Towards a Digital Reflexive Sociology: Using Wikipedia's Biographical Repository as a Reflexive Tool

2 authors:

Pablo Beytía
Humboldt-Universität zu Berlin
53 PUBLICATIONS   **220** CITATIONS

SEE PROFILE

Hans-Peter Mueller
Humboldt-Universität zu Berlin
154 PUBLICATIONS   **776** CITATIONS

SEE PROFILE

Towards a Digital Reflexive Sociology: Using Wikipedia's Biographical Repository as a Reflexive Tool

Pablo Beytía – Hans-Peter Müller

Humboldt University of Berlin

Abstract

We propose the development of 'digital reflexive sociology', understood as the use of digital methods and Big Data to reflect on the social and historical circumstances of sociologists and sociological thinking. To show this approach's potential, we employ Wikipedia as a 'reflexive tool', i.e., an external artefact of self-observation that can help sociologists to notice conventions, biases, and blind spots within their discipline. We analyse the collective patterns of the 500 most notable sociologists on Wikipedia, performing structural, network, and text analyses of their biographies. Our exploration reveals patterns in their historical frequency, gender composition, geographical concentration, birth-death mobility, centrality degree, biographical clustering, and proximity between countries, also stressing institutions, events, places, and relevant dates from a biographical point of view. Linking these patterns in a diachronic way, we distinguish five generations of sociologists recorded on Wikipedia and emphasise the high historical concentration of the discipline in geographical areas, gender, and schools of thought. Drawing on these results, we discuss the potential of using digital repositories and methods to enhance reflexivity within sociology.

Keywords: reflexive sociology, digital sociology, sociology of knowledge, computational social science, digital methods.

Introduction

We live in a strongly mediatised world, where both daily practices and the social construction of reality increasingly rest on communication technologies and are largely shaped by them (Hepp 2016; Couldry and Hepp 2018). The current 'wave of digitalisation' of communications (Couldry and Hepp 2018) and its corresponding 'datafication' process –i.e., the transformation of all things into a data format (Schäfer and Van Es 2017)– are generating an unprecedented record about people's lives that is currently being used to answer scientific questions of all kinds (D. Lazer et al. 2009; Pentland 2014).

The social sciences, however, have responded slowly to the implementation of computational methods to analyse large amounts of digital information (Lazer et al. 2009; Burrows and Savage 2014; Mützel 2015). Whereas the ability to collect and analyse massive data rapidly transformed fields such as physics and biology, the use of these procedures in the social analysis was initially developed only in large Internet companies (Lazer et al. 2009). Particularly in sociology, there were no widespread interest in the use of techniques to build or manipulate large digital databases (Mützel 2015). Unlike what happened a few decades ago –when sociologists pioneered the development of sampling theory, survey design, and qualitative interviewing– the discipline has been cautious in innovating methodologies to study digital information (Savage and Burrows 2007; Burrows and Savage 2014). Although this situation has been changing in recent years –with the consolidation of digital methods (Rogers 2013; 2015), computational social science (Salganik 2019; Lazer et al. 2020), and digital sociology (Lupton 2014; Marres 2017)– there are still many areas where sociology has barely explored these new approaches.

Reflexive sociology is one of them. Following the original formulation by Alvin Gouldner (1971: 494), reflexive sociology is the attempt by sociologists to deepen their awareness of who and what they are in a specific society and at a given time, and how both their social position and personal praxis affect their work as sociologists. It emerged as a critical response to positivism: rejecting its notion of knowledge as external information that can be 'discovered', reflexive sociology assumes that researchers and their objects are interrelated and mutually constituted, so that it is not possible to aspire to a value-free empirical analysis (Gouldner 1971). According to this perspective, all sociologists have adherences and adhesions, affiliation or filiation links, and have developed, throughout their education, prejudices, repressions, and omissions of which they

are hardly aware (Bourdieu 2016). That is why it is needed a practice where they objectify themselves, study their circumstances and determinations, and become attentive to their biases and blind spots. Reflexive sociology would be that form of 'epistemological surveillance' that some researchers consider a prerequisite for achieving rigorous social knowledge (Bourdieu and Wacquant 1992).

Initially, reflexive sociology was an innovative practice willing to use creative methods for scientific self-observation –such as correspondence analysis (Bourdieu 1988)–. However, there has been scarce research on reflexive sociology based on large digital information sources in the last two decades. That might signal that digital data does not help obtain self-awareness among sociologists. Nonetheless, that does not seem to be the case: in the same two decades, various areas of reflexive studies that work primarily with computational methods and large amounts of digital information emerged –e.g., bibliometrics, scientometrics, and culturomics–, which nowadays are even grouped in macro-reflexive areas of science, such as 'metaknowledge' (Evans and Foster 2011) or the 'science of science' (Fortunato et al. 2018).

Because of the high potential of large digital databases for understanding social problems and the relevance of reflexive research to achieve a self-critical perspective within sociology, we take up the challenge of performing what we call *'digital reflexive sociology*': to use digital methods and Big Data to systematically reflect on the social and historical circumstances of sociologists and sociological thinking. The term 'digital sociology' refers to a broad branch of sociology that combines professional digital practice, analyses of digital technology use, digital data analysis, and critical digital sociology (Lupton 2014). Here, we refer specifically to the analysis of digital data - i.e., materialised from bits or binary digits[1]- to reflect on the circumstances that have surrounded sociology, affecting its historical development and present configuration.

In concrete terms, we will employ the biographical repository on Wikipedia –the digital, polyglot, and free encyclopaedia– to explore the social and historical configurations of the world's best-known sociologists. We will focus on Wikipedia because it is currently one of the world's most comprehensive and influential multicultural repositories of knowledge. It contains more than 55 million articles (6.22 million biographies) in over 300 languages (Wikipedia 2022a; 2022b). It is the tenth most visited website globally and the first dedicated to systematically compiling

---

[1] The bit (acronym for "binary digit") is the minimum unit of information. It can be thought of as a response to a two-choice situation (Shannon and Weaver 1949): when a specific circumstance admits two possible alternatives (e.g. yes/no, on/off, true/false.), they can be encoded in one bit using the values 0 and 1.

knowledge (Alexa Internet 2022). Moreover, in recent years it has been increasingly used as a source of scientific research (e.g., Apic, Betts, and Russell 2011; Ronen et al. 2014; Schich et al. 2014; Reznik and Shatalov 2016; Menini et al. 2017; Rollin, Lages, and Shepelyansky 2019).

From this enormous source of global and multilingual 'social memory' (Olick and Robbins 1998), we select the sociologists (i.e., persons labelled as such in Wikipedia) who have their biographies written in ten or more languages and could, therefore, be considered the most 'multiculturally notable' [2] sociologists on this platform. Starting from these sociologists' biographical content, we subsequently carried out a 'prosopography': an "investigation of the common background characteristics of a group of actors in history by means of a collective study of their lives" (Stone 1971 p. 46). This method, roughly speaking, consists of investigating characteristics, attributes and/or biographical relationships of a universe of individuals and processing this information to describe the 'collective profile' of that group. In this case, we focus on describing 'emergent properties' –i.e., characteristics observable in the aggregate of biographies (not in particular cases), such as tendencies, patterns, distributions and/or networks– that allow us to collectively portray the most notable sociologists and to historically contextualise their construction of sociological thinking. We then perform a diachronic analysis of these properties to identify historical ruptures and generations in sociologists' history recorded on Wikipedia.

Despite its usefulness in identifying historical patterns, we do not use this method to produce a more reliable version of the history of sociology. Rather, we argue that Wikipedia is a massive repository of global and multicultural information that can be used as what we term a 'reflexive tool', i.e., an external artefact of self-observation (in the manner of a mirror) that allows sociologists to objectify their discipline without impregnating the data, from the outset, with biases of their own perspective (e.g., related to their language, culture, geographical position, or disciplinary sub-specialisation)[3]. Using a reflexive tool for scientific self-observation is valuable because it provides a potentially divergent perspective on our own discipline, which can cause dissonance with our knowledge, opinions, and beliefs as participating agents. Such a process can help to problematise conventions and reveal our biases and blind spots with greater clarity.

---

[2] We refer to notability in the same sense used in Wikipedia's editorial policies. A person would be notable if there is verifiable information about his or her life from reliable sources independent of him or herself (Wikipedia 2022b). A person would be 'multiculturally notable' if they meet the criteria of notability in multiple languages.
[3] Wikipedia, as a cultural product, can be understood as a "tool kit" (Swidler 1986), that is, a repertoire of symbols and stories from which actors (in this case sociologists) can select pieces to define their next lines of action.

This study has two complementary goals. Our first (and direct) objective is to *contribute to the disciplinary discussion on the collective circumstances in which the work of the most prominent sociologists has been situated*. Here, our focus is on *developing a collective profile of the most notable sociologists as they are conceived in Wikipedia (our reflexive tool)*. On the one hand, this implies faithfully displaying the information curated by Wikipedia and avoiding reworking the data according to criteria internal to sociology (e.g., selecting only sociologists with university training, publications in sociology journals, or membership in associations). On the other hand, it implies focusing on describing properties that emerge in the set of sociologists studied rather than in individual biographies (e.g., spatial distributions, gender compositions, typical birth and death trajectories, and networks of biographical ties).

Our second (and indirect) goal is *to illustrate the potential that digital repositories and methods have for reflexive sociology*. We propose that this approach significantly extends the research scope since it makes possible the observation of global, multilingual, historical, and associative patterns that can be linked to the lives and works of thousands (and potentially millions) of sociologists spread across many periods, countries, and cultures. This potential has been noticed, though in a fragmented way, in previous studies: they have shown the capacity of digital methods to study sociology on a global scale (e.g., Demeter and Toth 2020), to observe historical changes (e.g., Korom 2020) and/or to visualise associative patterns (e.g., Cárdenas 2021). Here we show that the use of large digital databases makes it possible to study these and other levels of research together and interrelatedly, thus obtaining a more comprehensive picture of the discipline.

This study is organised into six sections. The first one reviews previous studies that could be considered examples of digital reflexive sociology. The second section explains how the record of biographies is constructed in Wikipedia, discussing the meaning of this information and its usefulness for sociological research. The third one details the methods used in this study, while the fourth and fifth analyse the collected information from a comprehensive and generational perspective. Finally, we discuss this study's overall results, examining what a digital approach could offer to reflexive sociology.

Related work

Within the 'digital sociology', there have been reflexive studies that have examined the implications, for sociologists, of studying *and* participating in the digital sphere (Burrows 2012; Lupton, Mewburn, and Thomson 2017). However, they have not examined *the worldwide circumstances in which sociologists and sociological thinking have been historically situated.* From our perspective, this is an essential task of reflexive sociology, because it allows sociologists to objectify themselves and their work, recognise their own historical circumstances, and increase awareness of their biases and blind spots. The study of sociology's circumstances enables both an understanding of who sociologists are in a specific society and time (Gouldner 1971), as well as the development of an informed 'epistemological surveillance' (Bourdieu and Wacquant 1992).

A few studies have used large digital data sources to describe macro circumstances in sociology. Demeter and Toth (2020) studied, from university websites, the networks and educational paths of faculty members in 100 leading sociology departments worldwide. They found an over-representation of academics from Western countries in the Global North (such as the United States, the United Kingdom, and Germany) and a considerable male bias, especially in senior academic positions.

Cárdenas (2021) researched, also from institutional websites, the editorial boards of the 100 sociology journals with the highest impact. His conclusion was that a small number of editors with similar profiles belong simultaneously to many journals. In them, the participation of Anglo-Saxon universities prevails whereas the universities from the Global South are marginalized (Cardenas 2021).

Adams, Brückner and Naslund (2019) investigated whether the gender and race proportions of sociologists at North American universities correlate with that of American sociologists with Wikipedia biographies. They reported a consistent underrepresentation of women and people of colour, which is not explained by occupational elements such as academic rank, career length, or the reputation of sociology departments.

Finally, Korom (2020) collected a large corpus of sociological texts from journals, textbooks, handbooks and encyclopaedias to compare the most cited sociologists in 1970 and 2010. His conclusion was that the most prestigious sociologists in 1970 were mainly based in the USA

showing little mobility, while there is greater European participation and more inter-university movements today.

These studies are significant examples of the use of large digital databases and computational methods –i.e., computer-based techniques to analyse, model, and simulate phenomena– to understand the circumstances of sociology worldwide. They testify to the potential of digital reflexive sociology. Each investigation focused on a particular disciplinary object: faculty membership, journal committees, encyclopaedic representation, and citations in texts. But taken together, some cross-cutting trends stand out, especially the idea of an over-representation of countries from the Global North and a significant bias towards male participation.

Among these studies, only one was interested in *how sociology is perceived and communicated in an external environment* (Adams, Brückner, and Naslund 2019). As we said, we consider this approach to be promising. Not only because it allows detecting biases in the media describing sociology (as that study highlights), but also because it allows looking at the discipline from an external perspective, which has the potential to illuminate aspects that might not be easily perceived by its participants. It would be appropriate, however, to extend the scope of that study, as it investigated the academy of one country (United States), in one specific year (2014) and considering biographies in one language only (English).

On the other hand, there is one research that investigated historical processes in the discipline (Korom 2020), comparing two points in time (1970 / 2010). We believe that this processual perspective is fundamental for disciplinary reflexivity, since it allows us to account for the circumstances that have affected the production and reproduction of sociological knowledge at different times, and with it, a better understanding of the structures of knowledge accumulated by the discipline. It would also be promising to extend this processual analysis to the entire history of sociology. This would reduce the risk, for sociologists interpreting their discipline, of decontextualising meanings or producing ideological effects (Bourdieu 1993).

