Geocaching: Tracing Geotagged Social Media Research Using Mixed Methods

Submitted in partial fulfillment of the conditions
for the award of the degree of Doctor of Philosophy.

Judit Varga

School of Sociology and Social Policy
Faculty of Social Sciences
University of Nottingham

April 2021

Exploring how different ways of thinking and being meet,

In memory of my parents

Gábor Varga (1951 - 2012)

Headmaster of the Alsóerdősori - Bárdos Lajos School (1991-2011)

& Mathematics and physics teacher

& Honorary Citizen (Posztumusz Díszpolgár) of Erzsébetváros (Seventh District), Budapest, Hungary

After whom the Varga Gábor-díj (Award) was named

Magdolna Fellner (1953 - 2020)

Chemistry and pedagogy teacher

& Osztályfőnök ('class-lead teacher') at the Raoul Wallenberg School

Who spent their lives teaching, learning and doing diverse ways of knowing.

Abstract

This thesis explores the development of academic research with geotagged social media data (geosocial research) - an emerging computational, digital social research field - using 19 semi-structured interviews with scholars from diverse disciplines, participant observation at a geosocial research summer school and scientometrics. It asks: 'how can we study the development of geosocial research approaches through combining STS and scientometrics?' for five main reasons: to explore the diversity of computational social research; reflect on the ESRC's (2013) call to 'close the gap' between quantitative and qualitative human geography; contribute to methodological discussions in academic literature which call for combining STS and scientometrics; co-compose knowledge with distinct ways of knowing through mixing methods; and inform research methods curriculum development in the social sciences.

Using new forms of digital data (like social media posts) is core to contemporary social science. Scholars from diverse disciplines conduct geosocial research. It thus provides rich opportunities to study how diverse approaches to computational social research develop. I combine STS and diverse scientometric methods as part of a single case study iteratively to explore how they can co-compose knowledge.

The thesis contributes to literature which explores the STS – scientometrics interface. Most existing studies either reflect on diverse mixed methods approaches from theoretical or methodological perspectives, or provide worked examples using specific mixed methods designs. Conceptually, this thesis contributes by highlighting the need to develop and evaluate the affordances of computational methods for STS in light of the interpretative context - including research questions, characteristics of the studied research practice, theories and prior findings. I developed computational methods iteratively, in light of my theoretical and empirical knowledge about geosocial research. Empirically, the thesis first contributes by showing how diverse combinations of STS and scientometrics – including statistical and visual network analyses as well as descriptive statistics – can inform a single case study. Second, it offers three ways STS and scientometrics can co-compose knowledge by aligning their units of analyses, reflecting on how calculation acts inform qualitative analysis even when analytical units are not aligned, and using each method inductively.

I combined STS and scientometrics to study practices through which geosocial research approaches develop - including collaboration, developing (sub)-disciplinary communities and methods' mediation of geosocial research. I also identified geosocial research approaches and compared them using mixed methods. Finally, I combined insights from STS and scientometrics to highlight the construction of my own analyses.

Using mixed methods, the thesis argues that geosocial research is a collection of approaches rather than a coordinated community. I highlight fourteen practices that enable scholars to develop their approaches, including interdisciplinary collaboration; setting up distinct geosocial laboratories to experiment with geosocial data; reflecting on the data analysis process; and using local knowledge about spaces. I differentiate 'social', 'technical' and 'geographic' approaches, which differ in terms of the methods they use and spatial units they study. Finally, I illustrate approaches' heterogeneity - including their diverse computational approaches - and similarities, such as their urban studies focus.

Acknowledgements

I would like to thank my supervisors, Reiner Grundmann, Andrew Goffey and Alison Mohr for their advice.

I would like to thank my sponsors, the Horizon Digital Economy Centre for Doctoral Training (especially Sarah Sharples and rest of the Horizon CDT selection committee), the Engineering and Physical Sciences Research Council and Ordnance Survey for the opportunities. Special thanks to Jeremy Morley for his continued interest.

Special thanks to all my interviewees who generously shared their experiences and time.

Special thanks to Fred Cummins for encouraging me to pursue a PhD, his support and all the inspiring conversations, Brigitte Nerlich for her support and for always lending a helping ear, and Sujatha Raman for the encouraging confirmation review.

Special thanks to Nora Wikoff and Will Paterson for the teaching opportunities, their support, trust, and friendly welcome, and them as well as Macarena Orchard, Elena Genova, Esmeralda Bon and Karolina Trdlicova for the wonderful co-teaching experiences.

Special thanks to Veronica Blake and Andrea Haworth for all their invaluable, quick, expert and kind help with administration and extensions throughout the years, and at personally and globally challenging times. I am so grateful. Many thanks for Dawn Todd and Vicky Spencer for their kind and energetic presence.

This thesis would not have been possible without Matt Voigts. Special thanks for your invaluable advice, writing tips and support throughout these often crazy years.

Special thanks to Claudia Iacob and Shamal Faily for their wonderful company during my Southampton visits, and to Claudia for her support an encouragement.

Special thanks to the CWTS Community: to Ismael Rafols for connecting us, and Sarah de Rijcke and Thomas Franssen for believing in this project and in me. I am very grateful. Thanks to Sarah, Thomas, Petra van der Weel, Tjitske Holtrop, Wolfgang Kaltenbrunner, Jochem Zuijderwijk, Jackie Ashkin and Sarah Rose Bieszczad for their warm welcome and encouragement, and Guus Dix for the friendly and inspiring chats. Special thanks to the FluidKnowledge Team, especially to Jackie, Sarah Rose and Sonia Mena Jara for their collegial support in our virtual office(s) in the past year.

Special thanks to Federica Russo, Anne Beaulieu, Martha Lampland, Judit Gárdos and Csaba Szummer for the invigorating and encouraging discussions at crucial moments. Thanks to Nees Jan van Eck for the chat about matrix normalisation.

Special thanks to Rebeka O. Szabó, Srebrenka Letina, Johannes Wachs, Cristian Candia-Castro-Vallejos, Dávid Deritrei and Olga Perédi whom I met during my visit to CEU's DNDS department, as well as Linda Sīle for the inspiring chats, peer support, friendship and help - professional or personal.

This thesis would also have not been possible without the friendship, support and invaluable advice of Betti Bódi, Hannah Keding, Niki Cheong, Juliet Gryspeerdt, Simon Breden, Ashley Lewis, Shalaka Kurup, Richard and Rachel Ramchurn, Andrew Moffat, Neelima Sunil, Eleanor Hadley Kershaw, Mathilde Vialard, Theo Torres, Macarena Orchard, Jakob von Raumer and Anna Bánki. Thanks for sharing the Nottingham - PhD journey with me, for the fun times, and for your patient support during these crazy times!

Special thanks to Juliet, Niki, Betti and Anna for the always inspiring, energising and insightful conversations, and to Hannah for all this and also sharing in the institutional journey with such wisdom, energy and productive irony.

Special thanks to Katalin Varga, András Bánvölgyi MD-PhD who after all these years finally became a real dr., Lilla Horváth, Petra Vancsura, Márta Szrakityné Fellner, Marianna and Márk Szrakity, Márta Szrakity-Vaskó, András Vaskó, Miklós Losó, Judit and Gyuri Bánvölgyi, Merijn and Simon Gábor Marius Verschuur as well as Enikő Péter for all their help, always, and also in the past few years. Words cannot express my love and gratitude. I could not have done this without you!

Special thanks to Petra, Lili and Katus for being there no matter what, for always cheering for me, for your patience and advice. I would not be who I am without you. Lili, special thanks for bringing joy to and sharing in the last bit of this crazy ride!

Also huge thanks to Andi Perlusz, Éva Gombos and Péter Sirály for their invaluable support, friendship and (medical) help. Huge thanks to Eszter Kádár, Aliz Süki and the rest of the family; as well as Brian Blecher, Betti Bódi, Rob Taylor and Dave Edwards for their friendship, support and help at key moments with housing and a house move.

Last but not least, huge thanks for the friendship, support and patience of Gina Binét, Tiziana Callari, Zsuzsi Czinki, Viki Farsang, Andrea Kóbor, Anna Misovicz, Szandi Pankász, Juli Takács, Angéla Tóth and Lilla Üveges. Thanks for being there, cheering for me and not forgetting me while I've been absent the past few years! Special thanks to Gina for her ceaseless encouragement!

Contents

List of Tables

List of Figures

Chapter 1

Introduction

”The digital does not solve sociology’s problems, it unsettles and exposes the troubles and difficulties of social research, and, we should strive to [make] these available for public exploration.”
– Noortje Marres, Digital Sociology, p. 21.

“The social sciences found their telescope with the big data”, as one of my interviewees put it. This quote attests to digital traces’ growing use for social scientific research. Situated practices with ubiquitous digital technologies create digital traces of these activities, such as social media posts, online reviews, online purchase records and phone call logs (cf. Ruppert, Law, and Savage, 2013). As section 1.1 will discuss, recently established university programs, conferences and recent books that discuss digital or computational social research illustrate the interest in using such digital traces for academic research.

To study diverse uses of digital data and computational data analysis methods for social research, this project employs mixed methods to study the knowledge diversity of a digital, computational social research field. It combines STS, interviews and scientometrics to explore how academic researchers from diverse disciplines use ’geotagged’ social media data (geosocial data, for short). My definition of geosocial data include social media posts with diverse forms of geographic information, such as geo-coordinates, place tags or location mentions. I interviewed 19 scholars who use geosocial data and have backgrounds in diverse disciplines including anthropology, sociology, geography, computer science, physics and mathematics. I also conducted participant observation at a 10-day long summer school about geosocial data analysis and conducted diverse scientometric analyses, including visual and statistical network analyses and descriptive statistics. My methods were informed by concepts from Science and Technology Studies (STS). As section 1.2 will explain, I study the knowledge diversity of digital or computational social research inspired by STS scholarship that explores how dialogues can develop among distinct ways of knowing in the anthropocene, and due to my interest in social research methods curriculum development.

