Swiss Journal of Sociology, 51 (2), 2025, 317–336

Seeing Like a Field?

Katrin Herms* and Jörg Lehmann**

Abstract: This article introduces the notion of *seeing like a field* – the view from the inside of a Bourdieusian field in which agents and their social positions are located. Potentials and fallacies of concretizing field theory with social network analysis (SNA) are evaluated on the basis of two examples of current research approaches to big visual data: computer vision networks and image similarity analysis.

Keywords: Field analysis, computer vision networks, image similarity analysis, machine learning, predictive analytics

Voir comme un champ?

Résumé: Cet article introduit la notion de *voir comme un champ* – la vue de l'intérieur d'un champ dans lequel les agents et leurs positions sociales sont situés. Le potentiel et les pièges d'une concrétisation de la théorie des champs par l'analyse des réseaux sociaux (ou *Social Network Analysis* – SNA en anglais) sont discutés au travers de deux exemples d'approches de recherche actuelles par l'analyse du *big visual data*: les réseaux de vision par ordinateur et l'analyse de similarité des images.

Mots-clés: Analyse de champ, réseaux de vision par ordinateur, analyse de similarité des images, apprentissage automatique, analyse prédictive

Sehen wie ein Feld?

Zusammenfassung: In diesem Artikel wird das Konzept *Sehen wie ein Feld* eingeführt – der Blick aus dem Inneren eines Bourdieu'schen Felds, in dem sich Akteurinnen und ihre sozialen Positionen befinden. Potenziale und Fallstricke einer Konkretisierung der Feldtheorie durch soziale Netzwerkanalyse (SNA) werden anhand von zwei Beispielen aktueller Forschungsansätze für große visuelle Datensätze diskutiert: Computer-Vision-Netzwerke und Bildähnlichkeitsanalysen.

Schlüsselwörter: Feldanalyse, Computer Vision Netzwerke, Bildähnlichkeitsanalyse, maschinelles Lernen, prädiktive Analytik

* Université de Lausanne, Institut des sciences sociales, CH-1015 Lausanne, katrin.herms@unil.ch.

** Staatsbibliothek zu Berlin Preußischer Kulturbesitz, D-10772 Berlin, joerg.lehmann@sbb.spk-berlin.de.

DOI 10.26034/cm.sjs.2025.6906

1 Introduction[1]

Internet communication has become omnipresent. Images, texts, and videos retrieved in social media offer insights into everyday situations like cooking or spending time with family, friends, and pets. Digital stories are characterised by increasing imagery, as "people are three times more likely to engage with tweets that include visual content" (Alton, 2024). Users build social ties with other platform users when following, liking, commenting, or sharing their images. Since more and more people are posting pictures from their daily lives online, characteristics of circulating images, their production and reception in a social media environment form new fields of research (Ma & Fan, 2022). The social network Instagram is a platform that combines images and videos with microblogging. With more than 2.4 billion users since its launch in 2010, out of which 500 million users access the platform each day and who created over 990 million daily photo sequences that were published in 2023 as "stories" (Demandsage Instagram Statistics, 2024), Instagram can be seen to hold *big data.* It is an archive of everyday digital practices. Based on the available data (i.e. visual content and trace data including socio-demographic variables), the service offers a source for linking observations of aesthetic preferences with an analysis of users' characteristics and properties. Under the condition of data access, analysis enables to group users into categories of taste, explore their networks, and make statements about their lifestyles. Such an analysis can be done with economic interest or with sociological ambitions. Whilst sociologists study social dynamics and consequences of media use, ideally by following ethical standards of data protection, for instance to avoid user profiling, big tech platforms primarily long for profit maximisation based on trace data. Instagram is but one example of how the availability of *big data* – and not only big visual data, but also data on the users – changes the view of the big tech corporations onto their users. For them, it is now possible to *See like a market*, to cite Fourcade and Healy (2017), which indicates the ability to calculate, stratify, and apprehend user behaviour based on consumer tracking and measurements that are applied on massive collections of trace data: "As new techniques allow for the matching and merging of data from different sources, the results crystallise – for the individuals classified – into what looks like a supercharged form of capital" (Fourcade & Healy, 2017, p. 10). To give an example, Meta as the parent company of Instagram is able to combine images with further data sources from Facebook and existing telephone numbers from WhatsApp as well as the respective real names. The result is a precise set of digital records, a *data double* (Bouk, 2017) of all users active

1 The contributors would like to thank the anonymous peer reviewers for their valuable feedback and their guidance and support, as well as the editor of this volume, Sebastian W. Hoggenmüller, for his patience, perseverance and all his effort in bringing this article to publication. The contributors would also like to cordially thank Janna Joceli Omena, Elena Pilipets, Beatrice Gobbo, and Jason Chason for their kindness to provide illustrations for being discussed in this article. The first three illustrations (Figure 1.1–1.3) below stem from their article in the journal *Diseña*, published under the CC-BY-SA license.

on these platforms. Historically, capitalist markets and bureaucratic organisations in the service of nation-states have always strived to collect data about consumers and citizens. These data serve to apply rules and measures, to implement classificatory schemes and analyse consumption patterns and scores in order to make the real world legible to companies and states. However, in the 21st Century analytic capacities supported by computing power go way further. Predictive analytics – the identification of preferential consumption patterns – enables the anticipation of new needs, desires, and trends in real time during their formation. As big tech companies are able to distinguish consumers along categories such as riskiness or worth, such new stratifying technologies also have the power to discriminate in a way that was yet unimaginable in the 20th Century (Fourcade & Healy, 2017, p. 24). At the same time, big (visual) data are available for scientific purposes. Social media data are used in computational social science projects to test, among others, machine learning applications and to optimise models that are supposed to predict social behaviour. A risk of this development is the creeping economisation of digital sociology if researchers uncritically adopt and reproduce economic categories of user stratification. With the increase in publicly accessible trace data, another challenge is to choose meaningful heuristics for the interpretation of big (visual) data.

