 2006, *op. cit.* p. 5 - <http://cogprints.org/5217/1/Morin.pdf>.

50 Marta Boni, "Mondes sériels et complexité", in *Ecrans. L'analyse des séries télé*, edited by Jean-Pierre Esquenazi, 2, 4, 2015, pp. 55-68 - <http://dx.doi.org/10.15122/isbn.978-2-8124-6095-1.p.0055>.

a non-linear narrative, one that users can explore at their pace, following individual curiosity, instead of looking for a unified teleological direction. By browsing titles and clicking on specific points on the maps, users are able to see how, where and when a series colonizes the geopolitical and social space. Overall, the atlas gives us the image of a transnational proliferation, thus creating a new space, its own space. A map can be therefore relevant to explore the life of a series, or, better yet, to understand television series as dynamic life forms. The whole landscape of serial narratives contributes in creating a map, where practices transform spaces and spaces transforms practices. Digital tools ultimately prove the groundedness of a series in the social space, by bringing to the foreground something which otherwise may be held in the background, but also by generating an additional space where further social discourse can happen.

Biographies

Giulia Taurino is a Ph.D. candidate in Media Studies at the University of Bologna and at the University of Montreal. Her research interests mainly focus on narrative and cultural forms in contemporary media ecology. Building on neo-formalism, her research project notably addresses the role of the anthology form in internet-distributed television, while trying to assess its affordances and uses in the broader context of digital economy. She is currently collaborating with the research group Labo Télé (University of Montreal), where she works on a project for implementing data analysis and visualization in Media Studies.

Marta Boni is associate professor of Film and Television Studies at the University of Montreal. She has published *Romanzo Criminale. Transmedia and Beyond* (Ca' Foscari University Press, 2013), co-edited *Networking Images. Approches interdisciplinaires des images en réseau* (Presses de la Sorbonne Nouvelle, 2013), edited *World Building. Transmedia, Fans, Industries* (Amsterdam University Press, 2017), various essays in international scientific journals such as *Mise au point*, *MEI*, *SERIES*, *Cinergie*, and chapters in edited collections on television seriality, fandom, media paratexts, and transmedia.