As section 1.2 will explain in more detail, combining STS, interviews and scientometrics to study academic geotagged social media data research (geosocial research, for short) provides a rich opportunity to explore ways computational data analysis and (interpretative) social scientific research can be combined. STS and scientometrics have largely developed along different trajectories since the 1980s. Recently, there has been renewed interest in combining them, but no consensus exists on how to do so.

I use *geocaching* - mentioned in the thesis' title - as a metaphor to refer to studying geosocial research using mixed-methods. As part of the geocaching game, players seek 'treasures' hidden in public areas (e.g. boxes filled with toys or standalone objects like painted rocks or magnets) based on their GPS coordinates. Once found, players may modify the treasures, for example, by exchanging the contents of boxes, as well as painting or attaching letters to objects, but leave them in the same geo-coordinates for subsequent players. Tracing treasures - often hidden in non-urban nature areas or urban parks - requires players to precisely locate their GPS coordinates and also think out of the box by searching for them in unexpected places and unplanned ways, for example, in bushes, behind landmarks, in underground traps or on top of trees. I use geocaching as a metaphor to refer to both geosocial research and my study of it. Like geocaching, geosocial research creates diverse spaces. In geocaching, geo-coordinates signify treasures' map-able locations, enabling a collective, communal game. However, thanks to the diverse forms treasures take and players' adventures as they trace them, geocaching enacts technologically mediated, culturally and experientially diverse spaces. Similarly, geosocial scholarship often combines technologically enabled mapping capabilities with diverse conceptual approaches, creating various spaces even if they can be mapped using the same geo-coordinates. My efforts to trace geosocial research practices and the spaces they create is also similar to geocaching. It is my hope that through combining scientometric 'maps', interviews, participant observation and STS concepts, I can successfully follow the 'treasures' and 'traces' geosocial scholars create. Firstly, I hope to 'locate' their 'treasures' (geosocial research findings) and 'traces' (geosocial research practice) by describing geosocial research in ways that geosocial scholars recognise. Secondly, I hope to 'modify' 'treasures' - describe geosocial research findings and geosocial research practice in ways that differ from geosocial scholars' narratives and yield new insights about contemporary computational social research. Thirdly, akin to geocaching instructions, I document my own 'treasure hunt' - my mixed-methods journey - for subsequent 'players': the STS research community. I aim to develop mixed-methods solutions that guide me - benefiting from scientometrics' and STS' established methodological norms - and also help me do (re)search (study and conduct computational social research) in unexpected ways, through combining the distinct methodological norms of scientometrics and STS in diverse ways.

The thesis' main contribution is to the literature that explores the STS – scientometrics interface. Empirically, I contribute by showing how the diverse combinations of STS concepts, interview analysis and scientometrics can inform a single case study. Many existing studies that combine STS or sociology of science and scientometrics use either statistical or visual analyses, but rarely both. In addition, most advocate for the use of network analysis methods, disregarding the opportunities associated with descriptive statistics. This project's focus on science studies methods is in line with the discipline's recent self-reflection practices that position 'method' at the centre of attention (Smith-Doerr, 2017) and calls for STS to experiment with digital methods (e.g. Elgaard Jensen, 2013). In particular, scholars have called for STS to explore digital, computational scholarship, and to combine STS and scientometrics in practice (e.g. Wyatt, Milojević, et al., 2017; Marres and Gerlitz, 2016; Cambrosio, Bourret, et al., 2014; cf. Neff et al., 2017). The project's mixed-methods approach - and its experimentation with diverse computational methods - also aligns with digital STS' commitment to build bilateral bridges between STS and disciplines that use and make digital tools by welcoming concepts and methods 'inward' (in addition to STS' tendency to 'exporting' them) (Vertesi et al., 2019b).

Conceptually, I contribute to the above literature by highlighting the need to develop mixed methods approaches and assess the affordances of computational or scientometric methods 'for STS' (in single quotes to signal STS' diversity) in light of the interpretative context, including the conceptual framework, research questions, characteristics of the studied research practice and previous findings. Existing literature that reflects on the affordances of computational methods 'for STS' or related fields, including literature in digital sociology or digital anthropology, mainly evaluates methods in light of epistemological frameworks such as actor-network theory (e.g. Cambrosio, Bourret, et al., 2014) or ethnographic field work (theories) (e.g. Munk, 2019). My conceptual framework and previous interview and scientometric findings informed not only the interpretation of computational findings, but also the development of each computational method - and the aggregations or calculations therein. I argue that evaluating the affordances of computational methods in light of the interpretative context helps develop mixed methods research without restricting methods development on a type of method - such as visual or statistical, network analysis or descriptive statistics - *a priori*. This flexibility, in turn, helps explore how computational data analysis and STS or other interpretative social scientific research traditions *do* or *could* meet. Finally, given the importance of the interpretative context, my use of diverse mixed methods analyses in one case study helps assess their strengths and weaknesses.

Supported by, and combined with the conceptual framework and interviews, I do the following. I use the network modularity metric to show that scholars differentiate geosocial research approaches thorough collaboration and trace the differentiation of geosocial research approaches over time. I apply descriptive statistics to highlight that interviewees primarily belong to disciplinary and sub-disciplinary communities and that geosocial research is a collection of approaches but not a coordinated research community. I use the term-frequency inverse document frequency metric (TF-IDF) to show that geosocial research approaches differ in terms of methods they use and spatial units they study. I utilise visual, heterogeneous network analyses to show that computational methods mediate knowledge about spaces. Combining methods contributes to interview findings in two main ways. Firstly, it helps explore practices identified through interviews on extended spatial and temporal scales - better illustrating the differentiation of geosocial research approaches. Secondly, it helps me produce findings inductively: I identify geosocial research approaches, differences among them, and trace how computational methods mediate knowledge about spaces not hypothesised or studied through interviews.

1.1 Growing Popularity of Digital, Computational Social Research

As the examples below illustrate, the growing interest in using digital traces and computational data analysis for academic, social research is found in recently-established academic institutions, conferences and books that discuss digital and computational social research. Recent academic institutions dedicated to exploring the uses of digital traces in social research include undergraduate and postgraduate training programs as well as research institutes. The University of California, Davis, the University of Groningen and Maastricht University have recently launched undergraduate (minor) programmes dedicated to teaching a combination of data analysis techniques and critical 'data studies' skills to train graduates to work with data in diverse teams and reflect on the epistemic and political aspects of data analysis (Dumit and Nafus, 2018; *University of Groningen* 2019). All three programs - ran by STS scholars - emphasise the need to reflect on the political implications of data science, and position their programs as an avenue to impact societal 'big data' practices (e.g. *University of Groningen* 2019). Graduate degree programs organised around the analysis of new types of data for social scientific inquiry include Europe's first network science PhD program, which in September 2018 was expanded to form the Department of Data and Network Science at Central European University, co-founded by social scientists and physicists (Kertesz and Vedres, 2018). In addition, the Data Analytics and Society PhD program - an initiative of the UK's The Economic and Social Research Council (ESRC) - aims to foster research at the intersection of data science and social research, is based across 4 British Universities, and welcomed its first cohort of students in October 2017 (Pound, 2018). Finally, the Techno-Anthropology Lab at Aalborg University - which explores digital methods for STS (e.g. Elgaard Jensen et al., 2019) - provides undergraduate training and conducts research which combines STS or related fields and hands-on techno-scientific practice in diverse domains.

Recently-established conferences dedicated to exploring the interface between computational data analysis and social research include the first Data and Social Research (DSSR) conference, organised in 2018, and IC2S2, the International Conference on Computational Social Science, founded in 2015 (IC2S2, 2019). In addition, the 'Big Data and the Power of Narrative' workshop (organised at IT University of Copenhagen in March 2019) attracted an unexpected number of participants and resulted in a waiting list with over 100 people (Conference Organisers, personal communication). These complement conferences with a longer history that discuss computational social research, such as the Sunbelt Social Networks Conference, the yearly conference of the International Network for Social Network Analysis (INSNA) founded in 1981.

Recently-published books that discuss the combination of computational methods, digital data analysis and social scientific research include Knox and Nafus (2018), who explore the interface between digital knowledge practices and ethnography; Marres (2017a) and Lupton (2014), who discuss emerging data practices in 'digital sociology'; Salganik (2017), who discusses methodological and ethical considerations raised by computational social research methods using worked examples; and Wouters, Beaulieu, et al. (2012), who explore the theory, practice and infrastructure of digital scholarship in the humanities and social sciences. In addition, over a third of the scholars who responded to a recent survey studying the use of "big data" in social science conducted computational or digital social research, and half of those not engaged in such research at the time of the study intended to do so in the future (Metzler et al., 2016).

Finally, scholars have suggested diverse terms to refer to computational or digital social research, such as computational social science (Lazer et al., 2009), critical data studies (Iliadis and Russo, 2016) and digital sociology (e.g. Marres, 2017b). Each term is embedded in distinct disciplinary research heritages. As the proliferation of terms indicate, scholars who work in this area often disagree about how digital traces can or should be used for research. Some argue that novel digital knowledge practice can foster diverse ways of knowing (e.g. DeLyser and Sui, 2012) whilst others caution that it does the exact opposite, that it hinders diverse ways of knowing (e.g. Barnes and Wilson, 2014).