A classic epistemological framework for the analysis of everyday social practices was developed by the French sociologist Pierre Bourdieu. He was the first to ask how everyday situations are regulated without people consciously following predetermined rules. Inspired by literary descriptions and based on his empirical observations of French cultural idiosyncrasies in living, eating, dressing, and receiving art, Bourdieu unfolds his analysis in *la Distinction* (Bourdieu, 1979). Several researchers argued later on that social network analysis (SNA) would enable a concrete method of application for Bourdieu's field theory. While Bourdieu was amongst the first sociologists to employ correspondence analysis (Bourdieu, 1979, pp. 296, 596, 622), Wouter de Nooy (2003) proposed to transfer the contingency tables on which Bourdieusian correspondence analysis rests into adjacency matrices used for social network analysis. A more recent proposal is given by Stefan Bernhard (2008) who argues that the theoretical strengths of field theory should be combined with the empirical possibilities of network analysis in order to overcome the weaknesses of both approaches. In light of the growing possibilities for sociometry based on trace data, more and more quantitative and structural approaches have been developed to combine field theory and SNA. For example, Serino et al. (2017) suggested a practical implementation to combine blockmodeling and multidimensional data analysis in a case study on theatre industry as a field of cultural production. Nevertheless, concrete suggestions to use field theory for analysing the phenomenon of pictorial digital forms of expression with big visual data have not yet been developed. Therefore, and in allusion to Fourcade and Healy's critical study *Seeing like a market* (2017), we ask: Is it possible for sociology resting on big visual data to *See like a field*?

Our contribution discusses this question, first by introducing Pierre Bourdieu's central ideas on field theory. We then present two existing big visual data projects and discuss how technical approaches like computer vision networks and image similarity analysis can be used for an empirical concretisation of the field theory. With the concept of attention economy, we suggest a methodological extension. By providing technical insights into computer vision networks and image similarity clustering, the article explores some possibilities available for sociological analysis to extract, transform, and load such data. Which research results can realistically be achieved with it?

2 Excursion in Social Theory

2.1 Field, Struggles, Capital, Class – A French Social Theory of the 1970s

For Bourdieu, *fields* are practical and dynamic arenas of social struggles in which legitimate patterns of thought, perception, and action are fought for (Bourdieu, 1979, pp. 249–291). In that sense, society is seen as an ensemble of fields in which status and positions are negotiated, even if the agents involved are not aware of it. These fields can be imagined as spaces in which the individual agents occupy different positions and in which nevertheless structures prevail. With the concept of the field, Bourdieu aims to overcome the opposition of objectivism and subjectivism, i. e. he on the one hand wants to highlight objective structures, such as power relations, the distribution of different types of capital and labour or group formations, as well as developments that characterise the field as a whole, such as the emergence of a new fashion. On the other hand, he wants to throw the individual agents into perspective, their subjective view of the field and their individual contribution to its development. Every agent inevitably orients herself to the structures, purposes, and objects already present in these fields, since these spaces are characterised by an ensemble of objective power relations or structures that impose themselves as constraints on all those entering the field. At the same time, there is a permanent fight for the most influential social positions: "Les luttes des classements, individuelles ou collectives, qui visent à transformer les catégories d'aperception et d'appréciation du monde social et, par là, le monde social, sont une dimension oubliée de la lutte des classes" (Bourdieu, 1979, p. 564).

In addition, the different fields are characterised by four categories of capital. Economic capital follows the Marxian understanding. Social capital is formed by the networks, connections, and relationships that each agent has with other agents, e. g. relationships of knowledge and recognition or group affiliations. The third type of capital is cultural capital, which comes in three forms: Incorporated cultural capital (manners, etiquette, the ability to master certain narrative forms etc.), objectified cultural capital (paintings, books, instruments etc.), and institutionalised cultural capital (doctorates, professorships, honorary titles etc.). The last type of capital is symbolic capital, which can be described as the credit of social recognition granted by other agents; it is also referred to as social prestige, kudos, honour or reputation.

The positions of the agents in the various fields are characterised by the different combinations of these different types of capital, and, as groups of actors are formed, dominators and dominated emerge. In Bourdieu's conception, the terms *field* and *struggle* are closely related, as groups of actors struggle for the monopoly on the assertion of legitimate categories of perception and evaluation. The sequence of these struggles is what forms the history of a field, since the progress of these struggles characterises its evolution: "C'est la lutte même qui fait l'histoire du champ; c'est par la lutte qu'il se temporalise" (Bourdieu, 1992, p. 223). Analysis reveals this succession as a series of oppositions, i.e. the contrasts "entre les plus riches et les moins riches en capital spécifique, entre les dominants et les dominés, les tenants et les prétendants, les anciens et les nouveaux entrants, la distinction et la prétention, l'orthodoxie et l'hérésie, l'arrière-garde et l'avant-garde, l'ordre et le mouvement" (Bourdieu, 1979, p. 257). As can be read from this citation, the terms used for groups like the *established* and their challengers may vary and are exchangeable. Interestingly, Bourdieu gained this understanding during the genesis phase of his theory through an examination of Weber's concept of the religious field (Bourdieu, 1971) and transferred it from there to other fields of symbolic production such as art, literature, or fashion. In cultural markets, heretics (or avant-gardists) clash with the orthodoxy, challenging those who have arrived and form the dominant pole, questioning their creations by contrasting them with new works. Either the rules that characterise the field, *Les règles de l'art* (Bourdieu, 1992), are broken by the heretics and new ones are established; or the heretics are dismissed, and the orthodoxy remains in power. Identifying and describing this rotation process is tantamount to narrating the history of the field.

According to Bourdieu, class affiliation – the social position – results from the interplay of all the resources that are acquired in the course of life and for which socialisation in the family circle is the starting point. Bourdieu thus developed a social theory that expanded Marx's concept of class to cultural aspects. The concept of capital has the advantage of encompassing subjective forms such as mutual recognition as well as objectified and institutionalised forms. By analysing the four types of capital, the networks of relationships and the agents' resources for action as well as the gravity of the structures within fields can be captured. For Bourdieu, there is no ultimate determination by the economic, but fields in which distinction prevail. In this way, both can be analysed: The subjective view onto the field and the objective structure within the field.