The biographical record on Wikipedia and its reflexive potential

This article aims to perform 'digital reflexive sociology'. For this, we will use biographical information documented in multiple languages within Wikipedia. The employment of this data source is justified by 1) the enormous amount of information it gathers –more than 55 million

articles, including over 6.22 million biographies–, 2) the multicultural breadth of its editorial processes and the information generated –which is edited and published in 317 languages–, 3) the influence that this source of information has on daily life –as it is the tenth most visited website on the planet, and the most popular dedicated to systematising knowledge (Alexa Internet 2022)–, and 4) the possibilities it offers to extract multilingual information in structured data format –which are organised in Wikidata[4], the general data repository of the Wikimedia Foundation–.

Despite these reasons, the use of information from an encyclopaedia 'that anyone can edit' needs additional clarifications. We will then explain a) how the content in Wikipedia is selected and edited, b) how reliable its information is, c) what type of information is stored in this data, d) how the content should be interpreted, and e) what is its reflexive sociological potential.

a) How is the content selected and edited in Wikipedia? Wikipedia content is selected and edited by a network of 'actants' (Latour 2005) –including humans and non-human agents– that acts and interacts under the circumstances established by a specific socio-technical framework. This framework is composed of at least three elements: a technological infrastructure (the MediaWiki software, which defines the possibilities of editing and interaction between users), a normative framework (rules, guidelines, and institutional aspirations that seek to regulate behaviour and guide the selection of content), and an editorial stratification system (which facilitates coordination through the distinction of roles, responsibilities and editing rights). Under these circumstances, human and nonhuman editors establish processes of spontaneous coordination, conflict, or mandatory coordination (guaranteed by institutional mechanisms), which result in temporary consensus on each article[5].

Two editorial rules are especially relevant to the biography selection process. First, the 'no original research' rule, which states that Wikipedia is not and should not be a medium of primary sources, but a compendium based on verifiable secondary sources (Wikipedia 2019). Second, the 'notability test': 'If a topic has received significant coverage in reliable sources that are independent of the subject, it is presumed to be suitable for a stand-alone article' (Wikipedia 2022b). This second

---

[4] Wikimedia Foundation, "Wikidata." https://www.wikidata.org (reviewed 9 September 2019).

[5] To understand in detail the technological infrastructure, the normativity, the editorial stratification, the agency and the interaction dynamics of Wikipedia, see: Viégas, Wattenberg, and Dave 2004; Kittur et al. 2007; Lih 2009; Yang and Lai 2010; Jemielniak 2014; Biuk-Aghai, Pang, and Si 2014; Clément and Guitton 2015; Tsvetkova et al. 2017; Yenikent, Holtz, and Kimmerle 2017.

rule also insists that editors should avoid self-promotion and advertising through articles. In general, these norms aim to ensure –with greater or lesser success– that all people with biographies on Wikipedia have significant coverage in reliable, verifiable, and independent media.

b) How reliable is Wikipedia? Wikipedia's reliability has been tested at least by evaluating its articles on natural sciences (Giles 2005), medicine (Pender et al. 2008), pharmacology (Clauson et al. 2008), history (Rector 2008), philosophy (Bragues 2007), and political science (Brown 2011). Famously, the journal *Nature* accredited, through a blind experiment with experts, that Wikipedia has a low rate of 'serious errors' in scientific articles and it has them in a similar degree to the Encyclopaedia Britannica (Giles 2005).

However, two aspects of reliability must be distinguished: accuracy and completeness. Concerning the *accuracy*, several studies mention that Wikipedia has a good level of precision; even on political –and therefore conflicting– issues, it generally provides reliable information (Brown 2011; Giles 2005). However, various studies indicate that this encyclopaedia is *incomplete* in the sense that it omits relevant information (Bragues 2007; Brown 2011; Lavsa et al. 2011). In any case, the content documented in it should not be considered as neutral or value-free, since, despite the openness of its editorial processes, it reflects the views and interests of the people who edit it (Royal and Kapila 2009; Tkacz 2012). In this regard, Wikipedia has also proven to have some collective biases in selecting its content. For example, among its articles, there is a disproportionate amount of information about men, western countries, and events of recent historical periods (Graham et al. 2014; Wagner et al. 2015; Reznik and Shatalov 2016; Samoilenko et al. 2017; Beytía 2020; Beytía and Wagner 2022; Beytía et al. 2022).

To increase the reliability of this information, we took two methodological precautions in this research: 1) to reduce errors of completeness, we consider as the initial source the biographical record of sociologists written in *all* languages available on Wikipedia; 2) to mitigate errors of accuracy, we analyse only *barely equivocal or interpretable* biographical information –data on the birth (place and date), gender, death (place and date), and references between biographies–.

c) What type of information is stored in this data? Despite their reference to detailed historical facts and their considerable level of accuracy, the data and patterns found in Wikipedia information are

not historical data. What is observed is instead a type of 'social memory' (Olick and Robbins 1998) about past events.

As 'memory', it can never be a complete preservation of the past. In fact, one of the most relevant social functions of memory is the handling of oblivion (Luhmann 1999). What is remembered must be understood as a 'contemporised past' (Assmann 1995), a reconstruction that makes a society, in a specific time, and from its own frame of reference (Halbwachs 1997).

With the qualifier of 'social', on the other hand, we want to indicate that the reference of this type of memory is not in individual consciousnesses or minds (as in the case of the 'individual' memory), but rather in communicative operations that are socially organised (Luhmann 1999; Esposito 2008). That is especially clear in Wikipedia since its articles do not store someone's memory, but rather a collective editorial result permanently open to new collaborations and changes.

To be more specific, we can be distinguish two main types of social memory: communicative and cultural (Assmann 1995; 2011). While 'communicative' memory is based on everyday oral communication –thus having a temporal horizon reduced to a few decades–, 'cultural' memory is based on objectivized forms of culture –such as texts, images, rites, buildings, monuments, or cities– which allow memory to transcend the temporal horizon of everyday life. Following this distinction, the social memory documented in Wikipedia would be of a *cultural* type (Jara-Figueroa, Yu, and Hidalgo 2019) since it is constructed based on texts, images, and other physical forms of record.

d) How should the content extracted from Wikipedia be interpreted? Media always involves content selection. Therefore, this information must be clearly distinguished from the information obtained, for example, in a census. Assuming that a media cannot show everything that happens (it must select content) and its perspective is delineated by editorial positions (in this case collective), the content analysis must be especially aware of the criteria that each media uses for selection. Regarding the registration of biographies in Wikipedia, three selection criteria stand out: 1) the level of coverage in other sources that the person in question has, 2) the reliability and independence of that person that those sources have, and 3) that the information from the sources is verifiable. All these criteria revolve around the same element: the quantity and quality of the *sources* on which the articles are built.

On the other hand, researchers should approach this information assuming that the content on Wikipedia results from a 'double hermeneutic' process. Anthony Giddens (1984), thinking fundamentally in the processes of qualitative research, proposed that sociology necessarily works on the intersection of two mutually influencing frames of meaning: the actors interpret their social world (first-order observation) and the sociologist constructs metalanguages on those frames of meaning (second-order observation). Within Wikipedia there is also this double hermeneutic: given the norm of 'no original research', all articles are interpretations of interpretations (external sources). Thus, two levels of interpretation (external and internal) affect the result of the content. The sociologist's perspective would rather be in a 'third-order position': observing second-order interpretations.

e) What is the sociological potential of this information? Millions of people, over two decades, have written millions of articles on all themes and historical figures that seem relevant to them, and have done so in multiple languages, cultures, territories, educational training, and social circumstances. Today, we have computational techniques to visualise the results of that social process and the relationships that have been traced among all the articles, showing how this form of social memory has been structured. That is the great potential of this information: the possibility of systematically observing the organisation of a type of social memory that has been configured and depurated in long global and intercultural processes.

For reflexive sociology, this source of information offers at least three attractive opportunities:

1. *Automatic access to a broad record of biographical information*: it is possible to automatically extract (from a directory that exceeds 6,220,000 registered people) the list of all sociologists who have at least one biographical article in any Wikipedia language, along with basic information about their lives. It is also possible to trace the hyperlinks or references between those biographies. This multicultural information would be very difficult to obtain, at that level of detail, with nondigital methods.

2. *Access to intersubjective and multicultural information:* this type of record has not been generated based on a personal perspective. It is the unintended, intersubjective result of a worldwide network of editors from multiple cultures and social backgrounds. While this does not eliminate the existence of collective biases in information, it does mitigate some

information selection biases associated with the researcher's perspective (e.g.: ethnocentric biases, or disciplinary subspecialisation biases).

3. *Access to external social memory as a tool for sociological reflexivit*y: this type of information allows sociology to study itself without basing on itself, that is, considering an external structure of memory as a relevant source for disciplinary reflexivity[6]. Although it would be rare, this type of information may have historical inaccuracies. However, even if so, it remains a direct and accurate source of social memory, and in that sense, it has sociological value despite its potential inaccuracies or omissions.

In recent years, researchers have used information from Wikipedia in various scientific disciplines, such as media studies, geography of information, computer science, computational linguistics, social physics, or biology. It has been employed to study social phenomena, such as the relations between languages (Ronen et al. 2014), geopolitical instability (Apic, Betts, and Russell 2011), or the global influence of infectious diseases (Rollin, Lages, and Shepelyansky 2019). It has also been used to analyse historical processes, such as large demographic trends (Reznik and Shatalov 2016), birth-death migration trajectories (Schich et al. 2014), or mobility of famous people during their lives (Menini et al. 2017). Could it also be a valuable source of information for reflexive sociology?

Methods

We extracted the multilingual data of Wikipedia from Wikidata (Wikimedia Foundation 2020), the central repository of structured data of Wikimedia Foundation. It is an open source of information released and updated in real-time, collaboratively maintained by specialised editors and artificial intelligence. They are responsible for systematically collecting and organising the information continuously processed in the multiple languages of Wikipedia.

Considering its 317 language versions, Wikipedia contains more than 6.22 million biographies, which generally specify the portrayed people's occupation. From this Big Data repository, we constructed a discipline-specific database that comprises 9,658 biographies of

---

[6] Christian Borch (2012) used a similar approach, tracking the debates of crowds to study the history of the main sociological discussions in Germany, France and the United States.

people that Wikipedia catalogues as sociologists. From that universe, we selected the sociologists (persons who are identified as such) who fulfilled two requirements: a) they were born after 1798[7] and b) they have their biography written in ten[8] or more languages. This method to select the most 'notable' biographies, in the multilingual sense, has been used in previous studies (Jara-Figueroa, Yu, and Hidalgo 2019; Beytía 2020) and validated as an appropriate way to find historically relevant figures (Yu et al. 2016). We obtained a database with the 500 sociologists with biographies in more languages within Wikipedia.

Not all people identified as a sociologist in Wikipedia are sociologists of the *main occupation* (some, for example, are mainly recognised as politicians, philosophers, or writers). Some of them also did not have *formal training* as sociologists –this is the case by the way of classics and founders of the discipline, such as Comte, Marx, Durkheim, Weber, Simmel, and Mead. Nevertheless, *they have been recognised as sociologists in some of the Wikipedia versions*, either because of their formal education[9] or because of their professional trajectory relevantly linked to the discipline. Figure 1 shows the frequency and male percentage of these biographies by the decade of birth.

Figure 1. Frequency and gender proportion of sociologists disseminated in ten or more languages on Wikipedia by the decade of birth

---
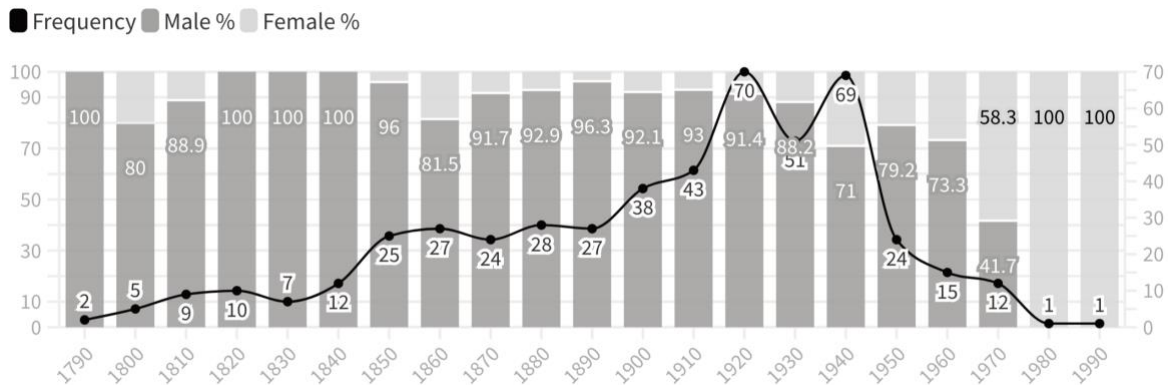
[7] We selected this reference year to include all the founders of sociology (e.g., Auguste Comte or Herbert Spencer) and the older classical referents (e.g., Alexis de Tocqueville or Karl Marx). Auguste Comte was born in 1798.

[8] We put up this language barrier to study sociologists with global relevance. Since there is no precise number of languages to define "global notability," we manually checked the database to verify the worldwide status of sociologists and the completeness of their biographical information in Wikipedia. From ten languages onwards, the information was complete, and we could identify at least one sociologist from the list.

[9] Two people with formal education as sociologists were discarded, because their dissemination on Wikipedia evidently has no connection with sociology: the Spanish football referee Alberto Undiano Mallenco, and the model and daughter of U.S. President Tiffany Trump.