1.2 Societal and Scientific Relevance

This project explores the knowledge diversity of geosocial research using mixed methods for five main reasons. Firstly, in line with STS' interest in exploring and fostering knowledge diversity, it helps explore how diverse forms of knowledge can be created in computational (social) research, popular in and outside of academia. Academic digital and computational knowledge practices reflect the popularity of 'big data' and computational research across industries. For example, Dumit and Nafus (2018) partially frame their data studies minor program mentioned in section 1.1 in terms of the economic relevance of these skills (cf. Elgaard Jensen, 2013). STS highlights the importance of nurturing diverse types of knowledges and being critical of ways of knowing that are positioned as 'factual' and 'correct'. All knowledge is partial - it favours some perspectives at the expense of others - and is thus inherently political. Ways of knowing framed as 'undoubtedly true' and 'objective' in the sense that they are 'independent of human judgement' obscure other ways of knowing and other experiences. Similar to Rieder's (2017) who argues that detailed studies of computing practices may help shape them, this project explores data practices to better understand how digital traces *do* and *could* support diverse ways of knowing.

Secondly, in the face of today's 'grand challenges' like climate change and the global rise of extreme political movements - which highlight our interconnectedness amid the difficulties they bring - STS scholarship emphasizes the importance of co-existing alongside, in dialogue with (e.g. Latour, 2013) or collaborating with one other across diverse ways of knowing (e.g. Haraway, 2016). STS scholars argued that social scientists can help 'co-compose' knowledge and foster dialogues in such situations where the terms of engagement and values are collectively negotiated, rather than critiquing them from distance (e.g. Latour, 2005a; Latour, 2010; Birkbak, Krogh Petersen, and Elgaard Jensen, 2015; Elgaard Jensen, 2012). As discussed below, the geosocial case study and my mixed-methods practice help explore dialogue and alliance building among diverse knowledge traditions, such as computational data analysis and (interpretative) social science.

Thirdly, as outlined earlier, combining computational data analysis and STS contributes to methodological discussions in science studies, including to recent literature that calls to combine STS and scientometrics, and digital STS' goal to build bilateral bridges between STS and digital methods practices.

Fourthly, I reflect on Economic and Social Research Council's (ESRC) (2013) call to close the gap between quantitative and qualitative human geography by emphasizing their complementarity in light of my findings about how scholars from diverse disciplines use computational data analysis for geosocial research. The ESRC (2013) calls to increase computational training in British geography and argues that education programs should emphasize the complementarity of these methodological traditions - for example, their ability to study phenomena at diverse scales - to change existing, dominant views that highlight their divisions.

Finally, I present my research in the hopes that it can inform social science research methods teaching. The recent surge in academic programs that seek to teach computational, digital social methods discussed above illustrates the goals' current relevance. During my PhD studies I taught undergraduate research methods modules to familiarise myself with existing curricula. I hope that my research will help me contribute to curriculum development in the future.

Geosocial research provides rich opportunities to explore the diversity of computational social research because it is practiced by scholars from diverse disciplines and is currently being formed. The abundance of geosocial data is a relatively recent phenomenon. Social media platforms that afford them - such as Twitter, Instagram, Panoramio and Flickr - were launched around and after 2005. Much as my interviewees have diverse disciplinary backgrounds, the composition of scientometric data about geosocial research this thesis uses - which comprises papers published in journals associated with the social sciences, computational sciences, health and environmental research - and the thesis' findings show scholars from diverse disciplines developing distinct approaches to geosocial research. As section 2.2 will illustrate, the use of geosocial data has spurred lively controversies among scholars. Geosocial research approaches are not (yet) institutionalised, it is a research field in formation.

The project's mixed methods approach which combines interviews, participant observation and the computational analysis of scientometric data provides opportunities to explore the diversity of computational or digital social research because, as as outlined above, STS and scientometrics have developed largely independently since the the 1980s. Combining them in multiple ways as part of one project allows me to experimentally explore alliances among research techniques associated with distinct knowledge traditions (cf. Marres and Gerlitz, 2016).

To study how the computational analysis digital traces *do* and *could* support diverse ways of knowing, this project explores the three research questions below. The sub-questions of the Third Research Question explore how, on the one hand, methods mixing help explore geosocial research, and on the other hand, how we can assess the affordances of computational methods 'for STS'.

1. How do different approaches to geosocial research develop?
2. How do approaches to geosocial research differ?
3. How can we combine scientometrics and STS to study geosocial research?
   3.1. How does mixing methods help study the development of and difference among approaches to geosocial research?
   3.2. How can we assess the suitability of computational methods for STS?

1.3 Overview of Chapters

Most chapters combine a discussion of geotagged social media data practices and mixed methods STS, and thus contribute to all three research questions.

Chapter Two briefly introduces geosocial research - outlining illustrative examples and controversies - and outlines the project's conceptual framework. I conceptualise science as practice and assume that scholars develop diverse approaches to geosocial research through relational practices, such as collaborating, exploring the affordances of geosocial data and dialogues with their academic communities. In addition, I interpret and develop scientometric methods in light of interviews and my conceptual framework.

Chapter Three outlines the project's methodology. It discusses my mixed methods case study approach and the data analysis infrastructure (including software I used for scientometrics and my interview analysis approach). It also outlines each computational method in detail, alongside my approach to interview analysis as well as ethical considerations and the limitations of my methodology.

Chapters Four - Eight discuss the project's empirical findings.

Chapter Four explores the First Research Question by discussing the following three practices that help interviewees develop their approaches to geosocial research. It presents a core finding: interviewees' unequivocally state that developing their approach to geosocial research requires them to combine computational data analysis and social scientific research, which they find challenging. In addition, the chapter outlines two practices that motivate interviewees to combine these knowledge traditions and thus develop their geosocial research approaches: their concurrent academic and non-academic employment, and their aesthetic appreciation of both social research and computational or social media data analysis.

Chapter Five explores *how* interviewees combine computational data analysis and social scientific research to develop their approaches to geosocial research. In response to the First Research Question, it highlights three practices that help interviewees combine these knowledge traditions: collaborating with scholars with complementary skills, setting up distinct 'geosocial laboratories' and experimenting with computational data analysis methods. In addition, it addresses the Second Research Question by arguing that interviewees with social scientific and technical backgrounds combine these knowledge traditions differently when they seek data patterns. While the former state they combine them iteratively - and identify data patterns through combining statistical, computational criteria and social theory - the latter claim they combine them sequentially and identify data patterns primarily based on statistical or computational criteria. Finally, it contributes to the Third Research Question by studying collaboration using a combination of interviews and scientometrics.

Chapter Six explores how interviewees make institutional homes for their geosocial research - which allows them to develop their approaches to it. To address the First Research Question, it highlights two such practices: imagining geosocial research in light of their disciplinary heritage; and social scientists' efforts to differentiate their approach to geosocial research from computational social science and geographic information systems science informed approaches. It also uses mixed methods to identify social and technical approaches by tracing their separation scientometrically. The latter helps answer the Second and Third Research Questions - studying the differences among geosocial research approaches using mixed methods.

Chapter Seven primarily explores how geosocial research approaches differ, using mixed methods. Through interviews - in response to the First Research Question - it argues that reflecting on how analytical decisions and social media platforms shape geosocial data and knowledge about spaces is essential for interviewees' development of their approach to geosocial research. However, I argue that social and technical geosocial research approaches entail different reflexivities. While the former highlights experiences and historicity, the latter foregrounds computations and assesses knowledge about spaces in demographic terms. This highlights a difference in their explanations of situated practices. In addition, using scientometrics - through clustering a citation network of geosocial papers, and analysing them using descriptive statistics and visual network analysis - it explores differences in the methods social and technical approaches use and the way they study spaces. My analyses highlight social geosocial research's relative focus studying diverse situated practices at specific locations, compared to technical geosocial research' focus on tracing 'senseable' practices at diverse spatio-temporal scales. Finally, through combining scientometrics and interviews, it identifies a third approach: geographic geosocial research. This helps highlight the diversity of geosocial research and the affordances of scientometrics to help study knowledge diversity inductively.

Chapter Eight, the last empirical chapter explores how research methods mediate my findings about the differences among geosocial research approaches, and geosocial scholars' findings about spaces. To study how methods mediate my knowledge about geosocial research approaches - and study the Second and Third Research Questions - it juxtaposes scientometric findings from Chapter Seven with findings obtained through clustering a second citation network and the noun phrase co-occurrence network of geosocial papers. Compared to the scientometric findings of Chapter Seven, the latter analyses highlight the diverse ways computational data analysis methods are used in geosocial research and approaches' joint focus on studying specific locations. To study how methods mediate geosocial scholars' findings about spaces, it uses visual, heterogeneous network analyses to study how two computational data analysis methods - machine learning and social network analysis - mediate knowledge about spaces. In addition, based on interviews, it argues that interviewees from diverse disciplines use local knowledge to study specific locations, such as London or Singapore. These help explore the First and Third Research Questions.

Chapter Nine concludes the thesis by summarising its main findings with respect to each research question. Importantly, it discusses the thesis' contribution to literature that explores combinations of STS and scientometrics as follows. It discusses the project's conceptual contribution - answering Research Question 3.2 - which argues for the importance of reflecting on how the interpretative context shapes scientometric analyses. Developing scientometric methods informed by the interpretative context allows me to study the differentiation of and differences among geosocial research using structural network analysis metrics, descriptive statistics and visual network analyses alike. In addition, it discusses ways scientometrics shaped my interview analyses. While in some cases I used scientometrics to study relational practices first hypothesised through interviews, in other cases I used scientometrics inductively and produced findings I did not hypothesise through interviews.