While the content of the last paragraph already indicates the transition from *Seeing like a market* to *Seeing like a field*, some differences between the two approaches must be pointed out. Bourdieu knew Marx well, and it can be said that the term *field* is in principle a reformulation of the term *market* (Bourdieu, 1992, pp. 121–126). On the other hand, Bourdieu distanced himself from Marx to the extent that he literally adopted Marx's concept of economic capital, but at the same time considered it to be deficient and therefore differentiated it into the four types of capital introduced above and insisted on the inherent logic of the symbolic. This in turn leads to new difficulties: While Marx operated only with an economic concept of capital, he was able to explain the regulatory mechanism between supply and demand as well as the fluctuation of market prices. Bourdieu's differentiation of four types of capital, on the other hand, creates the difficulty that symbolic capital, for example, is hard to quantify, and the market mechanism can no longer be explained. Therefore, and in the sense of Adam Smith's *unseen hand*, Bourdieu speaks of a miraculous "harmonie observée entre les producteurs et les consommateurs de produits culturels" (Bourdieu, 1992, p. 231). This has to be regarded as a theoretical blank, as Bourdieu was not able to appropriately explain the regulatory mechanism between supply and demand in cultural fields.

2.2 Scope of Social Network Analysis Applied to the Examination of Fields

In the 21st Century, the Internet has become the paradigm of a networked global society. With the emergence of platforms, the availability of data via application programming interfaces (APIs) and massive computational power and the expansion of analytic possibilities, SNA has developed into an effective methodological approach for the analysis of Bourdieusian fields alternative to the correspondence analysis employed by the sociologist himself. Proposals to concretise field theory with SNA approaches have therefore been discussed for around 20 years, in which quantitative approaches are dominant. The big promise to concretize field theory with computer science is that SNA offers a framework for the precise calculation of objective structures and individual agents' positions based on big data. Socio-demographic variables can be added to network data as attributes to calculate e.g. homophily (Centola, 2011). The network approach enables to shift the perspective from a general pattern to a specific context, thus providing a continuum of scale that ranges from a macro to a micro view (Ahnert et al., 2020). When Bourdieu studied statistical correlations between, e.g., the level of education and cultural taste, only a fraction of today's amount of data were available. Though Bourdieu used the computer for analysis from the 1960s onwards (Escofier, 2024, p. 268), today's multiplied computational power can calculate huge amounts of potential correlations between internet users' properties and predict future behaviour based on routines of the past. The availability of big data has even provided for methodological advances: *Quantifying reputation* has become possible and provides new insights into the distribution of symbolic capital and the valuation of art (Fraiberger et al., 2018).

However, individuals can change their lifestyles, opinions on politics, and usage of social media. The same accounts for music taste on streaming platforms like Spotify or Deezer: trace data do not cover individual future-related decisions if they contradict habits of the past. Another risk lies in the confusion of relational structures with society as a whole, as (1) society also provides values that give orientation for action (Trezzini, 1998), and (2) when Pierre Bourdieu spoke of *relations*, he rather meant distinctive ties than affiliations (Hennig & Kohl, 2012, p. 16). Furthermore, network structures, even if built on massive volumes of data, can only represent selectively chosen facets of relational ties between agents. During research, decisions about the variables to be compared have to be taken out of a much more complex ensemble of multilayer reality. Users also tend to appropriate the platform's functions and can rely on one and the same technical functionality for different reasons (Cha et al., 2010). At worst, SNA approaches could lead to structural determinism, meaning that individual preferences and discourse are only seen as a by-product of some network position, without the room for creativity that Bourdieu highlighted in his theory of practice (Janning, 1991, p. 30f.). Finally, digital networks do not map onto real-world networks on a one-to-one basis. While network data allow for the analysis of social and symbolic capital in a way that is not attainable via observation in the real world, such data are bereft of important dimensions of social interactions, such as the presentation of objectified and institutionalised cultural capital as well as the expression of body language that was so dear to Bourdieu (facial expression, gesture, or *hexis* and *kathexis* in his terms). Rather, the digital interactions facilitated by platforms have to be imagined as negotiations of global communities that conduct "la lutte pour le monopole de l'imposition des catégories de perception et d'appréciation légitimes" (Bourdieu, 1992, p. 223) deprived of face-to-face exchanges.

2.3 A Concrete Proposal for Big Visual Data Based on Attention Economy

The concept *attention economy* was coined by the Austrian economist Georg Franck (Franck, 1998). It assumes that various actors and institutions such as politicians, companies, or activists are involved in a societal competition for visibility and feedback. In digital environments, not only tech companies like Google and Meta have understood that it is worth investing in the rare resource of attention for selling products; also, average platform users participate in the struggle for attention and visibility (Pörksen & Detel, 2012). Inspired by this idea of implicit market rules that structure social interaction in social media, we suggest a new perspective on how to conceptualise field theory based on SNA. We therefore define the respective platform – such as X (before Twitter), Instagram, TikTok, or BeReal – as a social field. Here, different actors compete for visibility and develop forms of cultural expression that are both individually curated and standardised by platform design (Reckwitz, 2020, pp. 225–230). In continuity with Ferdinand de Saussure's linguistic structuralism to which Bourdieu also referred, the surface features of digital platforms provide the general grammar of what can be said (= *la langue*). Each user fills this space with individual re-combinations of cultural content (= *la parole*). If we assume attention economy as the implicit rule of interaction in a digital setting, we can look at concrete statements made by social media users in relation to the structural position occupied by agents in their social network. The central object of investigation would then be digital communication. The micro and macro level of social practice correspond to the cultural style with which individuals present themselves and how they compete for attention. Feedback would have the function of a resource. Those who generate a lot of clicks, replies and comments or followers – for example through particularly provocative statements, an appealing profile, or a linked visual – position themselves as central nodes in the overall network where social capital is quantifiable; less successful profiles are less visible and less well linked (Barabási, 2009, pp. 65–68). In this sense, the digital counterpart to cultural classes can be compared to collective evaluation practices that emerge when circulating images are commented on similarly by, say, many Instagram users. Transferred to ego-centred networks, a typology of cultural classes appears to be interesting with regard to so-called influencer profiles, which gather a community of followers around them with their personal style. Since images increase the impact of posts, as mentioned at the beginning, they are of specific interest for research.