We sorted the list of persons according to the number of languages (or 'multilingual ubiquity') in which each biography was available, ranging from 10 (e.g., Luc Boltanski) to 260 (Karl Marx). Subsequently, we extracted information for each sociologist on gender, birth (year, place, geographical coordinates), death (year, place, and geographical coordinates), and the institutions in which they were educated and employed. 87% of these sociologists had information from institutions where they were educated and 72% from organizations where they were employed. From the place of birth, we deduced the country in which they would have been born according to the current geopolitical boundaries[10]. The geographical coordinates of birth and death were also used to approximate relevant mobility trajectories. This indicator should not be understood as a complete representation of geographical mobility but rather as a sample of geographical movements that were permanent and definitive with a few exceptions[11].

Subsequently, we tracked the hyperlinks (or web links) among the 500 digital biographies of the most widely disseminated multilingual sociologists on Wikipedia, using the 'rvest' package available in the statistical software R (Wickham 2016). This method has also been used in previous research (Skiena and Ward 2014; Beytía 2020) and validated as an appropriate way to document the mentions or references between biographies within Wikipedia (Beytía and Schobin 2020). We extracted the hyperlink network from English Wikipedia because it is the language version with the most complete record: it has the most articles, biographies, edits, and editors (Nemoto and Gloor 2011; Aragon et al. 2012), and all biographies of the selected sociologists are available in this language[12]. Using this method, we detected 910 hyperlinks among the selected biographies.

Once we obtained the network of biographical references, we calculated various network centrality coefficients for each sociologist. To measure the degree of connectivity of every biography, we computed the PageRank of each biography, that is, a measure of the global

---

[10] To review trends in countries over time, we use a single global geopolitical division as a reference. We chose the current one, because it facilitates the communication of our findings. However, this poses challenges for the historical interpretation of our data: the sociologists we group in the same country today may have lived in different countries in the past, and those we separate today may have lived in the same country in the past. To avoid confusion, we will show geographical patterns wherever possible (without mentioning countries), and when we mention country trends, we will remember that our references are to current geopolitical borders.

[11] One exception is Talcott Parsons, who died in Munich after giving a lecture, but this did not mean a definitive displacement from the United States to Germany.

[12] This decision could imply a linguistic bias, recording better the connections between biographies of English-speaking natives or those linked to English-speaking countries. However, a large part of the hyperlinks between Wikipedia biographies crosses language barriers (Aragon et al. 2012), and probably that is more common among widely disseminated historical characters (since they tend to have a detailed biographical documentation).

importance of nodes calculated recursively by putting greater weight on incoming connections from central nodes (Brin and Page 1998; Page et al. 1999). This algorithm, which is similar to that used by Internet search engines to find the most relevant websites, was used as the primary approach to 'the biographical connectivity' of each sociologist within the biographical discourse on sociologists in Wikipedia. It considers the number of mentions each biography receives (such as the in-degree coefficient) and weighs those mentions according to the degree of centrality of the biographies that referred.

Considering the 'multilingual ubiquity' of each sociologist –approached by the number of languages in which their biography is present– and their 'biographical connectivity' within the most disseminated sociologists on Wikipedia –approached through the PageRank algorithm–, we subsequently calculated the Biographical Centrality Index (BCI). The BCI (Beytía and Schobin 2020) is an indicator of a *biography's positioning*, considering both its multilingual diffusion and its connectivity in the network of references generated within a specific group of biographies[13]. For example, if Auguste Comte's biography is available in 112 languages and has a PageRank of 0.0287 due to the number of mentions coming from other biographies of sociologists, the basis for calculating its BCI is 3.2144 (112 x 0.0287). That number is then standardised to bring all biographies within a range between 0 and 1. The BCI is an indicator of how likely it is that, within this group of biographies, a sociologist will be mentioned in a search of a random sociologist in a random Wikipedia language.

When sociologists are grouped according to *ascribed factors* such as gender, birth country, and birth period, the BCI shows a vast inequality. We therefore also calculated a *relative BCI*, which controls for these factors[14]. We used that indicator to identify sociologists who have high positioning on Wikipedia in proportion to their ascribed factors. Many of them do not appear in sociology textbooks because they do not belong to the discipline's dominant discursive groups. However, highlighting some of these cases will allow us to broaden our awareness of some of the discipline's blind spots.

---

[13] Considering the number of language versions of a biography (NL) and its PageRank (PR), the *non-normalized* BCI is the multiplication of both values ($NL \times PR$). This indicator is later normalized through Feature Scaling method (more details in Beytía & Schobin 2020).

[14] The relative BCI divides the positioning of each case by the total positioning of the cases with the same gender, birth country, and historical period (assuming five periods of sociology, which will be specified later). Formally: Relative BCI = individual BCI/sum of BCI in its gender-country-period category.

Thus, we used four main indicators to assess the dissemination and centrality of sociologists in Wikipedia. The *number of languages* is an indicator of multilingual dissemination, which can be affected by general processes of popularity at a global level and does not necessarily indicate prominence within sociologists' discourse. The *PageRank* approaches each biography's connectivity within the specific biographical discourse of the selected sociologists in English Wikipedia. The *Biographical Centrality Index (BCI)* attempts to balance the aspect of cultural ubiquity –which is open to recognition from multiple discursive fields (such as politics, philosophy, or literature)– with biographical connectivity within the discourse about sociologists and can be considered as a general indicator of the biographical positioning. Finally, the *relative BCI* is an indicator of each biography's overall positioning, controlling for the ascribed factors of gender, place of birth, and historical period of birth.

We then used 'spaCy', a Python library, to analyze the text of the set of biographies in English Wikipedia. This tool allowed us to identify the organizations, people, places, events, and dates most mentioned in the biographies, through a process called 'named-entity recognition', which allows finding and classifying predefined elements in unstructured texts. We performed that process using 'en_core:web_trf', the highest precision pre-trained model offered by spaCy for processing natural language in English.

We examined the information obtained with these methods using statistical, geographic, and social network techniques, which allowed the creation of 'cultural cartographies' (Lee and Martin 2015a; 2015b), i.e. formal representations of 'the whole' cultural phenomenon analysed (social memory about sociologists on Wikipedia) although selectively reducing the information to simplify the analysis of complex patterns (e.g., showing only the biographical connections between sociologists born in different countries, or only the trajectories from the place of birth to the place of death).

General view: diffusion, networks, distributions, flows, and mobility patterns

What are the most basic macro patterns about the sociologists with biographies in ten or more languages on Wikipedia?

Using the Biographical Centrality Index (BCI) as a criterion that synthesises both the multilingual ubiquity of sociologists and their biographical connectivity within the field, Table 1 lists the most widespread sociologists by region.

Table 1. The most notable sociologists on multilingual Wikipedia by region of birth [15]

Name	Gender	Birthcountry	Birthyear	Languages	PageRank	BCI
Africa						
Alain Badiou	male	Morocco	1937	44	0,0043	0,0097
Fatima Mernissi	female	Morocco	1940	28	0,0006	0,0006
Albert Memmi	male	Tunisia	1920	22	0,0006	0,0004
Pierre Lévy	male	Tunisia	1956	15	0,0006	0,0002
Buchi Emecheta	female	Burkina Faso	1944	14	0,0006	0,0001
Nouria Benghabrit-Remaoun	female	Morocco	1952	12	0,0006	0,0001
Khadija Arib	female	Morocco	1960	12	0,0006	0,0001
Eva Illouz	female	Morocco	1961	11	0,0006	0,0000
Asia						
Georgi Plekhanov	male	Russia	1856	60	0,0045	0,0141
Ali Shariati	male	Iran	1933	38	0,0050	0,0098
Ziya Gökalp	male	Turkey	1876	34	0,0030	0,0051
Georges Gurvitch	male	Russia	1894	21	0,0030	0,0031
José Rizal	male	Philippines	1861	71	0,0006	0,0019
B. R. Ambedkar	male	India	1891	60	0,0006	0,0015
Anna Hazare	male	India	1937	54	0,0006	0,0014
Andrey Korotayev	male	Russia	1961	44	0,0006	0,0010
Michel Aflaq	male	Syria	1912	34	0,0007	0,0010
Aleksandr Dugin	male	Russia	1962	42	0,0006	0,0010
Australia						
Elton Mayo	male	Australia	1880	33	0,0010	0,0014
Raewyn Connell	female	Australia	1944	12	0,0006	0,0001
Europe						
Karl Marx	male	Germany	1818	260	0,0724	1,0000
Max Weber	male	Germany	1864	135	0,0453	0,3247
Émile Durkheim	male	France	1858	103	0,0390	0,2132
Auguste Comte	male	France	1798	112	0,0287	0,1708
Herbert Spencer	male	UK	1820	98	0,0182	0,0947

---

[15] The quantitative ordering of these authors should not be understood as a ranking of importance. It only alludes to the multilingual diffusion and discursive connectivity of these sociologists on Wikipedia. The positions are also relative since they depend strongly on the indicator observed.

Friedrich Hayek	male	Austria	1899	93	0,0159	0,0781
Pierre-Joseph Proudhon	male	France	1809	86	0,0153	0,0694
Henri Bergson	male	France	1859	114	0,0111	0,0667
Rosa Luxemburg	female	Poland	1871	151	0,0059	0,0469
Karl Popper	male	Austria	1902	100	0,0084	0,0441
North America						
John Dewey	male	USA	1859	92	0,0166	0,0808
Talcott Parsons	male	USA	1902	55	0,0126	0,0366
Jane Addams	female	USA	1860	73	0,0066	0,0251
Thorstein Veblen	male	USA	1857	50	0,0082	0,0215
George Herbert Mead	male	USA	1863	40	0,0099	0,0207
Herbert A. Simon	male	USA	1916	57	0,0064	0,0192
C. Wright Mills	male	USA	1916	47	0,0055	0,0134
Marshall McLuhan	male	Canada	1911	69	0,0032	0,0114
Erving Goffman	male	Canada	1922	44	0,0044	0,0099
Robert K. Merton	male	United States	1910	41	0,0038	0,0080
South America						
Euclides da Cunha	male	Brazil	1866	19	0,0039	0,0036
Alan García	male	Peru	1949	61	0,0011	0,0032
Ernesto Laclau	male	Argentina	1935	29	0,0021	0,0029
Sílvio Romero	male	Brazil	1851	11	0,0039	0,0020
Darcy Ribeiro	male	Brazil	1922	11	0,0039	0,0020
Fernando Lugo	male	Paraguay	1951	58	0,0006	0,0015
Fernando Henrique Cardoso	male	Brazil	1931	48	0,0007	0,0014
Rafael Caldera	male	Venezuela	1916	35	0,0006	0,0008
Gilberto Freyre	male	Brazil	1900	15	0,0011	0,0006
José Carlos Mariátegui	male	Peru	1890	28	0,0006	0,0006

In most of the indicators of multilingual dissemination and centrality, Karl Marx's biography stands out as the most influential, followed with considerable distance by other classical sociologists such as Max Weber, Émile Durkheim, Auguste Comte, and Herbert Spencer. These sociologists, along with others such as Talcott Parsons, John Ruskin, Theodor Adorno, and W. E. B. Du Bois, are also highly mentioned in the text of the biographies (appendix, Figure A1).

Interestingly, a significant segment of this list comprises influential 20th-century philosophers who carried out sociology-related works (such as John Dewey, Friedrich Hayek, Pierre-Joseph Proudhon, Henri Bergson, Rosa Luxemburg, and Karl Popper). That suggests a usual tendency to associate sociologists with social philosophers, but also that in Wikipedia the identification of sociologists seems to be quite broad.

The classification of all these people categorised as sociologists is based, on the one hand, on the propagation of biographical information in multiple languages. On the other hand, it is grounded on each biography's centrality in the network of references (or hyperlinks) among biographies. Figure 2 represents this biographical reference network, distinguishing the groups of biographies with the greatest connectivity. In it, each biography is represented by a node that expresses two attributes: the size of the node represents biographical centrality (the larger, the more centrality in the global network), and its colour symbolises participation in a highly connected group. The lines between nodes represent the references between biographies. They should be widely understood as 'relevant associations of content' (supportive or confrontational) and do not necessarily imply effective social relations. The classification of groups highlights biographies that have high shared connections and thus could be participating in a common discourse in Wikipedia (e.g., by being located at the same time, place, or social group of reference)