Chapter 2

Literature Review

This chapter serves two main purposes: it briefly introduces geosocial research and outlines the project's conceptual framework. Part I introduces geosocial research by discussing its links to geography's 'quantitative' and 'critical' research traditions and providing examples which illustrate its methodological diversity. Parts II and III discuss the project's conceptual framework. Part II re-introduces the research questions and discusses the project's conceptual approach to three concepts fundamental to the project: scientific practice, digital data in the context of academic scholarship and space. It also reviews literature which combines STS and scientometrics - and discusses the evolving relationship between the two fields - which informed the project's mixed methods approach. Part III introduces literature related to the seven main themes discussed in Chapters Four through Eight: the relationship between social science and computational data analysis methods; the aesthetics of science and social media data; links between academic and commercial research; interdisciplinary collaborations; computational search for data patterns; reflexivity in computational research; and the role of local knowledge in mapping.

PART I

The first part of this chapter briefly introduces geosocial research. Section 2.1 discusses geosocial research's links to qualitative and quantitative geography and discusses examples which illustrate its methodological diversity. Section 2.2 discusses three sets of controversies about geosocial data, which show that uses of geosocial data for academic research are far from agreed.

2.1 Examples of Geosocial Research

This section briefly illustrates the diversity of geosocial research situated in the history of geography scholarship (although a detailed review of geosocial research is beyond the scope of this thesis). In the 1950 - 60s, many American geography university departments shifted their focus to 'quantitative geography' (e.g. Harvey, 1969). In contrast to earlier scholarship's focus on detailed descriptions of spaces and places, 'quantitative geography' primarily explored spatio-temporal patterns using quantifiable data with methods developed in physics and mathematics (Barnes and Wilson, 2014). The quantitative methodological tradition was further institutionalised through the development and widespread adoption of geographical information system (GIS) in the 1970s - 80s (Sheppard, 2005). In contrast, 'critical' human geographers developed concepts and methods to reflect on the experiential, political, aesthetic, affective and metaphysical aspects of spaces or places as well as geographic research and cartography. For example, they explore how lived experience (e.g. Tuan, 2001) and practices of governmentality create spaces and places (e.g. Crampton, 2011; Massey, 1994). About two decades ago, based on publication norms, Johnston (2003) argued that British geography was a collection of separate communities writing scholarship for different academic audiences, illustrating the co-existence of diverse methodological approaches.

Since the mid-1990s some scholars have actively been working on closing the 'gap' between the 'quantitative' 'critical' human geography (Sheppard, 2005). Recently, the UK's Economic and Social Research Council also called for fostering dialogues between these research traditions, highlighting the need to emphasize their complementarity and expanding quantitative methods training especially among early career researchers (ESRC, 2013). Geography scholars have recently argued that analysing new digital geographic data - such as social media data - and understanding the digitally mediated situated practices that create them necessitate dialogues between the 'quantitative' and 'critical' research traditions (e.g. Sui and DeLyser, 2012; DeLyser and Sui, 2012). As these examples which aim to close the 'gap' between quantitative' and 'critical' research show, the institutional divide between these traditions in academic geography still exist. In contrast, Mayhew (2011) highlights epistemic similarities and genealogical links between these traditions. He argues that depending on how one narrates the discipline's history, they can be seen either as distinct in origin and opposing, or interrelated and complementary. As section 2.7 will explain, I empirically explore how my participants position themselves in this terrain.

Current efforts to combine 'quantitative' and 'critical' human geography using social media data build on a rich geographic research tradition which reflects on the affordances of computational tools in light of diverse geography theories, sparked by debates associated with the above 'quantitative turn'. Next, I provide a few illustrative examples. Reflecting on the types of data gathered with quantitative and qualitative methods, Madden and Ross (2009) argue that the analysis of remote sensing data in combination with victims' narratives help communicate and study the extent and impact of the atrocities associated with the Conflict in Northern Uganda, and trace internally displaced persons. Reflecting on the certainty or contingency of knowledge claims produced with quantitative' methods and in 'critical' studies - as well as the types of relationships studies in these research traditions focus on - Bergmann, Sheppard, and Plummer (2009) highlight resonances between 'critical geography' and complex systems modeling methods. They highlight similarities by illustrating modeling methods' affordances to depict the flux, emergence and co-production of situated practices, as well as their inherent incompleteness and non-deterministic results contingent on analytical assumptions. They advocate for models' situated use, supported by modelers' practice of tracing the impact of analytical assumptions. Reflecting on the importance of mapping to account for diverse perspectives, Millington and Wainwright (2017) propose that (participatory) agent based modeling could help bridge research traditions given its affordance to represent scenarios comprising the interaction among diverse types of agents - including individuals or collectives. Human geography scholars also explored the affordances of participatory mapping methods to empower minorities and local communities (Pain, 2004).

Geosocial scholars use 'geotagged' social media posts which contain geographic information such as geo-coordinates, place tags or location mentions for academic research. They are relatively new digital information which result from digitally mediated, situated practices (Sui and Goodchild, 2011). The most popular social media platforms that afford them were founded since 2005. Examples include Twitter (launched in 2006), Instagram (launched in 2010) and Panoramio (launched in 2005). As the examples below and my interviewees' work show, scholars from diverse disciplines use 'geotagged' social media posts for research in part because they are multimodal - they contain traces of users' situated interactions depicted through geographic information, text or images.

The heterogeneous modalities of geosocial data coupled with the diverse computational and digital spatial analysis methods (e.g. GIS software, Python and digital mapping platforms such as Carto) produce diverse geosocial research agendas. Some researchers focus on developing computational methods to trace spatiotemporal patterns (e.g. Nikitopoulos et al., 2016) in relation to diverse topics. For example, they study crime (Wajid and Samet, 2016); migration (Simini et al., 2012), vernacular geographies (Brindley, Goulding, and Wilson, 2014); protests (Manovich et al., 2014) or data visualisation (Jia et al., 2016). Others place more emphasis on the social scientific concepts that inform their geosocial research. Shelton et al. (2014) draw on the relational socio-spatial ontology proposed by Jessop, Brenner, and Jones (2008) to analyse diverse sociospatial patterns of tweets associated with Hurricane Sandy that struck the eastern coast of the United States in 2012. Graham and Zook (2013) study digital power inequalities which shape whose voices can be heard through spatial media, through mapping the geo-linguistic inequality of geotagged GoogleMap content in Tel Aviv and Canada. Finally, whilst the above papers study places though their computer screens as 'observers', Boy and Uitermark (2017) combine ethnography or interviews with geosocial data analysis to study digitally mediated everyday practices. They illustrate how nuanced engagement with geospatial data can be curated through a collaboration between a researcher and local communities (cf. Currie et al., 2016).

2.2 Controversies about Geosocial Research

This section illustrates the contested nature of geosocial research through introducing three controversies. Firstly, scholars disagree about the innovative nature of geosocial research. González-Bailón (2013) argues that new forms of geographic data - such as geosocial data - can provide an increasingly nuanced and complex picture about social and geographic processes, not afforded by previous geographic data. In contrast, others emphasize the historical precedents of geosocial research. Barnes and Wilson (2014) argue that big data analytics in geography echoes ideologies and epistemologies associated with the quantitative turn of the 1950s and 1960s. Dalton and Thatcher

(2015) argue that geosocial data analysis is a result of developments in geodemographic research that intended to profile customers on an increasingly granular scales.

Secondly, scholars disagree about the epistemological characteristics of geosocial research. Some emphasize the advantages of such tools. González-Bailón (2013) argues that the spatial and temporal granularity of such data can help advance human geography theories, models and maps, and better understand the "multiple, nested layers of social life" (p. 293), including "individual and collective" (p. 295) social dynamics. Sui and DeLyser (2012) argue that social media analysis can help cross the divides between quantitative and critical human geography. Others emphasize the limitations of geosocial research. For example, digital social data is often non-representative due to the digital divide in terms of user groups (Haklay, 2012), device ownership (Graham and Foster, 2016) and differences in other geographic data - such as administrative information - available about spaces that can help interpret geosocial research findings (Dalton, Taylor, and Thatcher, 2016). Leszczynski and Crampton (2016) warn that the ubiquity of mapped or mappeable data can result in over-privileging a definition of spatiality as the longitude and latitude coordinates, which disregards the relational aspects of space. Similarly, Shelton (2016) warns that understanding events based on topics that trend on twitter can "promote an understanding of these events as "novel and fleeting" rather than as the gradual outcomes of social inequalities and disaffections rooted in historical geographies. Finally, Graham and Shelton (2013) argue that data driven research in geography might blur the boundaries between epistemological and ontological domains (i.e. due to large sample sizes, researchers might mistake the representation of the studied phenomena for the phenomena themselves).

Thirdly, researchers question the types of agency associated with geosocial data practices (cf. Kennedy, Poell, and van Dijck, 2015). Kitchin (2013) highlights the role of research funding as a motivator to engage with data-rich research in geography, and Dalton and Thatcher (2015) warn that remote geographic studies afforded by geosocial data can strip analysed groups from their agency. In contrast, others highlight the affordances of geosocial research to reconfigure the relationship between the mapper and the mapped. Taylor, Lindley, et al. (2014) and Currie et al. (2016) report their experiences about creating data analysis together with local communities who live in the studied area, and argue that geosocial data practices can become sites for the negotiation of values.