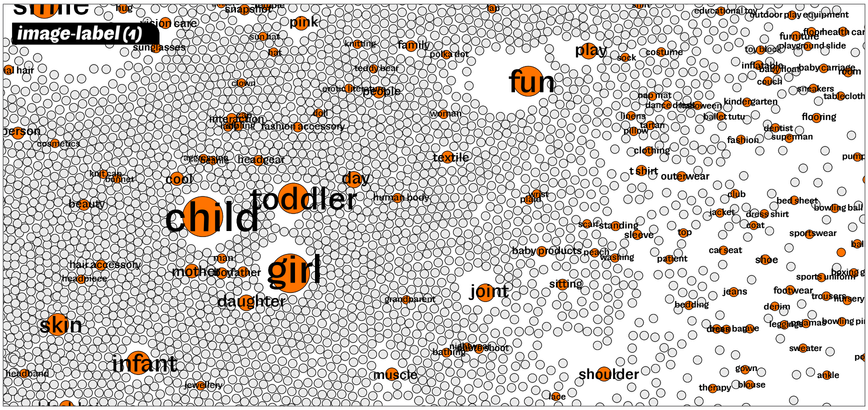
3 Two Examples of Big Visual Data Analyses

We now illustrate the applicability of a digital field interpretation in the sense of attention economy with two empirical projects: (1) computer vision networks, whereby positions in the network are derived from the frequency of image labels, image-entity links, and image sources (here websites), (2) image similarity clustering, where images are placed in a high-dimensional space.

3.1 Computer Vision Networks From Big Visual Data

The Google Image API is openly available and can be accessed by researchers to collect image data all over the web. The first chosen case study illustrates its usability by presenting a research project developed by Janna Joceli Omena and her colleagues (2021), who assembled online available image collections for a transformation into computer vision networks. The Image API delivers three kinds of metadata: (a) descriptive labels added to the images themselves, usually in the form of alt-text available in the html code; (b) web entities (e. g., a person, a place, or a thing) or textual descriptions added to images in the text surrounding the images,

Figure 1.1 Example of an Image-Label Network. The Size of Nodes Indicates How Often an Image Label Is Assigned (the Larger the More Frequent)

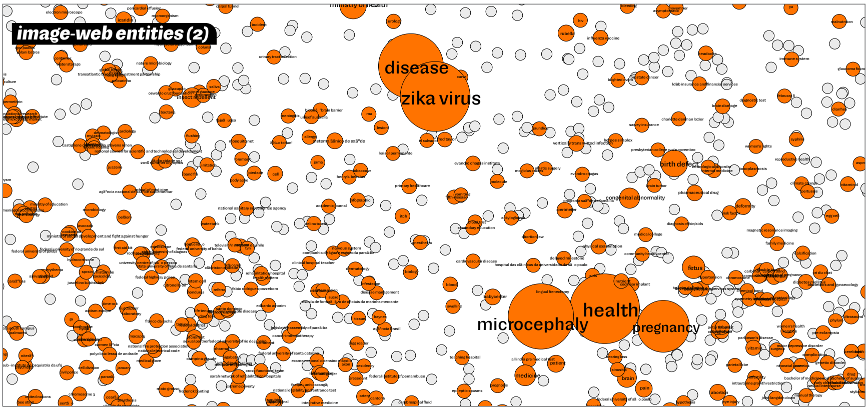


Source: Omena et al., 2021, p. 9.

a feature which is available only via Google's Image API; and (c) the respective sites of circulation on the web, i. e. the domain names (usually URLs) of the web pages on which the images can be found. With these data, three kinds of networks can be constructed: (a) image-label networks are built using the predefined tags or labels. For a research project, machine learning models can be used to analyse the content and visual composition on the basis of the annotations which are available as metadata. Researchers have to decide whether they prefer to use ready-made files provided by the given data collection methods or whether they prefer to construct networks on top of the web data performing search and filter operations. The latter approach uses a range of available software applications and thus facilitates the curation, building, visualisation, and analysis of the network. Images can be classified using the labels attached to them. These labels have been provided by the users of the various platforms or websites and enable the exploration of public attention for the issue in question. The network is constructed in a way that images and labels serve as nodes and the occurrences of labels serve as edges between them, thus grouping images with identical or similar labels together.

Secondly, (b) image-web entity networks display most frequently used named entities taken from the web context in which images were traced. They are built by Google Vision which annotates images according to the most common references with which visually similar images are captioned with. Such references are not necessarily descriptive with regard to the content of the image; rather, they relate to the images' context that is detected by Google's Knowledge Graph. It provides

Figure 1.2 Example of Image-Web Entities. Nodes Reveal Which Contexts or Knowledge Graph Entities Are Most Frequently Associated With an Image in Question (Here a Mosquito)



Source: Omena et al., 2021, p. 9.

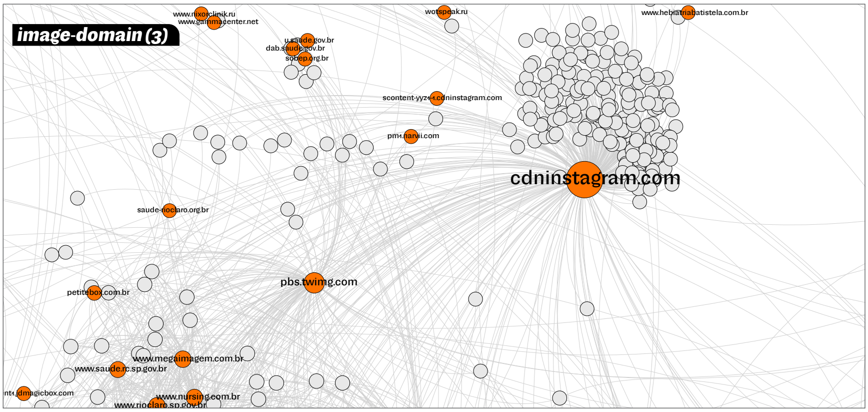
references from the image (here "mosquito") to entities such as persons, places, and things (here "zika virus", "microcephaly", "pregnancy", "ministry of health" etc). In this way, the Knowledge Graph goes beyond the content of the image itself and provides a structure to the network, where images and web entities serve as nodes and the occurrence of web entities (or other textual descriptions) in relation to images serve as edges.

The dichotomy between certain web entities (here *disease* versus *health*) may lead to the interpretation of a public controversy. As a methodological extension with respect to attention economy, we propose to combine this network approach with a qualitative analysis of image-text annotations to decide and evaluate what the entities refer to in specific situations.