Figure 2. The network of biographical references between the biographies of sociologists available in ten or more languages
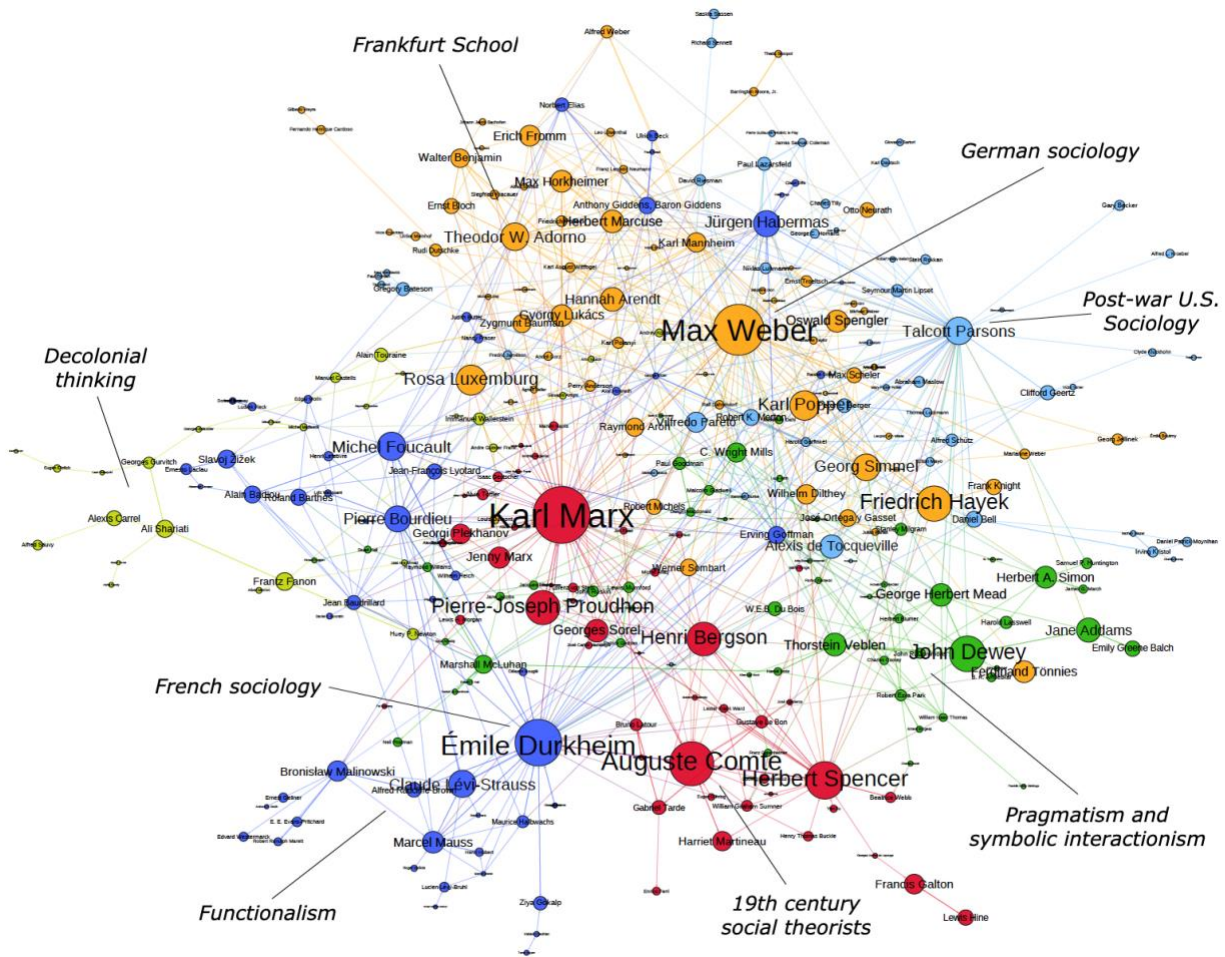
Note: We distributed the nodes using the ForceAtlas 2 algorithm. The size of each node represents its level of biographical centrality. The grouping of biographies was carried out inductively from the hyperlinks between them and using the Louvain method for community detection (Blondel et al. 2008)[16].

If some general contextualisation elements are considered –such as life span, native country, and sociologists' disciplinary focus– these discursive classifications make a considerable amount of sense. For example, the blue group comprises mainly French sociologists, although it incorporates a subgroup of functionalist researchers of different nationalities[17]. In the orange

---

[16] We selected this method because of its flexibility in testing community formation at different resolutions and its high efficiency in working with Big Data –which increases the replicability of these methods with more extensive databases of sociologists. The correct fit of the detected communities was evaluated with "ground truth information", i.e., by assessing the coincidence between the detected communities and real groups of highly linked sociologists.

[17] This makes sense considering, for example, the centrality of Radcliffe-Brown in British functionalism and its fundamental role in the dissemination of Durkheim's work in English-speaking countries (Lévi-Strauss 1945).

group, most German sociologists are included, and a more specific agglomeration is generated in the first generation of the Frankfurt School. Other groups in the sociological tradition can also be identified: a cluster of North American pragmatists and symbolic interactionists (marked in green), robust connectivity among nineteenth-century social theorists (in red), a group of leading sociologists in the United States after World War II (in light blue), and a small conglomerate of influential intellectuals in postcolonial studies (in yellow).

Figure 3 displays, in four generations of sociologists (which will be explained in the next section), the geographical distribution of their place of birth (heat spots) and their biographical links (lines from one place of birth to another). It shows that Western Europe has been the centre of disciplinary links from the beginning, and there has also been a relevant historical centre in the East of the United States. It also shows significant biographical links between these two regions, suggesting that they have not operated in isolation. Connectivity between the two centres has been extensive, at least since the generation of sociologists born between 1881 and 1910, which was subject to massive movements associated with the two world wars.

Figure 3. Birthplace of the most widespread sociologists and their biographical connections in four generations of sociologists
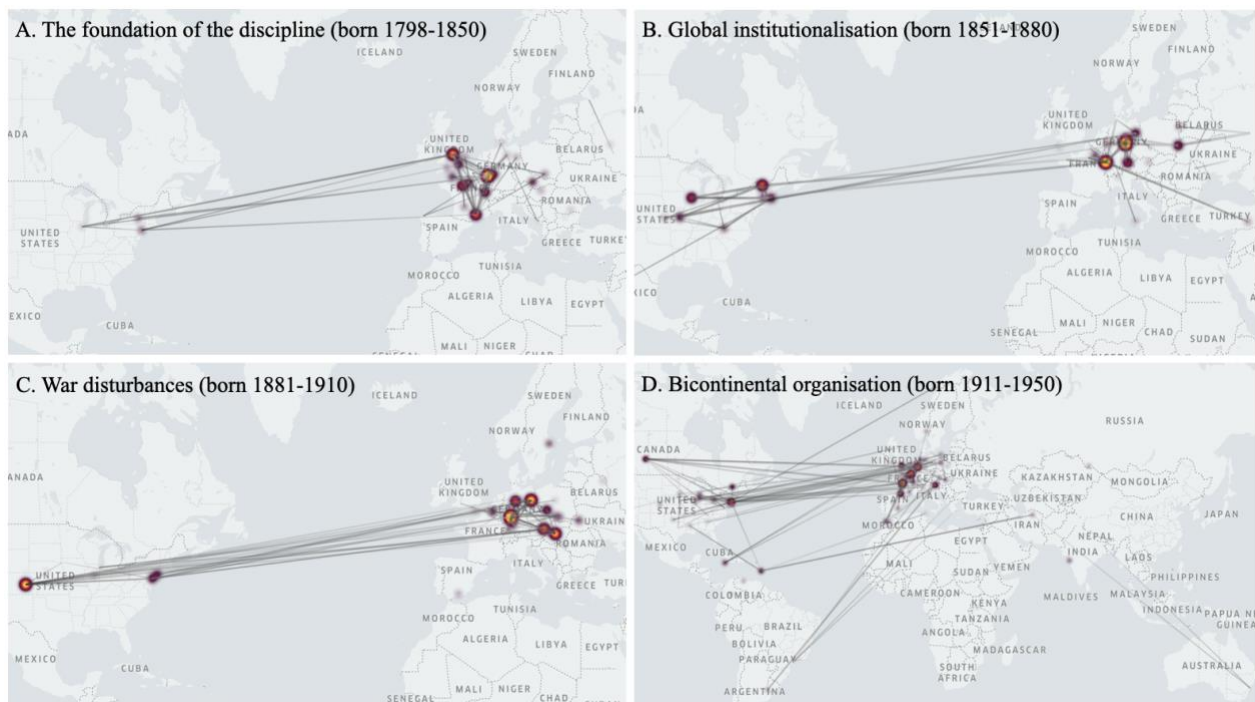
Note: The hotspots highlight the most crowded birthplaces of notable sociologists on Wikipedia. The black lines between two places represent biographical links between sociologists who were born in those places.

The overall concentration of prominent sociologists in certain countries and continents is shown most clearly in Figure 4. It displays the number of recognised sociologists born in the most typical countries (by current geopolitical divisions), their sum of biographies (sites) considering all languages, their aggregate PageRank, and their cumulative BCI. As can be seen, the United States, Germany, and France are the countries that have had the greatest multilingual impact on the discipline, followed by the United Kingdom, Poland, Austria, Russia, Spain, and Canada. On a continental level, after Europe and North America, there has been relevant participation in Asia and South America, while Africa and Australia have remained in a relatively marginalised position within the discipline (for more details on international comparisons, see Table A7 in the appendix).

Figure 4. The proportion of the top 20 countries in the group of notable sociologists
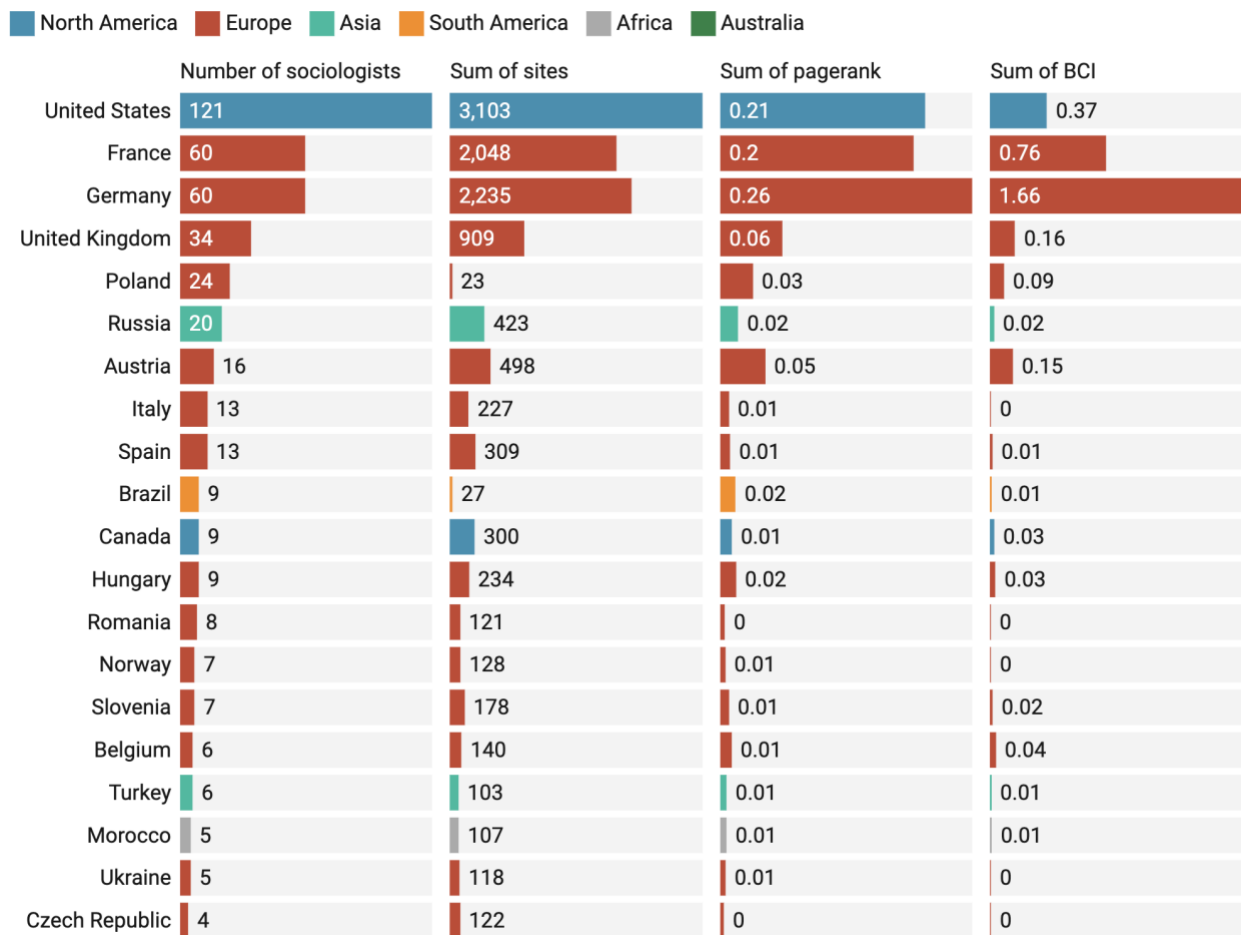
Note: The colour of each country represents its continent. The country of every sociologist was calculated according to the place of birth and the current geopolitical boundaries.

From a historical perspective, however, the integration of continents has varied as represented on Wikipedia. As Figure 5 shows, the first world-renowned sociologists were born in Europe, although soon (from the generation born in the 1910s) the discipline expanded into Asia and North America. The integration of South America into this group of sociologists has been maintained with some stability since the generation born in the 1850s. In contrast, Africa's inclusion began very late (roughly in the generation born in the 1920s), while Australia has generally had minimal integration, albeit with a small boom in the 1880s generation.

Figure 5. Historical variation in the proportion of sociologists by continent and decade of birth
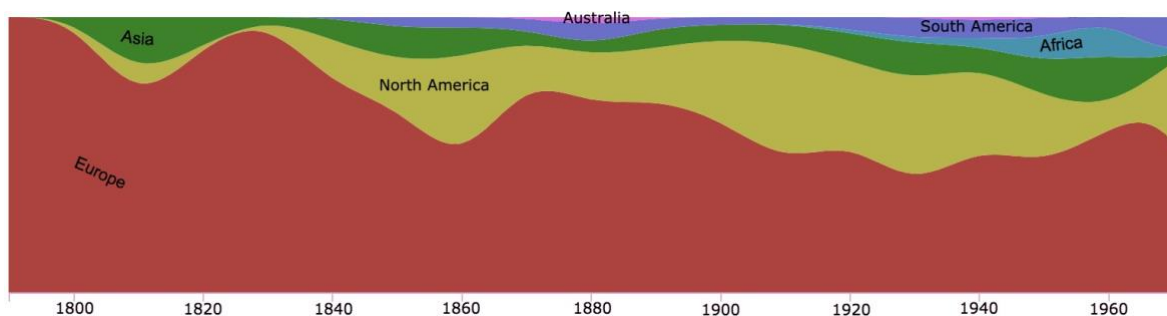

The relative participation of countries and continents can also be appreciated in biographical reference flows. Figure 6 shows from which countries these references come (left part of the graph) and which countries receive them (right part of the graph). The biographical reference flows between countries are understood here as the sum of the reference flows between the sociologists born in those countries. As can be seen, the main countries that receive biographical references (Germany, the United States, and France) are also those that generate the most references to other sociologists, and those with the most sociologists (see Figure 4). They are also the most mentioned in the text of this set of biographies (see appendix, Figure A6). In other words, these countries *quantitatively*, *relationally,* and *discursively* dominate the field of recognised sociologists. The United States is the country that provides the most biographical references, which is explained by its high number of renowned sociologists (121, as shown in Table A7). In contrast, Germany receives more biographical references, with about half of the representatives (60 sociologists).