PART II

The second part of this chapter outlines my conceptual framework to studying science as practice, digital data and space, and reviews literature which inform the thesis' mixed methods approach. As section 2.3 explains, I understand science as a set of relational practices. Section 2.5 explains that I assume that spaces - the object of geosocial research - are heterogeneous and brought forth through geosocial research, and digital mapping which has the potential to explore relations on diverse levels of aggregation. Section 2.4 discusses digital data as situated and digital methods' role in mediating research. Finally, section 2.6 explains my conceptual approach to combining STS and scientometrics.

2.3 Science as Practice

This section discusses my approach to conceptualising science as practice and discusses the research questions in light of this conceptual framework. Similar to Hutchins's (1995) distributed cognition framework, I assume that science, and thus geosocial research are achieved through scholars' intersubjective, embodied practices (e.g. Garnett, 2016), mediated by tools such as computational infrastructure and and data (cf. Bowker, 2005; Bates, Lin, and Goodale, 2016; Leonelli, 2016). In addition, similar to Latour (1988) and actor-network-theory (ANT) (Latour, 2005b), I assume that scientific practice comprises scholars' relational practices with diverse actors and collectives, including those that are and are not associated with academic institutions. Finally, I assume that science creates its objects and knowledge through such semiotic-material practices, rather than manipulating objects 'out there'. This applies to both to my understanding of geosocial scholarship and my own research. This framework is contrast with accounts that assume that science is determined by guidelines, procedures that directly 'represent reality'. (cf. Latour and Woolgar, 1986) As I discuss below, this relational framework informed my research questions and the project's methodology.

The case of geosocial research and my efforts to combine STS and scientometrics can also inform research about data driven social science and humanities. The latter fields, and their use of digital tools are largely under explored by STS with a few exceptions (examples STS studies of social science include Elgaard Jensen, 1999; Elgaard Jensen, 2019; and examples of studies of digital SSH scholarship include Antonijevic, 2015; Wouters and Beaulieu, 2006; Elgaard Jensen, 2020), given STS' predominant focus on the natural sciences, biomedicine and engineering (Wyatt, Scharnhorst, et al., 2013; Elgaard Jensen, 2019; cf. Fry, 2006). Characterising data-driven research in the social sciences can foster disciplinary comparisons, using existing research about digital infrastructures in biology (e.g. Levin, 2014; Leonelli, 2016; Hine, 2006), climate science (Edwards, 2010) and physics (Galison, 2011).

Similar to ANT approaches, the First Research Question seeks to study geosocial research by tracing the associations and relational practices through which it is accomplished. To identify units of analyses that help describe my interviewees' relational practices - rather than analysing them through predefined analytical units - I iteratively analysed interviews, conducted scientometrics and reviewed literature. Part II of this chapter reviews literature which informed the way I combined scientometrics with STS concepts. Part III introduces the explanatory concepts I identified through the above iterative process that informed my analyses of interviewees' practices.

As the Introduction explained, the thesis explores the diversity of geosocial research. I use the concept 'geosocial research approach' to refer to different ways geosocial research is practiced, without pre-defining the types of differences I seek to explore. The flexibility the concept 'approach' helped me account for differences in light of participants' narratives and the units identified through mixed methods analysis. For example, my definition of 'geosocial research approach' does not pre-define the extent to which the approaches - the differences in geosocial research I find - are homogeneous or heterogeneous; practiced by larger collectives or small groups of geosocial scholars; are formalised or institutionalised; or relate to methodological, organisational or substantive differences.

Altogether, the First Research Question asks how approaches to geosocial research develop, aiming to explore relational practices through which they are accomplished. A sub-research question asks how collaboration helps their development. As section 2.10 will discuss, collaboration is a relational practice I assumed was important for geosocial research, based on my literature review.

First Research Question

1. How do different geosocial research approaches develop?

1.1 How do scholars from different disciplines collaborate during geosocial research?

The Second Research Question explores differences among geosocial research approaches to study the diversity of contemporary digital or computational social research. I assume that knowledge about geosocial research approaches is comparative: approaches can only be identified or characterised with respect to one another because it is their relative differences which allow me to define them as distinct (cf. Maniglier, 2019 who argues that all critique is comparative). As outlined above, I do not know *a priori* where or what approaches are, and I assume that the approaches I find are contingent on my analytical decisions (cf. Bateson, 1972). Nevertheless, the differences I find help illustrate the diversity of contemporary digital or computational social research. (cf. Knorr-Cetina, 1999). As Knorr-Cetina (1999) argues, comparing research practices across fields can sensitise the analyst to observe and recognise "contradictions, discrepancies, variations, and differences" (p. 22) between settings, without the need to 'generalise'.

Second Research Question

2. How do geosocial research approaches differ?

The Third Research Question relates to this project's mixed methods methodology. As the Introduction outlined, this project's main contribution is its approach to combining STS, interviews, participant observation and scientometrics. The Third Research Question comprises two sub-research questions. Research Question 3.1 reflects on how methods mixing helps explore the development of geosocial research approaches (First Research Question) and their differences (Second Research Question). I summarise my findings for this research question at the end of each empirical chapter. Research Question 3.2 asks how we can assess the affordances of computational methods 'for STS' (in quotation marks because STS is itself diverse), which is the thesis' main conceptual contribution.

Third Research Question

3. How can we combine scientometrics and STS to study the development of and difference among approaches to geosocial research?

3.1. How does mixing methods help study the development of and difference

among approaches to geosocial research?

3.2. How can we assess the suitability of computational methods 'for STS'?

Table 2.1 summarises how the research questions help explore the thesis' five goals outlined in the Introduction. My units of analyses help attain the goals as follows. Studying how scholars from diverse disciplinary backgrounds conduct geosocial research without presupposing types of differences I find helps explore the diversity and diversification of geosocial research inductively. Combining STS and scientometrics iteratively, without choosing methodological approaches (e.g. homogeneous or heterogeneous network analysis; network analysis or descriptive statistics) *a priori* helps explore diverse ways STS and scientometics can be combined.

<table>
<tr><th></th><th colspan="2">Relevance</th><th>Research Questions</th></tr>
<tr><td>1.</td><td colspan="2">Study the diversity and diversification of computational (social) research (geosocial research case study)</td><td rowspan="2">RQ 1 & 2</td></tr>
<tr><td>2.</td><td colspan="2">ESRC's (2013) call to close gap between quantitative and qualitative human geography</td></tr>
<tr><td rowspan="2">3.</td><td rowspan="2">evaluate affordances of computational methods 'for STS'</td><td>Digital STS – 'import' methods</td><td rowspan="3">RQ 3</td></tr>
<tr><td>STS & scientometrics interface</td></tr>
<tr><td>4.</td><td colspan="2">Social sciences' 'compositionist' agenda (cf. Latour 2010)</td></tr>
<tr><td>5.</td><td colspan="2">Social research methods curriculum development</td><td>All</td></tr>
</table>

Table 2.1: Research Questions, Relevance and Units of Analysis

2.4 Digital Scholarship

This section introduces my assumptions about how digital data - including geosocial and scientometric data - mediate research. In line with Venturini, Bounegru, et al. (2018) I consider geotagged social media posts 'digital traces' to emphasize that they result from practices - and are created for purposes - other than academic research (cf. Madsen, 2015). I assume that digital traces are value-laden and political - they are shaped by the material-semiotic practices which produce them (cf. Iliadis and Russo, 2016). For example, the political economies, use practices and algorithms that constitute social media platforms shape traces. Social media traces capture the views or experiences of their users and omit the views of those who do not use them (e.g. Mellon and Prosser, 2017; Malik et al., 2015). Geo-tagged social media traces form a small subset of all posts (Sloan and Morgan, 2015), and are also subject to characteristics or changes of the Application Programming Interfaces (APIs) scholars use to access data (Morstatter et al., 2013) and social media platforms' policies. For example, in June 2019, Twitter stopped allowing tweet geolocation in terms of GPS coordinates (Benton, 2019).

Defining geotagged posts as digital traces helps study how scholars 'stage' them as data for geosocial research. I assume that digital traces become data as scholars perceive research opportunities associated with them, and analyse them - for example, they reformat or query data, and create variables or indicators in light of research questions and conceptual frameworks. Thus, like digital traces, geosocial data are 'never raw' (Gitelman and Jackson, 2013), but are 'always-already interpreted' (cf. Bateson, 1972, p. 15) and imbued with politics.

STS scholars have argued that data, research methods and infrastructures shape, but do not determine research (e.g. Rheinberger, 1997; Marres and Gerlitz, 2016). Data analysis and interpretation are intimately linked: data never speak for themselves, they need to be narrated (Dourish and Cruz, 2018; Muller et al., 2019). An illustrative example of methods' situated use is critical GIS scholarship discussed in section 2.1 which explores GIS methods through the lens of feminist and postcolonialist theories (Sheppard, 2005). At the same time, data's structure, and the information they contain and or omit shape analytical opportunities. For example, hiatuses in police involved homicide data in the United States - such as omitted atrocities and victims' personal information - limit understanding and change efforts (Currie et al., 2016). Marres and Moats (2015) provide another illustrative example, arguing that social media analysts need to reflect on the extent to which their findings illustrate social media platforms' mediation or the issues users post about.

STS scholars have also demonstrated that the introduction of new technologies is often associated with the re-negotiation of professional roles, identities and power relations. For example, Burri (2008) shows that the introduction of new visualisation technologies in radiology in the 1970s prompted radiologists to re-negotiate their professional identities and epistemic authority which came through the authority to create and interpret radiological images.

I assume that scholars perceive the affordances (cf. Gibson, 1979) - opportunities associated with use - of geosocial data and computational data analysis methods differently. Thus, geosocial data and computational data analysis methods - enacted through their use - are multiple (cf. Mol, 2002). These assumptions inform my study of geosocial research, because I study how participants use, create and modify geosocial data and computational data analysis methods, without assuming that data have inherent characteristics (e.g. 'bias') or that methods can only be performed in one way.