The third kind of network rests on the images in combination with the URLs on which the images have been published. Similar visual contents are suggested by the powerful Google Image Search. These image-domain networks (c) allow to explore and analyse the web origin of those images, whereby images and domains serve as nodes and the occurrence of domains in relation to images serve as edges.

In this way, similar or identical images can be traced back to scientific or social communities, or influential agents in terms of distributing images across the web can be identified. In a way, this is a double-edged sword, since the provision of URLs reflects Google's opaque ranking algorithm and the particular order in which the web content is delivered to users. Omena et al.'s method can be extended by an analysis of concrete and situated user interactions on the detected websites to better understand their digital sociability. It is also possible to combine the presented computer vision

Figure 1.3 Example of an Image-Domain Network. Nodes Display Web Domains Which Host the Chosen Images



Source: Omena et al., 2021, p. 9.

networks with each other (e.g. web-entities and image-domains). The potential for sociological research is enormous: Not only can networks of images available on the internet be explored at scale – up to tens of thousands of images are thinkable –, but image-web entity networks allow for the analysis of the circulation of the recognised entities and can be blended into social networks, whereas image-domain networks open up the possibility to identify domain-specific social communities gathered around the use of the images which have been selected by the researchers analysing such networks. With respect to image-label networks, a dependency on the data provided by the Google Image API has certainly to be noted. However, at least with respect to photographic material, alternative means to annotate images are currently available: Even without the labels attached to the images from the internet it is possible to first perform object detection, object localization, and object qualification by applying open source algorithms such as InceptionV3, a deep learning model capable to autonomously detect objects like persons, horses, or a car. In a second step it is possible to extract and represent semantic elements by establishing a code system which serves the needs of the researchers working on this material (Arnold & Tilton, 2019). The latter approach does not depend on the labels attributed by the users of the internet, but allows for a culturally and socially constructed code system that has to be established by the researchers according to their needs. Consequently, the analysis of the images themselves and the objects contained therein can be conducted according to the premise of a sociologist. In sum, the analysis of the three types of networks provides the basis for a transfer into social networks and thus into a modified field analysis in the Bourdieusian sense. Besides, the point of independent data collection with regard to photographic material should not be underestimated: While the knowledge necessary to use Google's Image API has been extensively documented (Omena & Currie, 2022a; 2022b) and the methodology to construct the three types of networks has been thoroughly described (Omena et al., 2021), data collection and therefore all subsequent steps depend on Google's Knowledge Graph and/or on Google's ranking systems and search capabilities: "researchers must understand that what they are seeing includes the layered structure of online connectivity through Google's eyes" (Omena et al., 2021, p. 20). However, currently available sophisticated methodologies using existing machine learning applications facilitate data collection independent from data provision by a big tech company wherever this deems suitable to the researcher (Smits & Wevers 2023).

3.2 Image Similarity Analysis on Digitised Cultural Heritage Images

The second example presented here does not necessarily rely on images to be found on the internet, even though a sociological perspective on contemporary societies might be a priority. If the emphasis of analysis is more on a historical perspective and focuses on longitudinal evolution of e. g., developments in the field of art, digitised visual cultural heritage assets might be preferred. Similar to the approach described in the section above, such material can equally be analysed with computer vision methods; however, the methodological approach is quite different. Digital cultural heritage datasets usually contain rich metadata; in the case of images from art history these metadata contain information on the artist, the year when the work of art was created, the place of creation or publication as well as further information like the artistic movement to which the artist belonged or stylistic or technical features characterising this specific work of art. In computer vision, image-text combinations are termed multimodal and allow for the deployment of deep learning models (Smits & Wevers, 2023). Such an image dataset can be analysed on a pixel level using a convolutional neural network. The computation of the location of each image in a high-dimensional vector space is subsequently possible, since the number of pixels with identical or similar colour intensities as well as their characteristic combinations within an image allows for the transformation of the image files into vectors consisting of numbers. This calculation also provides the basis for the determination of image similarities and the application of clustering algorithms (cf. on the logic of metapictures Hoggenmüller & Klinke in this Special Issue). While this approach resembles the sociological grouping of individuals on the basis of features characterising them, as it may be employed for example in lifestyle analysis, the methodology of creating a high-dimensional vector space for big data analysis has been developed in the field of natural language processing (NLP), where the number of word occurrences in a corpus of documents form the strictly quantitative basis for arranging each document in a vector space (Turney & Pantel, 2010). With the advent of neural networks, this method was implemented in computer vision applications. It not only has the benefit of computationally determining the position of each image in the high-dimensional vector space, but also in a visual space in which the clustered images can be explored as if they were located on a two- or three-dimensional Cartesian coordinate system. To achieve this, the numbers of each vector are computed using a technique of dimensionality reduction – from a high-dimensional vector space into a two or three dimensional one – using an algorithm called UMAP (McInnes et al., 2018; specifically for t-SNE see again Hoggenmüller & Klinke in this Special Issue). PixPlot is one of those computer vision applications. It uses the Inception deep neural network for analysis of image similarities, clustering and visualisation of large image datasets. It has been developed by the Yale Digital Humanities Lab. If metadata associated with an image are available, it is possible to display them in an interactive visualisation.

A couple of research projects, especially from the digital humanities, have employed PixPlot in order to visualise, explore, and analyse large visual datasets (see Frischknecht in this Special Issue as an example). Furthermore, several cultural heritage institutions have used the technology developed by Yale (see e.g. the implementations by the National Library of Denmark https://labs.statsbiblioteket.dk/pixplot/ as well as the National Gallery of Denmark https://pixplot.smk.dk/). The Surprise Machines project has visualised more than 200 000 images from the Harvard Art Museums usually inaccessible to visitors of the museum. While each of these digital objects are publicly accessible online, a visualisation of the extensive image collection was presented as an exhibition on the premises of the museum (Rodighiero et al., 2022). The title of the exhibition and of the research project is revealing: *Surprise Machines* refers to the often unexpected and unpredictable outcomes of machine learning applications, thereby generating surprises in the individuals exploring the visualisation of the vector space model. PixPlot has also been used to provide a panoramic overview of more than 1 000 digital images which have been created as illustrations of Dante Alighieri's *Divina Comedia* since the publication of this narrative poem in the 13th Century. Images are arranged in clusters based on their similarities and presented alongside the Comedy's structure and plot; metadata can as well be browsed by the users on the project's website at https://divinecomedy.digital/#/ (Bonera & Bardazzi, 2022). While these projects refer to similar 'clustering' approaches in art history like Aby Warburg's Atlas Mnemosyne (Warburg, 2020), the difference between the approach of an art historian with her rich contextual knowledge and the computational determination of image similarities is obvious.