Figure 6. Flows of biographical relationships between countries (by place of birth)
Source country   *Target country*
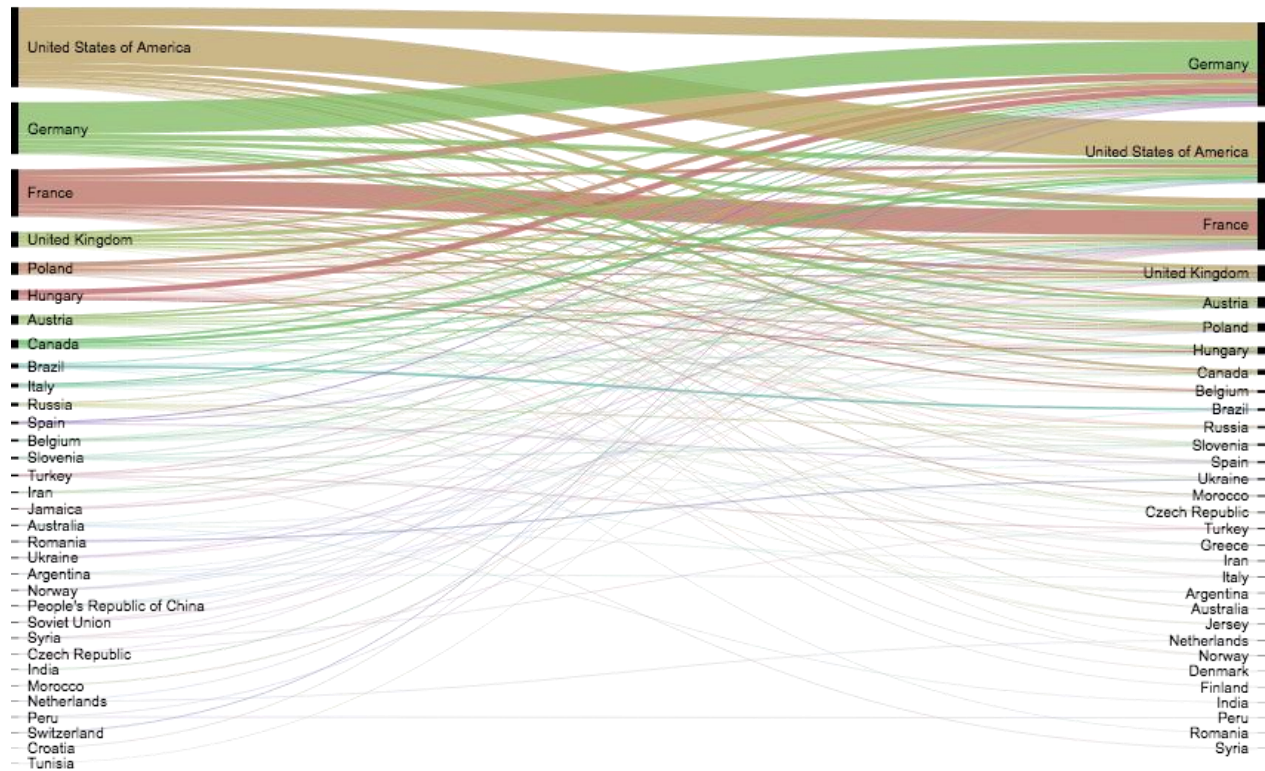
We can observe an interesting *self-reference* in the most influential nations: approximately half of the references coming from U.S. biographies are directed towards U.S. biographies, and this proportion of self-reference is similar in Germany and France. This could be interpreted as an indicator of autonomy of the discipline in those countries. In contrast, there are countries –such as Poland, Hungary, or Canada– where there is not a high degree of self-reference: Poland and Hungary have greater links to Germany than to their own countries, and the same happens from Canada to the United States. In those cases, the discipline seems to be more hetero-referential and dependent on other countries' development and disciplinary opportunities.

Finally, Figures 7 and 8 show the trajectories from place of birth to place of death in the four countries with the greatest historical influence: Germany, France, the United States, and the United Kingdom. In them, the sociologists' places of death are highlighted with black dots, their places of birth with grey dots, and a line between them indicates the birth-death trajectory. This information can be properly interpreted in connection with the leading institutional affiliations of these sociologists, both in terms of education and employment (see Table 2).

France's pattern is completely centralised: sociologists who move from their place of origin tend to settle and die in Paris (Figure 7). A closer analysis of this trend shows that notable sociologists born in France have tended to have affiliations in the Université de Paris (Sorbonne), the École Normale Supérieure, and the Ecole des Hautes Etudes en Sciences Sociales. Remarkably, several of them studied at the same school: the Lycée Louis-le-Grand (see Table 2). In the United Kingdom, there is a more moderate centralisation, which is established mainly in the south (London and its surroundings), probably because of the strong connection of British sociologists with the London School of Economics, Cambridge, Oxford, and the University of London. Similarly, a moderate centralisation of the discipline has developed in the United States, with the East Coast as its pole of attraction –considering cities such as Boston, New York, Philadelphia, and Washington (Figure 8)–. This could be explained by the fact that many renowned American sociologists were affiliated with Ivy League universities –particularly Harvard, Columbia, and Yale universities–. Germany's case is the most decentralised and only a few death destinations such as Berlin, Heildelberg, and Frankfurt stand out slightly (Figure 7). Again, this has to do with the more typical institutional affiliations. Many prominent sociological figures studied or worked at the Humboldt University of Berlin (e.g., Marx, Weber, Simmel, and Sombart), the Heildelberg University, and the Institut für Sozialforschung at the Goethe Universität Frankfurt (see Table 2).

From a wider institutional perspective, universities are typically the dominant institutions in the biographical narrative of these sociologists (appendix, Figure A5). But other relevant institutions also have a key place in their life narratives, especially professional and civic associations (ASA, NAACP, SDS), governmental organizations (states, parliaments, international organizations), and political parties.

Figure 7. Birth-death displacement in Germany, France, and the United Kingdom

Figure 8. Birth-Death displacement in the United States

Note: The black dots highlight death places, while the grey dots mark the places of birth. A line symbolises the birth-death trajectory.

Table 2. Top 20 institutions associated with the profiles of the sociologists with biographies in 10 or more languages

Rank	Education (87% with information)		Employment (72% with information)		Total	
	Institution	Mentions	Institution	Mentions	Institution	Mentions
1	Harvard University	51	University of Chicago	44	Harvard University	93
2	Columbia University	38	Harvard University	42	Columbia University	74
3	University of Paris	33	Columbia University	36	University of Paris	54
4	University of Chicago	32	University of California, Berkeley	25	University of Chicago	53
5	Humboldt University of Berlin	25	London School of Economics and Political Science	24	London School of Economics and Political Science	45
6	London School of Economics and Political Science	21	University of Paris	21	Humboldt University of Berlin	39
7	University of Vienna	21	School for Advanced Studies in the Social Sciences	20	University of California, Berkeley	38
8	École Normale Supérieure	19	Yale University	15	University of Vienna	34
9	Heidelberg University	17	Humboldt University of Berlin	14	Yale University	30
10	Yale University	15	Stanford University	14	School for Advanced Studies in the Social Sciences	29
11	University of California, Berkeley	13	Goethe University Frankfurt	13	Heidelberg University	26
12	Eötvös Loránd University	12	University of Vienna	13	Goethe University Frankfurt	25
13	Goethe University Frankfurt	12	Collège de France	12	University of Cambridge	23
14	Leipzig University	12	University of Cambridge	12	University of Pennsylvania	21
15	Lycée Louis-le-Grand	11	New York University	11	École Normale Supérieure	19
16	University of Cambridge	11	University of Pennsylvania	11	New York University	18
17	Faculty of Arts of Paris	10	École pratique des hautes études	10	Leipzig University	16
18	University of Pennsylvania	10	New School	10	Ludwig Maximilian University of Munich	15
19	City College of New York	9	University of Michigan	10	University of Warsaw	14
20	Ludwig Maximilian University of Munich	9	University of Oxford	10	Brandeis University	13

Five generations in the remembered history of sociologists on Wikipedia

To obtain reflexivity from the data collected in Wikipedia, the first step was to observe the global aggregated patterns (structures of emergent properties). Right after that, an historical analysis of this information is essential. The historical analysis allows us to face unconscious phenomena in sociology since it gives an account of the circumstances of production of the sociological work that may have been forgotten. For that reason, in this section we will describe the basic structural patterns of the historical evolution of prominent sociologists according to the multilingual record of Wikipedia. That will be done from a *generational perspective*: differentiating sets of sociologists born on similar dates since they shared relevant circumstances and a way of organising sociology at the world level. We will also highlight some sociologists who are not central in absolute terms within each generation but obtain a high relative positioning (relative BCI) when we control for ascribed factors that strongly skew the record (gender and birth country). Our intention with this is to raise awareness of sociologists who had not been born women or in peripheral places, might have had greater centrality in the discipline.

To differentiate these generations based on Wikipedia's information, we carried out an inductive (data-driven) analysis grounded upon *six structural aspects*. We ordered the sociologists according to their year of birth. We then looked for temporal disruptions in 'emergent properties' –characteristics observable in the aggregate of biographies and not in particular cases, such as tendencies, patterns, distributions and/or networks–. We specifically reviewed 1) the number of globally recognised sociologists, 2) their gender composition, 3) their geographical distribution, 4) their trajectories from a place of birth to a place of death, 5) their biographical links, and 6) their differentiation into groups with a greater degree of mutual references. The *joint variation* of several of these factors was our strongest indication of structural breaks between generations. To systematically identify these underlying disruptions, we developed a *simulation of historical advance* that makes it possible to visualise the evolution of these factors simultaneously[18].

We determined the limits of these generations and each sociologist's belonging to one of them by the *date of birth* and not regarding the moment in which they developed their main works

---

[18] This web application is available at: https://osf.io/9t6au/. The tool allows to visualise the historical evolution of the biographies studied. One can select variables (e.g., biographical links, or trajectories from a birthplace to a deathplace) and then observe, like a video, the patterns across time and space.

28

associated with the discipline. Thus, we distinguished *five generations*: the foundation of the discipline (sociologists born between 1798-1850), global institutionalisation (1851-1880), war disturbances (1881-1910), bicontinental organisation (1911-1950), and construction of a contemporary narrative (1951-present). Each of these stages will be reviewed below, highlighting the particularities that justify this division of generations.

First generation: the foundation of the discipline (born 1798 - 1850).

In this stage sociology was conceived in Europe as an academic area, albeit with rather diffuse disciplinary limits. The register of widely known sociologists is still relatively low and most of them were at the same time central references in other disciplines, such as philosophy, literature, economics, psychology, or anthropology (see Table A2). Some of the most characteristic and widespread sociologists born in this period are Auguste Comte (1798), Alexis de Tocqueville (1805), Pierre-Joseph Proudhon (1809), Karl Marx (1818), John Ruskin (1819), Herbert Spencer (1820), Francis Galton (1822), Wilhelm Dilthey (1833), Gabriel Tarde (1843) and Vilfredo Pareto (1848). An important portion of them is part of a network of widely connected biographical discourses, which we have previously called 'nineteenth-century social theorists' (see Figure 2).

In this period, sociology seems to be an almost uniquely European phenomenon, specifically in France, Germany, and the United Kingdom (see Figure 3-A). For example, it does not seem to be a relevant discipline in the United States, with only a few references (such as Lester Frank Ward, Lewis H. Morgan, and William Graham Sumner) that are biographically linked to a greater extent with Europeans than with the Americans themselves. Controlling, however, by gender and birth country (relative BCI), figures such as Harriet Martineau (United Kingdom), Max Nordau (Hungary), Johann Jakob Bachofen (Switzerland) and Valtazar Bogišić (Croatia) stand out.

If we look at birth-death mobility, Paris is the most explicit focus of (internal) migration: French sociologists tended to settle and die in France's capital. Less clearly, the UK's most renowned sociologists began to concentrate in the south (London and its surroundings). Germany, on the other hand, is characterised by complete territorial decentralisation. These three levels of geographical concentration –high in France, moderate in the United Kingdom, and low in Germany– have been maintained in these countries in all subsequent periods (see Figures 7 and 8),

which suggests that these are patterns rooted in the institutionalisation of the sociology of these nations.

Second generation: global institutionalization (born 1851 - 1880).

This generation institutionalised sociology worldwide as an autonomous academic discipline, although there were still no close collaborative links between nations. There is a significant growth in the number of notable sociologists, and within nations, the biographical links between them are densified, which can be considered as an indicator of greater mutual contact and disciplinary institutionalisation. For the first time, groups of sociologists whose biographies are especially connected at the national level are emerging (in France, Germany and the United States).

In this period, the classical systematisers of the discipline were born, who played a determining role in the foundation of sociology schools or associations, and many cases already considered themselves as sociologists. Some typical representatives of this stage are Émile Durkheim (1858), Ferdinand Tönnies (1855), Thorstein Veblen (1857), Georg Simmel (1858), George Herbert Mead (1863), Werner Sombart (1863), Max Weber (1864), Marcel Mauss (1872), Max Scheler (1874) and Maurice Halbwachs (1877). To this list, we could add other sociologists who did not have a dominant gender or birth country but still have a high relative positioning. Such is the case of Beatrice Webb (United Kingdom), Ziya Gökalp (Turkey), Edvard Westermarck (Finland), and Elton Mayo (Australia).