My assumptions about the mediation and multiplicity of data and computational methods also inform my own research practice. As Chapters Three and Nine will discuss in more detail, assuming that scientometric data and computational data analysis methods are contingent on their use helps me reflect on the analytical opportunities they afford in light of STS concepts and my interview findings as I combine methods iteratively. In addition, understanding that interpretation, analytical decisions and data infrastructures co-produce findings sensitises me to the importance of illustrating my data analysis process (for example, visually, using screenshots of the software user interface in Chapter Seven) and the contingency of my findings on analytical decisions and the data analysis infrastructure, discussed in section 9.3.2 (cf. D'Ignazio and Klein, 2020).

2.5 Space and Mapping

This section discusses the project's assumptions about spaces - the object of geosocial research. Similar to the above definition of scientific practice, I assume that spaces are multifaceted (embodied, experiential, historic-material) and dynamically come into existence on multiple temporal and spatial scales, such as minute-to minute intersubjective encounters, longer term and larger scale political, economic and institutional practices, climate trends, as well as changes in ecological landscapes and built environment (cf. Massey, 1994). I assume that binaries such as 'nature' versus 'politics' or 'culture' cannot capture the heterogeneity spaces and situated practices, which are at once political, material, cultural and natural. Thus, as section 3.4.3 will explain in more detail, I use the concept 'form of participation' (e.g. participation in commercial and civic practices) to describe differences in how geosocial research approaches study spaces in Chapters Seven and Eight.

I also assume that geographic research, including mapping makes spaces knowable by the mapper and brings forth, or create spaces as opposed to 'truthfully representing' them (Crampton, 2011). All mappings are partial - they highlight some aspects of spaces and obscure others, and make visible spaces 'actionable'. For example, mapping (and censuses) are instruments of state-making which help control territories (Crampton, 2010; Law, Ruppert, and Savage, 2011). In contrast, participatory mapping can help highlight diverse local experiences (Pain, 2004). This project explores how geosocial research creates spaces.

Digital mapping can help flexibly explore data at different levels of aggregation (e.g. Loukissas, 2016; cf. Munk and Elgaard Jensen, 2014). For example, Shelton et al. (2014) explore the tweets associated with Hurricane Sandy which hit the United States in 2012 aggregated at diverse levels and tracing diverse relations. They plot tweets at the country and region levels, study networked connections between tweets at geospatially distant locations and study lived experiences at specific locations through qualitative analysis. Such approaches highlight the practice of mapping and maps' situatedness and mutability (cf. Elgaard Jensen, 2011).

Finally, in line with my assumptions about data's and methods' mediating role outlined in section 2.4, I assume that research methods shape how scholars study spaces. For example, studying academic literature about research on urban "polycentric spatial structure[s]" (p. 1279) Meeteren et al. (2016) argue that research methods mediate knowledge about spaces and differentiate geographic research traditions. In contrast to dominant discourse which divides polcentricity literature based on the scale of the spatial units (e.g. inter-urban and intra-urban policentricity), the authors argue that policentricity research differs in terms of the methods they use and the "social context" of the spatial units they study. They argue that studies explore inter-urban policentricity through types of "actually existing urban region[s]" (p. 1292) mainly using conceptual descriptive analyses and desk research. In contrast, studies of itra-urban policentricity use spatial and regression *models*, and characterise existing places in terms of model fit and explained variance.

2.6 STS and Scientometrics

As the Introduction discussed, this project contributes to literature which explores the STS - scientometrics interface by evaluating computational methods' affordances in light of the *interpretative context* - including the characteristics of the research practice under investigation, the research questions, conceptual frameworks and previous findings - and combining them in diverse ways as part of a single case study. This section discusses the relationship between these two research traditions and reviews literature that combines them, which informs this project's mixed methods approach. As section 2.6.1 explains, STS and scientometrics have been growing apart in the past decades, but recently there has been growing interest in combining them. As section 2.6.2 argues, most studies which reflect on the affordances of scientometric or other computational methods for STS assess methods in light of theories, such as ANT or ethnographic theory, and study biomedical research with mixed methods. Finally, section 2.6.3 discusses literature which combines STS or social research methods (such as interviews) and scientometrics or network analysis. As the examples will illustrate, studies which combine STS and scientometrics primarily explored the biomedical sciences.

2.6.1 The Changing Relationship of Scientometrics and STS

This section briefly outlines the origins and historical connections of scientometrics and STS. Science has been studied through quantitative analysis of scientific literature for over a century (Hood and Wilson, 2001). For example, a century ago Gross and Gross (1927) recommended that college libraries should obtain periodicals through analysing references in a leading chemistry journal. The development of scientometric databases in the 1960s, such as the Science Citation Index was a turning point in the field. Over time, multiple concepts have been introduced to refer to quantitative studies of published literature, including scientometrics, bibliometrics, informetrics (Hood and Wilson, 2001) and more recently science of science (Fortunato et al., 2018) and quantitative science studies (Waltman, Larivière, et al., 2020). This project uses the term 'scientometrics' to situate its analysis of scientific publications in the intellectual tradition of scholarship discussed below which developed dialogues between STS and scientometrics over the past decades. However, it is important to note that this project's scientometric analyses are informed by STS' constructivist school of thought (cf. Cambrosio, Bourret, et al., 2014).

Although the philosophical study of science has a long past, Science and Technology Studies (STS) has a relatively short history (cf. Farr, 1991). Like bibliometrics, scientometrics and informetrics, the term STS was coined around 1970: in academia it became the label for studies of the politics of science and technology understood "as problematic social institutions" (p. 10) and collective practices (Sismondo, 2010). STS is inextricably linked to anti-nuclear political activism of the 1970s, committed to promoting responsible science and technology. It also reacted to analytical philosophical studies of science and the normative perspective of Mertonian sociology of science developed in the 1940s which argued that the norms of Communalism, Universalism, Disinterestedness, and Organized Skepticism ('CUDOS') distinguish science from other activities. (Wyatt, Milojević, et al., 2017) Generally, STS studies take an anti-essentialist position in relation science and technology. STS rejects understanding science, scientific method or technology as natural kinds or fixed, unambiguous entities. Rather, it highlights their situated, discursive, material and intersubjective construction, emphasizing their inextricable links to culture and politics. It shows that diverse audiences practice and value science and technology differently (Sismondo, 2010). Diverse disciplines such as anthropology, sociology, geography, history and philosophy and governance studies contribute to STS' diversity.

STS and scientometrics share interests in exploring the social and epistemic aspects of science, science evaluation and the science-society or science-policy interfaces. They have cross-fertilised each other since at least the 1970s, and some scholars argue that they share historical origins. Wyatt, Milojević, et al. (2017) locate the origins of both qualitative and quantitative science studies in Robert K. Merton's scholarship in the 1940s. During the early days of science studies between 1960 and 1980, scholars combined quantitative and qualitative methods to explore opportunities associated with the then-new Science Citation Index (founded in the 1960s) (cf. Glaser and Laudel, 2015). For example, in the late 1970s, scientometric studies of the social and cognitive structure of science (e.g. Small, 1977; Lenoir, 1979) or the meaning of citation practice (Small, 1978) were published in Social Studies of Science, a leading STS journal. In addition, ANT scholars experimented with computational methods since the early 1980s (Munk and Elgaard Jensen, 2014) in part to develop techniques for controversy mapping (e.g. Latour, Mauguin, and Teil, 1992) and to trace heterogeneous associations among actors. For example, the founders of ANT were among the pioneers of mapping approaches using co-word analysis (Cambrosio, Bourret, et al., 2014). Informed by the sociology of 'translation', (Callon et al., 1983) developed co-word analysis to study the development and changes in problems scientists explore by relating them. Teil and Latour (1995) argue that term co-occurrence analysis is a "quali-quantitative" method (p. 18) which can achieve both fine grained analysis and aggregation or synthesis to explore large datasets, which helps overcome the divide between micro and macro-scale analyses and qualtiative and quantiative social sciences (cf. Latour, Jensen, et al., 2012). Dialogues between STS and scientometrics continued during the 1980s. Early STS fieldwork benefited from scientometrics. For example, in their seminal study of Roger Guillemin's laboratory at the Salk Institute, Latour and Woolgar (1986) traced the acceptance of scientific findings - and their development into scientific 'facts' - by showing that over time, as 'facts' develop, the original papers which publish them receive less and less citations, but more and more papers mention the 'fact' in their titles (p. 109). Finally, a special issue of the journal Scientometrics published in 1989 explored the links between STS theories and scientometrics mainly to explore how scientometric maps capture scientific practice and the meaning of scientometric indicators used in policy and science evaluation (Leydesdorff, 1989).

However, since the 1980s, STS and scientometrics have been growing apart, each associated with their own journals, conferences and professional identities (Rotolo, Hicks, and Martin, 2015). Scholars name multiple reasons for their separation. Glaser and Laudel (2015) note the difference in the units of analysis their preferred methods enable. They argue that science policy and bibliometrics focus on quantitative methods to study "field-level or system level dynamics" (p. 327), whereas sociology of science "prefers qualitative methods because" they can "provide in-depth explanations of the mechanisms that produce changes in knowledge production" (p. 327) and that shape "field-level processes" (p. 327). Wyatt, Milojević, et al. (2017) emphasize that qualitative STS and scientometrics require different skills and differ in their epistemological assumptions, "research practices, norms and standards" (p. 87). White (2011) argues that in the 1970s two schools of thought started to compete for the attention space of science studies scholars: "American sociology of science centred on Merton and constructivist accounts of science centred on Bruno Latour" (p. 278). According to him, this opposition is partially maintained by the journals, 'Scientometrics' and 'Social Studies of Science', the former more closely associated with the Pricean and Mertonian tradition, while the latter inspired by the constructivist tradition of thought. As Edge and MacLeod (1986) note in their Editorial to Social Studies of Science, most submissions that emphasized quantitative methods got referred to the journal Scientometrics as a result of a "friendly but informal arrangement" (p. 5) among the editors. Noting the journal's breadth, they argue quantitative science studies research is the "only exception" (p. 5) excluded from Social Studies of Science.