In contrast, the *Poscatálogo* (or *Postcatalog* in English) project took a different approach. The starting point here was Alfred H. Barr's famous diagram representing the most important artistic movements of the first decades of the 20th Century,

Figure 2.1 Barr X Inception CNN Visualisation of Clustered Image Data in a Vector Space Model, Which Can Be Understood as a Field in Bourdieusian Terms, Here a Full View of the Visualisation.

Source: Available online at https://digital-narratives.versae.es/ [Screenshot, Herms and Lehmann].

which charts formal influences and thus interprets the evolution and genealogy of art of modernity. This diagram was transformed into a high-dimensional vector space using the metadata accompanying the image dataset, especially the dates when the works of art were created. In a second step, an Inception convolutional neural network (CNN) was used to group the dataset of about 2 000 images into clusters and map them onto the diagram (Rodríguez Ortega et al., 2021). The resulting visualisation combines hand-curated metadata and computed clusters, thus allowing for an immersive exploration of the grouped images and their various relationships. The two-dimensional display of these grouped images enables an analysis of the development of the field over time and supports – in sociological terms – the exploration of the tensions between and sequence of "orthodox" and "heretical" artists (Bourdieu, 1979, p. 257). However, the exploration of structuring oppositions – such as cubism vs. surrealism – does not unfold in a straightforward way, because the images are clustered on the basis of their similarity and not according to their affiliation to a modernist movement, the painting style of which might have changed over time. This insight is put in a nutshell by using the term *Poscatálogo* – it implies the detachment from the traditional, man-made cataloguing systems and the interpretations provided by researchers. By contrast, it inaugurates visible and physically perceptible approaches to large image datasets grouped according to their computed similarities.

Figure 2.2　The Same Visualisation as in Figure 2.1, With a Zoom into One of the Clusters, in This Case Cubist Paintings.

Source: Available online at https://digital-narratives.versae.es/ [Screenshot, Herms and Lehmann].

Both figures illustrate the potential of image similarity approaches for arranging large amounts of images from art history, if they are combined with metadata. In the first figure, the elder modernist movements (e.g. impressionism) can be found on the left, while younger movements (e.g. abstract expressionism) are to be found on the right. The second figure shows how the clustering algorithm works, since it arranges similar images into one group (e.g. cubist paintings). However, not all cubist paintings are accumulated, since motifs and the colours used may vary.

Compared to traditional cultural sociological approaches, the potential of image similarity analyses to scale up from a narrow, qualitative approach to big visual data is particularly evident here. They could be further elaborated if the neural networks would be trained on traits like brushwork, stroke weight, composition or else. A further extension would be realised by including the likes or other kind of feedback which the used images received by retrieving such attention economy data via the Pinterest REST API, thus providing the clustering algorithm with further data.

At the same time and as it is often the case, the digital methodologies used in the named projects bring man-made epistemologies as cultural constructs and traditional representations of art history to the surface. In this way, they question customary classification systems and genealogical narratives, conceptions of creativity, originality, and influence with a long pedigree in art history, or the determination of similarity and difference as they have been established and systematically elaborated by art historians. The digital turn in art history might therefore unfold its disruptive potential in a way which may be perceived as unsettling by scientific researchers – the promise of advancing human knowledge by using computational approaches might be accompanied by the shattering of long-established art historical or sociological methodologies, a collateral damage that is not always welcomed if it contests the foundation of art history as a discipline (Rodríguez Ortega, 2019). However, the example of *Poscatálogo* shows that the clustering of similar images presents a productive provocation to art history oriented towards Bourdieu, because it would have to clearly identify what exactly "orthodoxy" and "heresy" (Bourdieu, 1979, p. 257) mean in terms of content. The metadata collected by cultural heritage institutions use categories that were established beforehand; it would be possible to group the images according to these descriptive labels, for example to visually arrange Cubism as opposed to Surrealism. Meanwhile, this would only result in a visualisation of an interpretation established beforehand. Image similarity clustering, by contrast, reveals new visual contexts, challenges inherited interpretations and concretises the visual rules established in the field. It thus supports the analysis of big visual data as a field in the Bourdieusian sense.

4 Conclusion

The two presented examples have underlined the significant potential of analytic possibilities for sociological analyses resting on big visual data. The first case study exemplifies how image-label networks can be created, answering the question *What does the users' visual attention focus on?* With respect to image-web entity networks, it has to be taken into account that those entities are added by Google and linked to the Google Knowledge Graph. As such, they enable the identification of various perspectives onto related topics, revealing image-related debates from a Google Knowledge Graph perspective. Image-domain networks, finally, enable the exploration of the socio-technical origin of the circulating images, sometimes pointing to domain-specific audiences gathered around the use of specific images, thus answering the question *Which social groups participate in the visual debate?* This methodology can be understood as an alternative to the contingency tables used by Bourdieu to analyse aesthetic preferences: The size of nodes in all three types of presented computer vision networks reflects different levels of image visibility, be it through image-labels, web-entities or image-domains. Visual attention can thus be interpreted as a resource that makes circulating images gain influence. Therefore, we propose that attention economy may be used as a helpful concept to interpret computer vision networks from a Bourdieusian perspective.

The second case study does not take images drawn from the web into focus, but image collections from cultural heritage institutions alongside with metadata curated by cultural heritage practitioners. The pixelwise analysis of the image allows, alongside with the available metadata, a clustering and visual arrangement of large amounts of images (tens or hundreds of thousands). Subsequently, an analysis in terms of content of the image clusters and the resulting visual arrangement can be performed, thus identifying the position of the groups of images in a virtual space.