At least two formal factors suggest a process of global institutionalisation of the discipline. First, globally recognised sociologists appear for the first time in South America (Silvio Romero and Euclides da Cunha in Brazil, Luis Alberto Riart in Paraguay, and Laureano Vallenilla in Venezuela), East Asia (Yan Fu and Zahng Binglin in China, and José Rizal in the Philippines) and Australia (Elton Mayo). Second, a much more relevant mass of research began to exist in the United States (this generation was the most active in the American Sociological Association's foundation in 1905).

Considering both birth-death mobility and the level of biographical referencing, there is a relative disconnection between the European sociology –which is still the undisputed centre of the discipline– and that of the United States. On the one hand, there are few birth-death displacements between the two regions (the exceptions are Mary Parker Follet and Emily Greene Balch, who

moved from the United States to the United Kingdom, and Franz Oppenheimer who was born in Berlin and died in Los Angeles). On the other hand, there was a relevant degree of communication between both regions, but the biographical connections are relatively low compared to the subsequent periods.

Third generation: war disturbances (born 1881 - 1910).

When analyzing the text of the biographies, the most prominent events are wars or crises associated with them: World War I (154 mentions), World War II (124), the Holocaust (53), and the Cold War (41). Wars appear to be biographically crucial events, which decisively affected the life course of these sociologists. They are mentioned about five times more than other relevant disruptive events, such as revolutions or social crises (see appendix, Figure A4).

The sociologists born in this period belong to a generation extraordinarily affected by the two world wars. At the beginning of the first war, they were between 4 and 33 years old, while in the second they were between 29 and 58 years old. Without doubt, this experience profoundly marked their life trajectories and transformed their work practice. This generation was subject to important geographic and social movements reflected in mobility patterns and biographical links. Under these political circumstances, only one group of broadly biographically interconnected sociologists emerges (which is the densest in the database): the first Frankfurt School (Adorno, Horkheimer, Benjamin, Fromm, Marcuse, Bloch, Pollock, Neumann, Wittfogel, Löwenthal, Kracauer), also linked to other intellectuals of Jewish origin (Lukács, Mannheim, Elias, Arendt). Controlling for the relevance of birthplace and gender, we can mention other relevant sociologists such as Harold Innis (Canada), Gilberto Freyre (Brazil) and José Carlos Mariátegui (Peru).

In Europe, the most remarkable change occurs in the pattern of birth-death mobility. In general, there is a process of emptying of Central Europe intellectuals. No sociologist born in this period in Eastern Germany, Poland, or Ukraine dies in his/her country of origin. The trajectories from a place of birth to a place of death occurred in two ways. a) Intra-European displacement: developed primarily in a north-south direction. For example, Walter Benjamin died when he moved from Germany to Spain (although he intended to reach the United States), while Theodor Adorno and Erich Fromm died in Switzerland. An exception to this internal pattern is Karl Mannheim, who moved westward (to the United Kingdom). b) Transatlantic displacement: it was much more

common than intra-European mobility, and was developed primarily by intellectuals born in Germany, Austria, and Poland. That is the case, for example, of Hannah Arendt, Bronisław Malinowski, Karl Polanyi, and Alfred Schütz, who mainly settled in the United States (except for Polanyi, who died in Canada).

The biographical connections between intellectuals born in Europe and the United States exploit in this period, probably due to the transatlantic movements. There is a relevant increase in the intercontinental connection of the discipline, which could be an unintended consequence of the world wars. In this period, moreover, the rate of recognised sociologists born per year increased progressively.

Fourth generation: bicontinental organisation (born 1911-1950).

This generation is the most active in the development of post-war sociology and was characterised by a new foundational impulse, in the sense of critically assessing the theoretical approaches and dominant methodologies of the discipline. After the strong waves of displacement from Europe to the United States, for the first time, the United States became a worldwide sociological centre, developed mainly on the East coast of the country (Boston, New York, Philadelphia, and Washington). The development of this new centre also affects sociology's massiveness: this is the period with more sociologists disseminated in multiple languages. Moreover, most of these sociologists were very active between 1950 and 1980 –particularly in the 1960s– that is, in the period most referenced in Wikipedia's biographies of notable sociologists (appendix, Figures A2-A3).

Two indicators can show the greater centrality and independence that the United States acquired in this period. On the one hand, birth-death mobility began to be directed to the East coast. That suggests a new attraction of sociologists towards specific geographical areas that could compete with the European disciplinary centres from now on. On the other hand, the American network of biographical references became significantly denser, indicating greater disciplinary autonomy at the national level. The increase in mobility and biographical links within the United States was accompanied by an inverse trend towards Europe. As the density of American relations grew, birth-death displacements and biographical references from the United States to Europe

decreased. This lesser relative linkage can be interpreted as an expression of the same process, in which the sociology of the United States gained greater autonomy and sustainability.

As a result, world sociology began to have a decidedly bicontinental focus (see Figure 3). A strong focus in the United States (mainly at the East Coast) joined the traditional Western European centre (which was established primarily in Germany and France). In Europe, the most characteristic sociologists of this generation were Michel Foucault, Jürgen Habermas, Pierre Bourdieu, Zygmunt Bauman, and Anthony Giddens, while in the United States, C. Wright Mills, Daniel Bell, and Clifford Geertz, plus some Canadian intellectuals who migrated to that country (Marshall McLuhan and Erving Goffman) stand out. Sociologists who are relatively highly positioned, despite not having a dominant gender or birth country, are Saskia Sassen (Netherlands), Agnes Heller (Hungary), Stuart Hall (Jamaica) and Fatima Mernisi (Morocco). The first two spent a significant part of their working lives in the United States, while Stuart Hall in the United Kingdom. Mernissi completed her university studies in France and the United States.

The general overview of biographical relationships (Figure 2) hides the specific grouping patterns of this generation, which is probably the most influential in contemporary sociology[19]. These relations are clarified in Figure 9. A group of theoreticians densely associated obtains the highest centrality and biographical diffusion, led by Foucault, Habermas, and Bourdieu, and including Giddens, Goffman, Badiou, Luhmann, Beck, and Offe.

---

[19] This suggests the high number of times the 1950s, 1960s and 1970s (and the specific years of that period) are mentioned in the biographies (appendix, Figures A2-A3).

Figure 9. **Biographical relationships among fourth-generation sociologists (born between 1911 and 1950)**

Note: sociologists with the greatest biographical connections are marked in red, and the second group with highest connectivity is highlighted in pink. The larger the size of each node, the greater the centrality of the biography.

In France, Paris remains the most attractive city for sociologists' migration, although it is primarily French mobility (e.g.: Michel Foucault, Pierre Bourdieu, and Jean Baudrillard). In Germany, there is still a tendency towards concentration in the Western are, that is reflected in the high birth-death mobility from Berlin to other German cities or countries in the southwest.

Fifth generation: construction of a contemporary narrative (born 1951 - to date).

This period is characterised by sociologists' low global visibility, probably since this generation is still in an early or intermediate stage within the cycle of dissemination of their works and professional experiences. The number of sociologists recognised in multiple languages is notably less than in the previous period, and this exclusive group also has a low discursive centrality within the network of biographies. Some of them have excelled in the intellectual field –such as Judith Butler (61 languages), Naomi Klein (46), Andrey Korotayev (44), and Aleksandr Dugin (42)–,

while others have excelled in the political field –such as Fernando Lugo (58 languages), George Papandreou (56), and Iveta Radičová (47)–.

Despite the low overall visibility of this generation, we can recognise some relevant patterns so far. First, we see a decisive increase in female participation in the most widespread sociologists group, especially since those born in the 1970s (see Figure 1). Correlatively, in the early decades of this generation (1950 - 1960), it was the first time that sociologists born in North Africa (Morocco and Tunisia) acquired worldwide relevant participation (see Figure 5). Finally, the bicontinental pattern in the distribution of disciplinary recognition can still be observed: the main centres of world-renowned sociologists remain Western Europe and the Eastern United States.

Since this generation's sociologists still have little visibility, this stage's patterns are just preliminary generalisations. For example, only five sociologists –Loïc Wacquant, Malcolm Gladwell, Arno Tausch, John Bellamy Foster, and Judith Butler– have been mentioned in other biographies of renowned sociologists. Additionally, in this generation it is not yet possible to observe patterns in birth-death mobility, given that a few sociologists have died and none of them had a migration from their place of birth.

Intergenerational trends

Except for the last stage –which is still an emerging period–, there is a clear tendency towards the increase of *disciplinary massiveness*: with generations' passing, there is a greater frequency of globally recognised sociologists. However, this does not imply that the new generations are more influential than the previous ones. There is also an evident impulse towards the *valuation of the classical*: the older the sociologists are, the more their biographies tend to be present in a greater number of languages and to have more connectivity with biographies of recognised sociologists, which makes them on average more influential for the discipline (Figure 10).

Figure 10. Frequency of sociologists and average of the Biographical Centrality Index (BCI) in the five generations
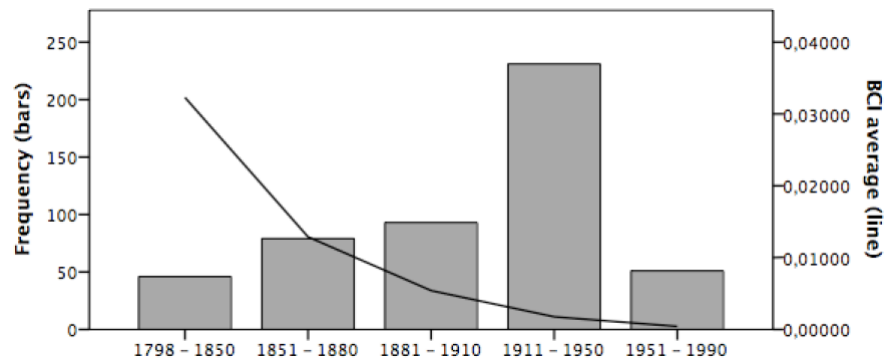
Note: The bars represent the frequency of recognised sociologists, and the line shows the average of Biographical Centrality Index (BCI) for each period.

Since gender and birth country are factors that strongly structure this record of biographies and have also affected the configuration of sociology departments, editorial boards and textbooks. (Demeter and Toth 2020; Korom 2020; Cárdenas 2021), we also examine the combination between the generations and those ascribed factors. Table 3 specifies, from an intersectional logic, which combinations of birth country, gender, and generation (g) accumulate the sociologists with the best positioning in Wikipedia (BCI Sum). As can be seen, the classical (first and second) generations of male sociologists born in Germany and France are the best positioned globally, followed by the group of US-born male sociologists of the second generation. Interestingly, in the third generation (war disturbances period) stands out the intersection of male sociologists born in Austria, Belgium, Hungary, and Poland.

Table 3. The top twenty intersectional combinations of notable sociologists with the best positioning (BCI sum)

Tipology	Number	BCI Mean	BCI Sum
Germany-Male-g1	9	,1130	1,0166
Germany-Male-g2	11	,0392	,4310
France-Male-g2	12	,0261	,3137
France-Male-g1	11	,0282	,3100
United States-Male-g2	13	,0112	,1451
Austria-Male-g3	7	,0200	,1400
Germany-Male-g3	17	,0071	,1211
France-Male-g4	25	,0048	,1192

United Kingdom-Male-g1	8	,0144	,1153
United States-Male-g4	58	,0017	,1006
United States-Male-g3	22	,0034	,0738
Poland-Female-g2	1	,0469	,0469
Germany-Male-g4	14	,0031	,0437
United States -Female-g2	6	,0061	,0364
Belgium-Male-g3	2	,0176	,0352
Hungary-Male-g3	2	,0159	,0317
Poland-Male-g3	5	,0049	,0243
Germany-Female-g3	1	,0241	,0241
Canada-Male-g4	4	,0057	,0226
Germany-Female-g1	1	,0175	,0175

Conclusion

This article aimed to show some of the research opportunities opened by what we have called 'digital reflexive sociology': the use of digital methods and Big Data to systematically reflect on the social and historical circumstances of sociologists and sociological thinking. Particularly, we studied biographical pattens of sociologists with high global recognition on Wikipedia, understood as those who have their biographies in ten or more languages. We used computational methods to extract data on these sociologists' birth, gender, and death and trace the network of references (or hyperlinks) between their biographies. That information was used, in the manner of prosopography or multiple career-line analysis, to describe the *structural narrative* on the history of sociologists on Wikipedia. We constructed 'cultural cartographies' that exhibited 1) the variation in the frequency of globally recognised sociologists, 2) trends in gender composition, 3) geographical concentrations of the discipline, 4) disciplinary differentiation in clusters with high biographical connectivity, 5) biographical reference flows between countries, and 6) birth-death mobility trajectories. We also developed a web simulation of these aspects to sketch the macro-historical narrative of sociologists in Wikipedia, distinguishing five generations of people who shared similar collective circumstances (such as a degree of disciplinary institutionalisation, a specific territorial organisation, a structure of birth-death mobility, or a particular configuration of biographical clusters).

Our analysis provides novel evidence. Previous studies have used digital information and computational methods to analyse prominent sociologists. There are studies on a global scale (Demeter and Toth 2020; Cárdenas 2021; Korom 2020) and on the evolution of the discipline over the last 50 years (Korom 2020), but none of them studied sociology globally and through its entire history. There is also research on the representation of sociology in Wikipedia, but it focused only on the selection biases of English-language biographies in the current US academy (Adams, Brückner, and Naslund 2019). In contrast, we offer *a global, full-history analysis of sociologists according to Wikipedia's multilingual record*. We argue that the exploration of the circumstances of the best-known sociologists in this encyclopaedia can contribute to the reflexivity of sociology, to the extent that sociologists examine these results and confront them with their own knowledge, opinions, and beliefs on the discipline.