Recently, there is a growing interest in combining STS and scientometrics, exemplified by growing number of mixed methods publications and a special issue of the new journal Quantitative Science Studies (QSS) published in August 2020 dedicated to discussing the interface between quantitative and qualitative science studies (Leydesdorff, Ràfols, and Milojević, 2020). This trend might follow in the footsteps of mixed method research's growing popularity in the social sciences in the past two decades, especially since the mid 2000s (Timans, Wouters, and Heilbron, 2019). It also develops in parallel (e.g. the QSS special issue) or dialogue with (e.g. Elgaard Jensen, 2019; Marres and de Rijcke, 2020) the digital STS tradition for example, through scientometrics' use in controversy mapping (Munk and Elgaard Jensen, 2014). Digital STS links a community of STS scholars - and those from related disciplines - interested in studying contemporary uses of digital tools, exploring their affordances for STS scholarship and participating in their design informed by STS' theories. Digital STS has been developing since the late 2000s - early 2010s. Elgaard Jensen (2020) cites the development of the Issue Crawler web network mapping software - introduced by Marres & Rogers in 2008 - as the first and by now classic example of digital STS research and tool development. Vertesi et al. (2019c) note that the digital STS community building started at the 4S conference in 2011 thorough subsequent workshops, conferences, meetings and (online) peer discussions, yielding the DigitalSTS handbook published in 2019 (Vertesi et al., 2019a). Digital STS strives to build bilateral bridges between STS and other "fields that have embraced digital studies and making" (p. 3), which the authors contrast with STS' predominant tendency to 'export' insights to other disciplines, thus building on and extending STS scholarship. (Vertesi et al., 2019b)

My thesis focuses on the use of computational data analysis methods, such as scientometrics and social network analysis 'for STS', which I treat as a subset of digital methods. I differentiate these to highlight my thesis' focus on computational data analysis methods, whereas digital methods can encompass a range of digitally mediated practices. Next, I discuss three main benefits of combining STS and scientometrics.

Firstly, combining methods can help understand large scale, collaborative and digitally mediated research practices such as contemporary biomedicine. On this account, neither ethnography nor scientometrics by itself can capture the heterogeneous and distributed relations that comprise such research practices. (Cambrosio, Bourret, et al., 2014). This argument is similar to earlier ANT programs' use of computational methods to trace the heterogeneity of associations that 'solidify' or 'stabilise' facts, technical objects or cultural features (e.g. Teil and Latour, 1995). Teil and Latour (1995) note that neither quantitative methods (which obscure the type of associations) nor ethnographic studies (which obscure links between cases) can account for such heterogeneous associations.

Secondly, given the increase in digital data which capture situated practices, there is a growing interest in exploring their affordances for the social sciences and humanities - including STS (Cambrosio, Bourret, et al., 2014; e.g. Ruppert, Law, and Savage, 2013). In science studies, the number of digital bibliometric databases has increased since the 1960s. Recent examples include Google Scholar and Scopus, both launched in 2004 (Wikipedia, 2018a; Wikipedia, 2018b) and the free and open access Microsoft Academic Graphs launched in 2016. In addition, 'altmetric' indicators have recently been introduced that capture mentions of scientific outputs in news media as well as Web 2.0 platforms, such as social media, blogs or online reference management tools (Robinson-García et al., 2014).

Thirdly, investigating the strengths and weaknesses of scientometrics methods can help assess the limitations of citation based metrics used for research evaluation (cf. Hicks et al., 2015; Wouters, 2014; De Rijcke, Wouters, et al., 2016) and create alternative evaluation methods (e.g. De Rijcke, Holtrop, et al., 2019).

As the Introduction and section 2.3 outlined, this project mainly combines STS and scientometrics to help explore the diversity of computational or digital social research and how dialogues could be built among ways of knowing. Next, I discuss how existing studies evaluate the affordances of scientometrics 'for STS'.

2.6.2 Evaluating the Affordances of Computational Methods 'for STS'

This section discusses recent literature which conceptually reflects on the strengths and weaknesses of computational data analysis 'for STS' or related fields. Most existing computational STS studies discuss specific methods, and few reflect on the affordances of different combinations of STS and scientometrics. Notable exceptions are three papers this section discusses which reflect on the affordances of computational methods 'for STS' or digital social research in light of epistemological frameworks, such as actor-network-theory (Cambrosio, Bourret, et al., 2014; cf. Venturini, Munk, and Jacomy, 2019) or ethnographic field work (theory) (Munk, 2019). Such reflection is essential given that digital tools may bring epistemologies which differ from STS theoretical sensibilities. As section 9.3 will discuss, I contribute to this line of research by emphasizing the need to develop and evaluate the affordances of computational methods in light of the broader interpretative context. In a recent study published a few months ago Elgaard Jensen (2020) reaches a comparable conclusion, reflecting on two collaborative digital STS projects. I conclude this section by briefly summarising his argument.

Cambrosio, Bourret, et al. (2014) and Venturini, Munk, and Jacomy (2019) reflect on the adequacy of network analysis methods that originate in the physical sciences (e.g. Barabási, 2012) and are used in scientometrics - for "S&TS research agendas" (p. 18). Cambrosio, Bourret, et al. (2014) use network analysis to study contemporary biomedical science, a large-scale collaborative science. Drawing on ANT, they study the changing heterogeneous relations that constitute biomedical research practices - among scholars, institutions and biomedical actors (e.g. molecules and diseases). They seek methods to produce 'surprising' insights about such collective, heterogeneous agencies. They highlight four network analysis tips 'for STS'. Firstly, they call for characterising networks beyond quantitative and structural indicators which, they argue, cannot capture the unfolding of diverse relations highlighted by ANT and related approaches. Secondly, they highlight the importance of heterogeneous network analysis. In contrast with traditional social network analysis which assumes that homogeneous actors interact in a unified social space "within which social ties can be properly measured and described" (p. 20), they argue that heterogeneous network methods can highlight heterogeneous relations or forms of engagement. They argue that clustering (for more details about clustering see section 2.6.3.3) heterogeneous networks is more likely to create surprising insights where homogeneous network clustering yields well-known findings. Thirdly, they call for using multiple data visualisations to "destabilize conventional readings, generate a feeling of analytical strangeness, and record unexpected events" (p. 27). Fourthly, they note the danger of obscuring data infrastructures at play. For example, scientometric networks are primarily derived from the analysis of texts, which contrasts ANT studies' reliance on non-textual information about scientific practices. As a remedy, they call computational STS studies to diversify data sources which highlight human and non-human actors' diverse connections.

Similar to Cambrosio, Bourret, et al. (2014), Venturini, Munk, and Jacomy (2019) assess the affordances of network analysis 'for STS' in light of ANT. Highlighting the importance of not conflating network analyses or visualisations with actor-networks, they note four main differences between them. Firstly, digital traces analysed with network methods cannot trace all aspects of a phenomena and are shaped by the material-semiotic features of the practices that create them. Secondly, they note the discrepancy between graph theory and associated mathematics which treats all nodes and edges as homogeneous and ANT's heterogeneous networks. In graph theory, even weighted nodes and edges have the same type of agency. In addition, graph theory has difficulties depicting 'negative relations', such as oppositional alliances, also key to ANT. Thirdly, they note network analyses' theoretical separation of 'individual nodes' and 'global networks', contrasting their co-constitutive framing by ANT. However, the authors note that empirical network analyses often overcome such binary logic because key properties of nodes depend on overall network topology and vice versa. Finally, like Cambrosio, Bourret, et al. (2014), they note the challenge of providing temporal accounts using formal network analysis, and such methods' focus on "movement through networks" - which hinges on a separation between network structure and content - rather than ANT's focus on the "movement of networks" (p. 517).

Munk (2019) proposes four digital sociology methods to address a methodological question: the 'meaning problem' in digital sociology. Digital sociology uses digital traces to study collective practices: Munk analysed hyperlinked websites to study the Nordic Food Movement in different Nordic Countries. Munk (2019) asks how to analyse and interpret digital traces to gain insights about practices that created them. He compares this to the 'meaning problem' in early anthropology - as Malinowksi noted, recording the details of everyday life does not guarantee insights about lived experiences. Munk argues that the collection of hyperlinks (the 'onlife traces') provide information about associations among actors (e.g. relationship among communities who maintain websites), as well as what and when the communities talk about. However, they do not reveal the meaning of these actions. Whilst ethnographers' embodied presence in the field helps them develop interpretations that reflect participants' experiences, digital sociologists - being removed from the field - have to develop alternative methods. He proposes four such methods.