Coming back to the main question of our contribution – is it possible for sociology resting on big visual data to *See like a field*?, we highlight that the methodologies presented above need to be complemented by qualitative, contextualised, and interpretative approaches. In this way, both examples open the door widely for the analysis of big visual data as a field in a Bourdieusian sense: The network analyses do not only enable interpretation about what these visual data are about, but also which social groups are engaged in the visual debates. Furthermore, if available attention economy data such as likes, emojis, or reposts are integrated as well, aesthetic preferences of the users can be analysed. Cultural styles around circulating images will then materialise in the process of analysis. Image similarity clustering draws the attention of the researcher to the content of the images themselves, thereby enabling an interpretation of who would constitute orthodox and heretic groups of image creators in the given image sample, as well as fostering an interpretation of what the visual exchange is really about. Taken together, these methodological approaches allow to a large extent for an interpretation of big visual data as Bourdieusian fields.

However, there are limitations to what can analytically be achieved with an approach that tries to *see like a field*. These limitations result from the notable gaps between researchers and big tech companies in terms of data access, ethical standards and strategic goals of big data analysis. While companies long for prediction of consumers' behaviour, researchers analyse trace data with the aim to better understand social dynamics in a platform context. By contrast, the asset of behavioural data and derived analytic products like identified patterns of aesthetic consumption and production as well as preferences of taste and social relationships are what enable big tech companies to match user dispositions and advertisements and thus to generate profits out of data analysis. This process of "commodifying people's behaviours" (Fourcade & Healy, 2017, p. 16) has been analysed by Shoshana Zuboff (2019) in her study *The Age of Surveillance Capitalism*. The economist analyses digital capitalism as a novel market form and develops concepts like "behavioural surplus" (Zuboff, 2019, pp. 63–97), the "epistemic coup" (i.e. the claim of ownership of knowledge in society by tech corporations, Zuboff, 2019, pp. 176–195, 495–595) and "instrumentarian power" (Zuboff, 2019, pp. 351–444). The first one is relevant in the context discussed here. As Zuboff explains, the identification of behavioural patterns allows for predictive analytics: "These machine intelligence operations convert raw material into the firm's highly profitable algorithmic products designed to predict the behaviour of its users" (Zuboff, 2019, p. 65). In combination with the trace data on language, cultural assets, traditions, customs, and preferences in terms of taste available only to big tech companies, such analytic products enable even the prediction of cultural production and consumption. This is a central point in Zuboff's argumentative framework, since it is exactly here that surveillance capitalism transcends the market logic as it was conceived so far, namely as an opaque mechanism where supply and demand are balanced: "Surveillance capitalism thus replaces mystery with certainty as it substitutes rendition, behavioural modification, and prediction for the old 'unsurveyable pattern'. This is a fundamental reversal of the classic ideal of the 'market' as intrinsically unknowable" (Zuboff, 2019, p. 497).

The availability of complete trace data at the hands of big tech corporations thus has significant consequences for the division of knowledge production especially in Western societies. Predictive analytics fills the blank in Bourdieusian field theory which we marked above, but it cannot be performed on the basis of attention economy data alone. If the power of disposition over complete trace data is decisive for the analysis of big visual data, sociological research endeavours are to be found on the dominated side of the division of knowledge production that has taken place since the advent of the Internet. For scientific research, it may be extremely attractive to explore cultural creativity as patterns of behaviour and to predict cultural production, but it has to be admitted that those analytic products rely on trace data which are beyond the reach of scientific knowledge. Big tech corporations will not make such data available since such patterns of consumption are eminently exploitable and thus serve the maximisation of profit they are looking for. As long as researchers don't have full access to trace data, it will not be possible for sociology to *see like a field*, to the same extent as platform companies *see like a market*.

5 References

Ahnert, R., Ahnert, S. E., Coleman, C. N., & Weingart, S. B. (2020). *The Network Turn: Changing Perspectives in the Humanities* (1st ed.). Cambridge University Press. https://doi.org/10.1017/9781108866804

Alton, Liz (2024, April 4). 7 tips for creating engaging content every day. https://web.archive.org/web/20240404010929/https://business.twitter.com/en/blog/7-tips-creating-engaging-content-every-day.html

Arnold, T., & Tilton, L. (2019). Distant viewing: Analyzing large visual corpora. *Digital Scholarship in the Humanities*, *34*(Supplement_1), i3–i16. https://doi.org/10.1093/llc/fqz013

Barabási, A.-L. (2009). *Linked. How everything is connected to everything else and what it means for business, science, and everyday life* ([18. edition] Plume print; authorized reprint of a hardcover ed. publ. by Perseus Publ.). Plume.

Bernhard, S. (2008). Netzwerkanalyse und Feldtheorie. Grundriss einer Integration im Rahmen von Bourdieus Sozialtheorie. In C. Stegbauer (Ed.), *Netzwerkanalyse und Netzwerktheorie: Ein neues Paradigma in den Sozialwissenschaften* (pp. 121–130). VS Verlag für Sozialwissenschaften. https://doi.org/10.1007/978-3-531-91107-6_8

Bonera, M., & Bardazzi, A. (2022). Data Visualization as a Tool to Experience the Legacy of Dante's Divine Comedy and its Influence on the Cultural Heritage. *Bibliotheca Dantesca: Journal of Dante Studies*, *5*(1), 288–297.

Bouk, D. (2017). The History and Political Economy of Personal Data over the Last Two Centuries in Three Acts. *Osiris*, *32*(1), 85–106. https://doi.org/10.1086/693400

Bourdieu, P. (1971). Une interprétation de la théorie de la religion selon Max Weber. *European Journal of Sociology / Archives Européennes de Sociologie / Europäisches Archiv Für Soziologie*, *12*(1), 3–21.

Bourdieu, P. (1979). *La distinction. Critique sociale du jugement*. Les Éd. de Minuit.

Bourdieu, P. (1992). *Les Règles de l'Art. Genèse et Structure du Champ Littéraire*. Éd. du Seuil.