The most relevant findings of our analysis of Wikipedia's multilingual social memory are the following:

1. We confirmed the disciplinary centrality of some classic authors –such as Marx, Weber, Durkheim, Comte, or Spencer– as well as of some theoreticians of the second half of the XX century –such as Foucault, Habermas, and Bourdieu–. Along with them, an important role of some social philosophers stands out. That could indicate a biographical closeness between sociologists and social philosophers, but also that the classification of sociologists in Wikipedia is rather blurred.

2. We observed a high geographical concentration of the discipline in three countries – Germany, France, and the United States. These nations, moreover, seem to be the ones with the highest degree of disciplinary autonomy, since they have the highest percentages of biographical self-reference. This over-representation of the Global North is consistent with what has been reported by studies on the composition of the top sociology departments and the editorial boards of journals worldwide (Demeter and Toth 2020; Cárdenas 2021).

3. These three countries have very different patterns in their disciplinary geographic distribution. In France, sociology seems to have been highly centralised (in Paris), in the United States moderately centralised (on the East coast), and in Germany very decentralised.

4. We found extensive clusters of connected biographies, as in the case of German and French sociology. On a more specific level, six strong biographical clusters turn out: nineteenth-

century social theorists, functionalism, pragmatism and symbolic interactionism, the Frankfurt School, the main sociologists in the United States after World War II, and some intellectuals connected to post-colonialism.

5. From the formal patterns of the history of sociologists in Wikipedia, we can distinguish five generations of sociologists: those who participated in the foundation of the discipline (born 1798-1850), in its world institutionalisation (born 1851-1880), in the war disturbances (born 1880-1910), in the bicontinental organisation (born 1911-1950), and in the construction of a contemporary narrative (born 1951 - to date). From an intersectional logic, the best positioned sociologists are German and French men of the first and second generation.

6. The two world wars are completely disruptive phenomena in the formal narrative of these sociologists and the most referenced events in their biographies. Over three decades, European sociologists migrated massively, especially from Central Europe to North America. Only after this process did the United States achieve a significant level of disciplinary autonomy, suggesting that the wars were relevant to its internal development and the bicontinental configuration of global sociology.

7. There is a trend towards greater centrality and multilingual diffusion of the first generations' sociologists ('valuation of the classical'). In contrast, the new generations tend to register a greater number of globally recognised sociologists ('disciplinary massiveness').

8. Despite the huge concentration of the discipline according to gender and place of birth, there is a growing trend towards greater inclusion and disciplinary diversity. On the one hand, sociology began in Europe and successively incorporated notable intellectuals from Asia, North America, South America, Australia, and Africa (in that order). On the other hand, the proportion of globally recognised female sociologists has been growing since the generation born in the 1970s.

Taken together, these patterns suggest that sociology should be fully aware that its historical narrative involves a *high disciplinary concentration*. From a *geographical* perspective, only three countries have concentrated the history of prominent sociologists recorded in Wikipedia –in terms of quantity (number of sociologists recognized worldwide), relations (centrality in the network of hyperlinks), and discourse (number of times they are mentioned in the biographies)–. Regarding

gender, 86.2% of sociologists recognised in various languages are men. Furthermore, only a few *schools of thought* have achieved a level of massiveness and grouping that allows them to be clearly identified in this sample of sociologists: the Frankfurt School, functionalism, and symbolic interactionism. Groups of prominent sociologists who belong to this record but not to any of these concentrations (of gender, place of birth, or schools of thought) tend to participate in a position of 'exclusionary inclusion' (Agamben 1998; Decoteau 2013): although they are not outside the group, they are left aside in the formal configuration of this social memory[20].

Certainly, according to the Wikipedia record, both territorial and gender concentrations have decreased along with the global spread of sociology. However, both concentrations continue to exist at very high levels. For example, Demeter and Toth (2020) reported that the world's top 100 universities (according to the QS rankings) have high biases towards hiring male sociologists and from countries in the Global North. Moreover, the trend we have found towards a greater 'valuation of the classical' implies that the recent growth in diversity does not immediately have a high discursive weight.

The latter does not mean that these discursive trends are static. The Wikipedia record is in constant motion and the world-wide organisation of sociology can be structurally affected by social phenomena external to the discipline –as we illustrated with the world wars. Therefore, it is possible to imagine future structural changes in the world organisation of sociology and its social memory structure, even though there is a clear impulse towards stabilising a 'sociological canon' concentrated on men from Western Europe and the United States. Nevertheless, reconstructions of social memory also imply an effort of disciplinary self-observation that allows sociology to critically incorporate its non-dominant discourses. For this reason, we have highlighted in our historical account some sociologists who might be central to the discipline if it were not almost exclusively dominated by Western European men. Such is the case, among others, of Harriet Martineau, Max Nordau, Valtazar Bogišić, Beatrice Webb, Ziya Gökalp, Elton Mayo, Gilberto Freyre, José Carlos Mariátegui, Saskia Sassen, Agnes Heller, Stuart Hall, and Fatima Mernisi.

The clarity and diversity of these findings make it possible to visualise the methodological and informative potential of digital reflexive sociology. With the help of 'natively digital' methods (Rogers 2013) –such as web scraping, hyperlink tracking, and automatic text analysis– we were

---

[20] More than half (269) of these sociologists are not referenced by other notable sociologist's biography, so they are immediately outside the clusters identified within the discipline.

able to obtain a global and multicultural list of highly disseminated sociologists systematically associated with spatial, temporal, relational, and personal attribute information. The development of this *global, multicultural, historical, and associative perspective* would be extremely difficult to achieve without the current digital repositories and computational methods that make it possible to obtain and relate such data.

Perhaps that may explain why the most influential studies in reflexive sociology have significantly narrowed their research focus. Some have studied key authors (Aron 2017; 2018; Coser 1971), the foundational period of the discipline (Nisbet 1993), specific national cases (Lévi-Strauss 1945; Lipset 1994; Halsey 2004; Heilbron 2015), national cases in particular periods (Käsler 2013; Lüschen 2013), or even the reception that specific countries have of sociologists from other specific countries (Ollion and Abbott 2016). *Digital* reflexive sociology, by contrast, seems to be called upon to develop an opposite but complementary approach: to progress in a global and multicultural perspective, which has no technical impediments to observing information of millions of sociologists, spread and linked in multiple epochs, countries, and cultures. This big picture is critical for detecting large-scale biases in sociology and could also contribute to the contextualization of more specific studies on the conditions of sociological work (e.g., Warczok and Beyer 2021).

In terms of its breadth, perhaps the closest analysis to this approach was obtained by analysing the most influential books of sociology in the twentieth century –according to the vote of the members of the *International Sociological Association* in 1997– (Lamo de Espinosa 2001). However, this interesting analysis has high validity and representativeness limitations[21] and is restricted to a specific period (20th century) and aggregate trends (excluding associative and spatial patterns). It is also a comparatively less feasible approach, as it was made possible only by a highly contingent situation: the impulse of critical evaluation that an international organisation had on the verge of a new century.

Besides its scope and feasibility, digital reflexive sociology's innovation has to do with the *type of information* it has access. The case of Wikipedia illustrates well how some current digital repositories –wikis, search engines, social networks sites, collaborative software, and other

---

[21] It was based on a survey of association members in 1997 and with a response rate of 16%. More details on this survey on ISA "Books of the XX Century": https://www.isa-sociology.org//en/about-isa/history-of-isa/books-of-the-xx-century (reviewed 9 September 2020).

platforms– are continually storing traces of social memory. Instead of being forms of record structured in individual consciousness (or in a sum of them), they often show results iteratively constructed by a network of actants (human and nonhuman agents) that acts and interacts under circumstances established by a specific socio-technical framework. Perhaps, we could understand this type of information as 'collective intelligence', that is, an intelligence 'distributed everywhere, constantly valued, coordinated, and mobilised in real-time' (Lévy 1997). In any case, as a collective phenomenon, it has a greater outreach to obtain information from hard-to-reach groups and mitigates some biases stemming from individual research decisions (Bonabeau 2009). For example, collective information sources are less likely to incorporate perspective biases (e.g., due to ethnocentrism, or disciplinary sub-specialisation), because they are based on a multiplicity of points of view, and therefore, of beliefs, interests, motives, desires, affections, and dispositions, which in the case of Wikipedia are also diversified in different languages, countries, and cultures.

That does not mean that Wikipedia provides unbiased information, compiled from a 'Neutral Point of View', or devoid of truth claims (Tkacz 2012; 2014). As we mentioned, Wikipedia and other digital platforms provide a *collective perspective* established by coordinating, sometimes in quite conflicting ways, multiple individual viewpoints. But their biases are not based on an individual perspective, what can be useful for sociologists who seek to start their research without delimiting a priori their object of study with their personal point of view.

The Wikipedia record, then, is a form of knowledge with specific claims to validity and truth. For this reason, we do not propose to use it as a privileged form of access to 'historical truth'. Rather, we consider this information helpful as a heuristic tool, which sociologists can use *to confront their own knowledge, opinions, and beliefs about their discipline*, and thus become more aware of some of their own biases and blind spots.

References

Adams, Julia, Hannah Brückner, and Cambria Naslund. 2019. "Who Counts as a Notable Sociologist on Wikipedia? Gender, Race, and the 'Professor Test.'" *Socius* 5: 2378023118823946.

Agamben, Giorgio. 1998. *Homo Sacer: Sovereign Power and Bare Life*. Translated by Daniel Heller-Roazen. Edición: 1. Stanford, Calif: Stanford University Press.

Alexa Internet. 2022. "Alexa Top 500 Global Sites." 2022. https://www.alexa.com/topsites.

Apic, Gordana, Matthew J. Betts, and Robert B. Russell. 2011. "Content Disputes in Wikipedia Reflect Geopolitical Instability." *PloS One* 6 (6): e20902.

Aragon, Pablo, David Laniado, Andreas Kaltenbrunner, and Yana Volkovich. 2012. "Biographical Social Networks on Wikipedia: A Cross-Cultural Study of Links That Made History." In *Proceedings of the Eighth Annual International Symposium on Wikis and Open Collaboration*, 19. ACM.

Aron, Raymond. 2017. *Main Currents in Sociological Thought: Montesquieu, Comte, Marx, Tocqueville and the Sociologists and the Revolution of 1848*. Routledge.

———. 2018. *Main Currents in Sociological Thought: Volume 2: Durkheim, Pareto, Weber*. Routledge.

Assmann, Jan. 1995. "Collective Memory and Cultural Identity." *New German Critique*, no. 65: 125–33.

———. 2011. "Communicative and Cultural Memory." In *Cultural Memories*, 15–27. Springer.

Beytía, Pablo. 2020. "The Positioning Matters. Estimating Geographical Bias in the Multilingual Record of Biographies on Wikipedia."

Beytía, Pablo, Pushkal Agarwal, Miriam Redi, and Vivek Singh. 2022. "Visual Gender Biases in Wikipedia: A Systematic Evaluation across the Ten Most Spoken Languages." *AAAI Conference on Web and Social Media (ICWSM)*.

Beytía, Pablo, and Janosch Schobin. 2020. "Networked Pantheon: A Relational Database of Globally Famous People." *Research Data Journal for the Humanities and Social Sciences* 5.

Beytía, Pablo, and Claudia Wagner. 2022. "Visibility Layers: A Framework for Facing the Complexity of the Gender Gap in Wikipedia Content." *Internet Policy Review*. https://doi.org/10.31235/osf.io/5ndkm.

Biuk-Aghai, Robert P., Cheong-Iao Pang, and Yain-Whar Si. 2014. "Visualizing Large-Scale Human Collaboration in Wikipedia." *Future Generation Computer Systems* 31: 120–33.

Blondel, Vincent D., Jean-Loup Guillaume, Renaud Lambiotte, and Etienne Lefebvre. 2008. "Fast Unfolding of Communities in Large Networks." *Journal of Statistical Mechanics: Theory and Experiment* 2008 (10): P10008.

Bonabeau, Eric. 2009. "Decisions 2.0: The Power of Collective Intelligence." *MIT Sloan Management Review* 50 (2): 45.

Borch, Christian. 2012. *The Politics of Crowds: An Alternative History of Sociology*. Cambridge University Press.

Bourdieu, Pierre. 1988. *Homo Academicus*. Stanford University Press.

———. 1993. *Sociology in Question*. Vol. 18. Sage.

———. 2016. *Questions de Sociologie*. Minuit.

Bourdieu, Pierre, and Loïc JD Wacquant. 1992. *An Invitation to Reflexive Sociology*. University of Chicago press.

Bragues, George. 2007. "Wiki-Philosophizing in a Marketplace of Ideas: Evaluating Wikipedia's Entries on Seven Great Minds." *Available at SSRN 978177*.

Brin, Sergey, and Lawrence Page. 1998. "The Anatomy of a Large-Scale Hypertextual Web Search Engine." *Computer Networks and ISDN Systems* 30 (1–7): 107–17.

Brown, Adam R. 2011. "Wikipedia as a Data Source for Political Scientists: Accuracy and Completeness of Coverage." *PS: Political Science & Politics* 44 (2): 339–43.

Burrows, Roger. 2012. "Living with the H-Index? Metric Assemblages in the Contemporary Academy." *The Sociological Review* 60 (2): 355–72.

Burrows, Roger, and Mike Savage. 2014. "After the Crisis? Big Data and the Methodological Challenges of Empirical Sociology." *Big Data & Society* 1 (1): 2053951714540280.

Cárdenas, Julián. 2021. "Networking among Scientific Journal Editors. A Network Analysis among the Top 100 Sociology Journals." *Revista Española de Investigaciones Sociológicas (REIS)* 175 (175): 27–63. https://doi.org/10.5477/cis/reis.175.27.

Clauson, Kevin A., Hyla H. Polen, Maged N. Kamel Boulos, and Joan H. Dzenowagis. 2008. "Scope, Completeness, and Accuracy of Drug Information in Wikipedia." *Annals of Pharmacotherapy* 42 (12): 1814–21.

Clément, Maxime, and Matthieu J. Guitton. 2015. "Interacting with Bots Online: Users' Reactions to Actions of Automated Programs in Wikipedia." *Computers in Human Behavior* 50: 66–75.

Coser, Lewis A. 1971. *Masters of Sociological Thought: Ideas in Historical and Social Context*. Houghton Mifflin Harcourt P.

Couldry, Nick, and Andreas Hepp. 2018. *The Mediated Construction of Reality*. John Wiley & Sons.

Decoteau, Claire Laurier. 2013. "Exclusionary Inclusion and the Normalization of Biomedical Culture." *American Journal of Cultural Sociology* 1 (3): 403–30.

Demeter, Marton, and Tamas Toth. 2020. "The World-Systemic Network of Global Elite Sociology: The Western Male Monoculture at Faculties of the Top One-Hundred Sociology Departments of the World." *Scientometrics*, 1–27.

Esposito, Elena. 2008. "Social Forgetting: A Systems-Theory Approach." *Erll/A. Nünning (Hg.)*, 181–89.

Evans, James A., and Jacob G. Foster. 2011. "Metaknowledge." *Science* 331 (6018): 721–25.

Fortunato, Santo, Carl T. Bergstrom, Katy Börner, James A. Evans, Dirk Helbing, Staša Milojević, Alexander M. Petersen, Filippo Radicchi, Roberta Sinatra, and Brian Uzzi. 2018. "Science of Science." *Science* 359 (6379): eaao0185.

Giddens, Anthony. 1984. *The Constitution of Society: Outline of the Theory of Structuration*. Univ of California Press.

Giles, Jim. 2005. "Internet Encyclopaedias Go Head to Head." *Nature* 438: 900–901.

Gouldner, Alvin W. 1971. *The Coming Crisis of Western Sociology*. Heinemann.

Graham, Mark, Bernie Hogan, Ralph K. Straumann, and Ahmed Medhat. 2014. "Uneven Geographies of User-Generated Information: Patterns of Increasing Informational Poverty." *Annals of the Association of American Geographers* 104 (4): 746–64.

Halbwachs, Maurice. 1997. *La Mémoire Collective*. Albin Michel.

Halsey, Albert Henry. 2004. *A History of Sociology in Britain: Science, Literature, and Society*. Oxford University Press.

Heilbron, Johan. 2015. *French Sociology*. Cornell University Press.

Hepp, Andreas. 2016. "Pioneer Communities: Collective Actors in Deep Mediatisation." *Media, Culture & Society* 38 (6): 918–33.

Jara-Figueroa, C., Amy Z. Yu, and César A. Hidalgo. 2019. "How the Medium Shapes the Message: Printing and the Rise of the Arts and Sciences." *PloS One* 14 (2): e0205771.

Jemielniak, Dariusz. 2014. *Common Knowledge?: An Ethnography of Wikipedia*. Stanford University Press.

Käsler, Dirk. 2013. *Die Frühe Deutsche Soziologie 1909 Bis 1934 Und Ihre Entstehungs-Milieus: Eine Wissenschaftssoziologische Untersuchung*. Vol. 58. Springer-Verlag.

Kittur, Aniket, Bongwon Suh, Bryan A. Pendleton, and Ed H. Chi. 2007. "He Says, She Says: Conflict and Coordination in Wikipedia." In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 453–62. ACM.

Korom, Philipp. 2020. "The Prestige Elite in Sociology: Toward a Collective Biography of the Most Cited Scholars (1970-2010)." *The Sociological Quarterly* 61 (1): 128–63. https://doi.org/10.1080/00380253.2019.1581037.

Lamo de Espinosa, Emilio. 2001. "La Sociología Del Siglo XX." *Reis*, 21–49.

Latour, Bruno. 2005. *Reassembling the Social: An Introduction to Actor-Network-Theory*. Edición: 1st. Oxford: Oxford University Press.

Lavsa, Stacey M., Shelby L. Corman, Colleen M. Culley, and Tara L. Pummer. 2011. "Reliability of Wikipedia as a Medication Information Source for Pharmacy Students." *Currents in Pharmacy Teaching and Learning* 3 (2): 154–58.

Lazer, David MJ, Alex Pentland, Duncan J. Watts, Sinan Aral, Susan Athey, Noshir Contractor, Deen Freelon, Sandra Gonzalez-Bailon, Gary King, and Helen Margetts. 2020. "Computational Social Science: Obstacles and Opportunities." *Science* 369 (6507): 1060–62.

Lazer, David, Alex Pentland, Lada Adamic, Sinan Aral, Albert-László Barabási, Devon Brewer, Nicholas Christakis, Noshir Contractor, James Fowler, and Myron Gutmann. 2009. "Computational Social Science." *Science* 323 (5915): 721–23.

Lee, Monica, and John Levi Martin. 2015a. "Coding, Counting and Cultural Cartography." *American Journal of Cultural Sociology* 3 (1): 1–33. https://doi.org/10.1057/ajcs.2014.13.

———. 2015b. "Response to Biernacki, Reed, and Spillman." *American Journal of Cultural Sociology* 3 (3): 380–415. https://doi.org/10.1057/ajcs.2015.11.

Lévi-Strauss, Claude. 1945. "French Sociology." *New York*.

Lévy, Pierre. 1997. *Collective Intelligence: Mankind's Emerging World in Cyberspace*. Perseus books.

Lih, Andrew. 2009. *The Wikipedia Revolution: How a Bunch of Nobodies Created the World's Greatest Encyclopedia*. Hachette Books.

Lipset, Seymour Martin. 1994. "The State of American Sociology." In *Sociological Forum*, 9:199–220. Springer.

Luhmann, Niklas. 1999. *Die Gesellschaft Der Gesellschaft*. Suhrkamp Verlag.

Lupton, Deborah. 2014. *Digital Sociology*. Routledge.

Lupton, Deborah, Inger Mewburn, and Pat Thomson. 2017. *The Digital Academic: Critical Perspectives on Digital Technologies in Higher Education*. Routledge.

Lüschen, Günther. 2013. *Deutsche Soziologie Seit 1945: Entwicklungsrichtungen Und Praxisbezug*. Vol. 21. Springer-Verlag.

Marres, Noortje. 2017. *Digital Sociology: The Reinvention of Social Research*. John Wiley & Sons.

Menini, Stefano, Rachele Sprugnoli, Giovanni Moretti, Enrico Bignotti, Sara Tonelli, and Bruno Lepri. 2017. "Ramble on: Tracing Movements of Popular Historical Figures." In *Proceedings of the Software Demonstrations of the 15th Conference of the European Chapter of the Association for Computational Linguistics*, 77–80.

Mützel, Sophie. 2015. "Facing Big Data: Making Sociology Relevant." *Big Data & Society* 2 (2): 2053951715599179.

Nemoto, Keiichi, and Peter A. Gloor. 2011. "Analyzing Cultural Differences in Collaborative Innovation Networks by Analyzing Editing Behavior in Different-Language Wikipedias." *Procedia-Social and Behavioral Sciences* 26: 180–90.

Nisbet, Robert A. 1993. *The Sociological Tradition*. Transaction publishers.

Olick, Jeffrey K., and Joyce Robbins. 1998. "Social Memory Studies: From 'Collective Memory' to the Historical Sociology of Mnemonic Practices." *Annual Review of Sociology* 24 (1): 105–40.

Ollion, Etienne, and Andrew Abbott. 2016. "French Connections: The Reception of French Sociologists in the USA (1970-2012)." *European Journal of Sociology / Archives Européennes de Sociologie* 57 (2): 331–72. https://doi.org/10.1017/S0003975616000126.

Page, Lawrence, Sergey Brin, Rajeev Motwani, and Terry Winograd. 1999. "The PageRank Citation Ranking: Bringing Order to the Web." Stanford InfoLab.

Pender, Michael P., Kaye Lasserre, Lisa Kruesi, Christopher Del Mar, and Satyamurthy Anuradha. 2008. "Putting Wikipedia to the Test: A Case Study."

Pentland, Alex. 2014. *Social Physics: How Good Ideas Spread-the Lessons from a New Science*. Penguin.

Rector, Lucy Holman. 2008. "Comparison of Wikipedia and Other Encyclopedias for Accuracy, Breadth, and Depth in Historical Articles." *Reference Services Review* 36 (1): 7–22.

Reznik, Ilia, and Vladimir Shatalov. 2016. "Hidden Revolution of Human Priorities: An Analysis of Biographical Data from Wikipedia." *Journal of Informetrics* 10 (1): 124–31.

Rogers, Richard. 2013. *Digital Methods*. MIT press.

———. 2015. "Digital Methods for Web Research." *Emerging Trends in the Social and Behavioral Sciences: An Interdisciplinary, Searchable, and Linkable Resource*, 1–22.

Rollin, Guillaume, José Lages, and Dima L. Shepelyansky. 2019. "World Influence of Infectious Diseases From Wikipedia Network Analysis." *IEEE Access* 7: 26073–87.

Ronen, Shahar, Bruno Gonçalves, Kevin Z. Hu, Alessandro Vespignani, Steven Pinker, and César A. Hidalgo. 2014. "Links That Speak: The Global Language Network and Its Association with Global Fame." *Proceedings of the National Academy of Sciences* 111 (52): E5616–22.

Royal, Cindy, and Deepina Kapila. 2009. "What's on Wikipedia, and What's Not...? Assessing Completeness of Information." *Social Science Computer Review* 27 (1): 138–48.

Salganik, Matthew. 2019. *Bit by Bit: Social Research in the Digital Age.* S.l.: Princeton University Press.

Samoilenko, Anna, Florian Lemmerich, Katrin Weller, Maria Zens, and Markus Strohmaier. 2017. "Analysing Timelines of National Histories across Wikipedia Editions: A Comparative Computational Approach." In *Eleventh International AAAI Conference on Web and Social Media*.

Savage, Mike, and Roger Burrows. 2007. "The Coming Crisis of Empirical Sociology." *Sociology* 41 (5): 885–99.

Schäfer, Mirko Tobias, and K. F. Van Es. 2017. *The Datafied Society: Studying Culture through Data*. Amsterdam University Press.

Schich, Maximilian, Chaoming Song, Yong-Yeol Ahn, Alexander Mirsky, Mauro Martino, Albert-László Barabási, and Dirk Helbing. 2014. "A Network Framework of Cultural History." *Science* 345 (6196): 558–62.

Shannon, Claude E., and Warren Weaver. 1949. "The Mathematical Theory of Communication,

117 Pp." *Urbana: University of Illinois Press*.

Skiena, Steven, and Charles B. Ward. 2014. *Who's Bigger?: Where Historical Figures Really Rank*. Cambridge University Press.

Stone, Lawrence. 1971. "Prosopography." *Daedalus*, 46–79.

Swidler, Ann. 1986. "Culture in Action: Symbols and Strategies." *American Sociological Review*, Vol 51, Nº 2, 273–86.

Tkacz, Nathaniel. 2012. "The Truth of Wikipedia." *Academic Research into Wikipedia*.

———. 2014. *Wikipedia and the Politics of Openness*. University of Chicago Press.

Tsvetkova, Milena, Ruth García-Gavilanes, Luciano Floridi, and Taha Yasseri. 2017. "Even Good Bots Fight: The Case of Wikipedia." *PloS One* 12 (2): e0171774.

Viégas, Fernanda B., Martin Wattenberg, and Kushal Dave. 2004. "Studying Cooperation and Conflict between Authors with History Flow Visualizations." In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 575–82. ACM.

Wagner, Claudia, David Garcia, Mohsen Jadidi, and Markus Strohmaier. 2015. "It's a Man's Wikipedia? Assessing Gender Inequality in an Online Encyclopedia." In *Ninth International AAAI Conference on Web and Social Media*.

Warczok, Tomasz, and Stephanie Beyer. 2021. "The Logic of Knowledge Production: Power Structures and Symbolic Divisions in the Elite Field of American Sociology." *Poetics* 87: 101531.

Wickham, Hadley. 2016. "Rvest: Easily Harvest (Scrape) Web Pages, 2016." *R Package Version 0.3* 1.

Wikimedia Foundation. 2020. "Wikidata." 2020. https://www.wikidata.org/wiki/Wikidata:Main_Page.

Wikipedia. 2019. "Wikipedia:No Original Research." In *Wikipedia*. https://en.wikipedia.org/w/index.php?title=Wikipedia:No_original_research&oldid=887957640.

———. 2022. "Wikipedia:Notability." In *Wikipedia*. https://en.wikipedia.org/w/index.php?title=Wikipedia:Notability&oldid=888716184.

Yang, Heng-Li, and Cheng-Yu Lai. 2010. "Motivations of Wikipedia Content Contributors." *Computers in Human Behavior* 26 (6): 1377–83.

Yenikent, Seren, Peter Holtz, and Joachim Kimmerle. 2017. "The Impact of Topic Characteristics and Threat on Willingness to Engage with Wikipedia Articles: Insights from Laboratory Experiments." *Frontiers in Psychology* 8: 1960.

Yu, Amy Zhao, Shahar Ronen, Kevin Hu, Tiffany Lu, and César A. Hidalgo. 2016. "Pantheon 1.0, a Manually Verified Dataset of Globally Famous Biographies." *Scientific Data* 3: 150075.