Centola, D. (2011). An Experimental Study of Homophily in the Adoption of Health Behavior. *Science*, *334*(6060), 1269–1272. https://doi.org/10.1126/science.1207055

Cha, M., Haddadi, H., Benevenuto, F., & Gummadi, K. (2010). Measuring User Influence in Twitter: The Million Follower Fallacy. *Proceedings of the International AAAI Conference on Web and Social Media*, *4*(1), 10–17. https://doi.org/10.1609/icwsm.v4i1.14033

de Nooy, W. (2003). Fields and networks: Correspondence analysis and social network analysis in the framework of field theory. *Poetics*, *31*(5), 305–327. https://doi.org/10.1016/S0304-422X(03)00035-4

Demandsage Instagram Statistics. (2024). https://www.demandsage.com/instagram-statistics/

Escofier, J.-P. (2024). *Petite histoire des mathématiques*. Dunod.

Fourcade, M., & Healy, K. (2017). Seeing like a market. *Socio-Economic Review*, *15*(1), 9–29. https://doi.org/10.1093/ser/mww033

Fraiberger, S. P., Sinatra, R., Resch, M., Riedl, C., & Barabási, A.-L. (2018). Quantifying reputation and success in art. *Science*, *362*(6416), 825–829. https://doi.org/10.1126/science.aau7224

Franck, G. (1998). *Ökonomie der Aufmerksamkeit: Ein Entwurf*. Hanser.

Frischknecht, M. (2025). Through the Eyes of the Machine: Exploring Historical Photo Collections with Convolutional Neural Networks. *Schweizerische Zeitschrift für Soziologie, 51*(2), Special Issue hrsg. von S. W. Hoggenmüller, Big Visual Data als neue Form des Wissens: Potenziale, Herausforderungen und Transformationen.

Hennig, M., & Kohl, S. (2012). Fundierung der Netzwerkperspektive durch die Habitus- und Feldtheorie von Pierre Bourdieu. In M. Hennig & C. Stegbauer (Eds.), *Die Integration von Theorie und Methode in der Netzwerkforschung* (pp. 13–32). VS Verlag für Sozialwissenschaften. https://doi.org/10.1007/978-3-531-93464-8_2

Hoggenmüller, S. W., Klinke, H. (2025). Metabilder als Forschungswerkzeuge: Zur Kontingenz und algorithmischen Bedingtheit ihrer Herstellung. *Schweizerische Zeitschrift für Soziologie, 51*(2), Special Issue hrsg. von S. W. Hoggenmüller, Big Visual Data als neue Form des Wissens: Potenziale, Herausforderungen und Transformationen.

Janning, F. (1991). *Pierre Bourdieus Theorie der Praxis. Analyse und Kritik der konzeptionellen Grundlegung einer praxeologischen Soziologie.* (Vol. 105). Westdeutscher Verlag.

Ma, X., & Fan, X. (2022). A review of the studies on social media images from the perspective of information interaction. *Data and Information Management*, *6*(1), 100004. https://doi.org/10.1016/j.dim.2022.100004

McInnes, L., Healy, J., Saul, N., & Großberger, L. (2018). UMAP: Uniform Manifold Approximation and Projection. *Journal of Open Source Software*, *3*(29), 861. https://doi.org/10.21105/joss.00861

Omena, J. J., & Currie, M. (2022a). *Collecting Data Using APIs Part 1: How to Understand APIs and Navigating API Documentation* [How-to Guide]. SAGE Research Methods: Doing Research Online. https://doi.org/10.4135/9781529611441

Omena, J. J., & Currie, M. (2022b). *Collecting Data Using APIs Part 2: How to Communicate With and Access APIs* [How-to Guide]. SAGE Research Methods: Doing Research Online. https://doi.org/10.4135/9781529611458

Omena, J. J., Elena, P., Gobbo, B., & Jason, C. (2021). The Potentials of Google Vision API-based Networks to Study Natively Digital Images. *Diseña*, *19*, Article 1, 1–25. https://doi.org/10.7764/disena.19.Article.1

Pörksen, B., & Detel, H. (2012). *Der entfesselte Skandal: Das Ende der Kontrolle im digitalen Zeitalter*. Halem.

Reckwitz, A. (2020). *Die Gesellschaft der Singularitäten: Zum Strukturwandel der Moderne*. Suhrkamp.

Rodighiero, D., Derry, L., Duhaime, D., Kruguer, J., Mueller, M. C., Pietsch, C., Schnapp, J. T., Steward, J., & metaLAB. (2022). Surprise machines: Revealing Harvard Art Museums' image collection. *Information Design Journal*, *27*(1), 21–34. https://doi.org/10.1075/idj.22013.rod

Rodríguez Ortega, N. (2019). Digital Art History: The Questions that Need to Be Asked. *Visual Resources*, *35*(1–2), 6–20. https://doi.org/10.1080/01973762.2019.1553832

Rodríguez Ortega, N., de la Rosa, J., Ruiz Gonzálvez, G., Ortiz Tello, M., & Crespillo Marí, L. (2021). Poscatálogo, o de cómo y por qué experimentar físicamente un espacio vectorial de alta dimensión. *CIMED21 – I Congreso internacional de museos y estrategias digitales*, 823–833. https://doi.org/10.4995/CIMED21.2021.12487

Serino, M., D'Ambrosio, D., & Ragozini, G. (2017). Bridging social network analysis and field theory through multidimensional data analysis: The case of the theatrical field. *Poetics*, *62*, 66–80. https://doi.org/10.1016/j.poetic.2016.12.002

Smits, T., & Wevers, M. (2023). A multimodal turn in Digital Humanities. Using contrastive machine learning models to explore, enrich, and analyze digital visual historical collections. *Digital Scholarship in the Humanities*, *38*(3), fqad008. https://doi.org/10.1093/llc/fqad008

Trezzini, B. (1998). Theoretische Aspekte der Sozialwissenschaftlichen Netzwerkanalyse. *Schweizerische Zeitschrift für Soziologie / Revue Suisse de sociologie, 24*(3), 511–554.

Turney, P. D., & Pantel, P. (2010). From Frequency to Meaning: Vector Space Models of Semantics. *Journal of Artificial Intelligence Research*, *37*, 141–188. https://doi.org/10.1613/jair.2934

Warburg, A. (2020). *Aby Warburg: Bilderatlas Mnemosyne – The Original* (R. Ohrt, A. Heil, B. M. Scherer, W. H. Sherman, & C. Wedepohl, Eds.). Hatje Cantz.

Zuboff, S. (2019). *The Age of Surveillance Capitalism. The Fight for a HumanFuture at the New Frontier of Power* (First edition). Public Affairs.