Firstly, like Cambrosio, Bourret, et al. (2014), he notes the possibility of interpreting network analysis in light of fieldwork. Secondly, referring to ANT and Latour, Jensen, et al. (2012), he argues that onlife traces afford thick analysis on a 'single level', to study how actors are constituted by their relations - blurring the micro/macro distinction. For example, he notes he could study how the connections between a set of tightly linked websites associated with the Nordic Food Movement (which he detected as a 'macro' data pattern) are enabled by diverse sharing practices on the 'micro' level, such as the way users share hyperlinks and write about their products. By showing a 'macro' pattern's contingency on 'micro' practices, he argues that this analysis blurs the micro/macro distinction. Through this example, he reminds the reader of ethnographers' responsibility to define the field themselves, as proposed by Marcus (1995). Thirdly, invoking Richard Rogers' "critical analytics" notion, he advises to carefully select onlife traces and evaluate their meaning in (temporal) context. For example, he manually curated the hyperlinks that form part of his network, rather than collecting links with fully automated web crawlers. Finally, he proposes 'algorithmic sense making': using computational pattern recognition, such as community detection to identify data patterns. He argues that both computational pattern recognition and ethnography can help trace regularities which might contradict established theories. For example, informed by theories, he focused on exploring geographical differences in Nordic Food Movement practices. However, through clustering a network of websites, he found that in some cases, instead of geography, shared thematic interests accounted for websites' connections.

As section 9.3 will discuss, I contribute to literature which reflects on the affordances of scientometric (or other computational) methods 'for STS' by emphasizing the need to develop and evaluate the affordances of these methods in light of the interpretative context, including the research questions, characteristics of the research practice, the conceptual framework and prior empirical findings. In a recent study Elgaard Jensen (2020) reaches a similar conclusion, reflecting on the practice of combining STS and computational methods in two collaborative digital STS projects which involved diverse participants, including computational analysts, STS scholars and issue experts. The first project explored notions of 'environment' in obesity research through analysing and mapping highly cited papers. The second project developed digital methods to visualise ambiguity - rather than clear cut data patterns - using the dataset from the first project. Elgaard Jensen (2020) argues that in both cases, introducing project specific criteria for *evaluating* methods' success - and issue experts' involvement who could provide immediate feedback - was key to establishing trading zones between STS and computational methods. As Chapters Three through Nine will describe, I developed computational methods iteratively, in light of my theoretical and empirical knowledge about geosocial research. As Chapter Nine will argue, studying geosocial research with computational methods also highlighted different affordances compared to computational STS studies of biomedicine. Elgaard Jensen (2020) notes a second practice to develop the trading zone: STS scholars' reflection on the digital object (such as a co-word network map of scientific literature) in "different terms than those of the tool maker" (p. 23). For example, reflecting on digital objects through STS theories can bring attention to practices 'behind' (e.g. underlying notions of obesogenic environments that guide obesity researchers and partially manifest in scientometric maps) or 'before' (e.g. contingency of obesity research approaches on infrastructures and instruments) digital objects. Next, I discuss examples of computational STS studies that inform this project's mixed methods design.

2.6.3 Examples of Computational STS

This section discusses studies which combine STS, other relational social theories or interviews with computational methods, which inform my mixed methods approach, including relationalist social network analysis (section 2.6.3.1); STS (inspired) studies of biomedicine using heterogeneous network analysis (section 2.6.3.2); (STS uses of) network clustering (section 2.6.3.3); and approaches to developing mixed methods units of analyses by aligning them with the aim to 'generate surprises' (section 2.6.3.4).

2.6.3.1 Tracing Homogeneous Associations

This section briefly introduces 'relationalist social network analysis' (R-SNA, for short) and Navon and Shwed's (2012) temporal homogeneous network analysis which informed my homogeneous network analyses in Chapters Five and Six. I study geosocial scholars' collaboration and the differentiation between geosocial research approaches with homogeneous network analysis and interpret these in light of the interviews.

Social network analysis has developed diverse approaches to study homogeneous networks (e.g. people, organisations, countries) since the 1950s (Wasserman and Faust, 1994). Erikson (2013) differentiates between formalist and relationalist social network analysis (R-SNA). Below, I situate my approach to homogeneous network analysis in the R-SNA tradition. Firstly, according to Erikson (2013), in R-SNA, the basic units of analyses are often interactions where the relationship and their meaning are inextricably linked. Similarly, I study collaborative interactions using mixed methods. Through interviews, I explore who participants' (ideal) collaborators are, and why or how they collaborate. Secondly, R-SNA analyses networks in the context of the practices they study, aiming to "interpret one particular setting" (p. 229), in contrast to formalist network analysis which aim to define universal networked patterns. Similarly, I use network analyses to study geosocial research practices: I do not aim to make 'universal' claims about science. Thirdly, while formalist social network analysis assumes the existence of fixed structures and individuals, R-SNA assumes that individuals and collectives are "different manifestations of similar processes", rather than essential categories (Erikson, 2013, p. 233; cf. Latour, Jensen, et al., 2012). Similarly, I assume that collective practices - (approaches to) geosocial research - and geosocial scholars are co-produced through relational practices, such as collaboration. As section 2.3 explained, I study collective practices - (approaches to) geosocial research - through the relational acts that comprise them, and do not assume their fixed existence separate from scholars' practices. Finally, while formalist social network analysis primarily defines actors' agency in terms of their network position, in R-SNA, actors' experiences play a key role in providing an account of their agency. Similarly, I study participants' collaboration experiences through interviews, and do not interpret their agency in terms of their network positions. Altogether, my homogeneous network analyses follow the R-SNA tradition. Next, I discuss Navon and Shwed's (2012) temporal homogeneous network analysis which informed by network analysis which traced the differentiation of geosocial research approaches over time in Chapter Six.

Navon and Shwed (2012) combine structural, statistical citation analysis, document analysis and ethnographic fieldwork to investigate how a genetic mutation – "the microdeletion at chromosomal locus 22q11.2" (p. 1633) – transformed understandings of several rare clinical syndromes, related medical research fields, and gave birth to a new clinical category. They compare changes in the modularity of two networks over time - all papers in their sample *(N1)*, and a network generated using a subset of their data which omits papers which mention the genetic mutation *(N2)*. For statistical baseline which is required when comparing the modularity of networks, they calculate the modularity of 100 randomly generated counterfactual networks whose size equals to *(N2)* but which omit random papers. Network modularity is a network structure metric which expresses the extent to which networks are cohesive or to the contrary, divided or 'modular'. Comparing the modularity of the citation networks which do and do not contain papers about the genetic mutation allows the authors to determine whether research about the genetic mutation changes network structure. They argue that an increase in modularity indicates that divisions between the communities who pertain to their field have become more salient, whereas a decline in modularity indicates their convergence. Their results suggest that the genetic mutation 'holds the literature together', acting as a boundary object and enabling the emergence of a qualitatively different research field. This paper informed my homogeneous analysis discussed in Chapter Six and outlined in section 3.4.4, which traces the differentiation of geosocial research approaches. Next, I discuss STS uses of heterogeneous network analysis.

2.6.3.2 Tracing Heterogeneous Associations

This section discusses three key studies of (biomedical) science which use statistical or visual heterogeneous network analysis informed by STS insights about heterogeneous agencies. These studies inform my heterogeneous network analyses in Chapter Eight which studies how methods' mediate geosocial research.

Shi, Foster, and Evans (2015) develop statistical methods to predict the development of biomedical research by studying how new associations form among heterogeneous actors, including scholars, methods and molecules. Using MeSH terms (a standardised vocabulary to index life science literature available through the MEDLINE database), they extracted the chemicals or diseases papers study, the methods they use, and their authors. Their analysis is inspired by ANT's focus on heterogeneous agencies, and Herbert and Simon's "garbage-can model" of decision making which assumes that solutions that ""stick to" nearby problems are deemed successful" (Shi, Foster, and Evans, 2015, p. 74). Using network statistics generalized to hypergraphs (to depict heterogeneous relations) they argue that the majority of new links form among actors within close graph distance. Their analysis foregrounds network structure - they study new associations among nodes based on their network position - and fixes the type and identity of actors over time: actors' type (e.g. a chemical, method or author) is defined by their fixed classification. They conclude that heterogeneous network analysis increases the prediction of new associations superlinearily (at a rate above linear) because actors "connect things through other types of things" (p. 84).

In contrast to Shi, Foster, and Evans (2015) who use statistical, structural network analysis, Bourret et al. (2006) and Cambrosio, Keating, and Mogoutov (2004) analyse heterogeneous networks of biomedical or health research visually to study how associations among human and non-human actors 'hold together' or enable these collective, collaborative and distributed research practices (cf. Weisz, Cambrosio, and Cointet, 2017). They study empirically and theoretically demarcated, collaborative practices. Cambrosio, Keating, and Mogoutov (2004) investigate the role of research academic and commercial institutions and biomedical substances in the development of a standardised nomenclature for immune cell-surface markers, through analysing paper submissions at relevant international workshops. Like Navon and Shwed (2012), they used heterogeneous maps as part ethnographic research. Cambrosio, Keating, and Mogoutov (2004) highlight three roles the heterogeneous maps played. Firstly, they helped structure and visually illustrate their argument. Secondly, they helped elicit information when used as interview prompts. Thirdly, they helped trace the evolving relationships among a large set of actors - the research institutes and the biomedical substances they produce. This, they argue, can help explore the role heterogeneous actors play in collaborations when "the structure of" collaboration is less codified than in their case study (p. 357).

Bourret et al. (2006) use heterogeneous maps - alongside homogeneous co-authorship and thematic maps, as part of ethnography and as discussion prompts - to study ”the founding and development of a French bioclinical collective - the Groupe Génétique et Cancer (GGC). GGC coordinates most actors in French ”cancer genetics and operates simultaneously in the clinical, research, and regulatory domains” (p. 432). They use heterogeneous maps to interrogate how researchers and non-human entities that papers discuss (e.g. breast cancer, gene mutations) come into relation and coordinate the development of GGC, using network representations of the collective at different time points in two ways. Firstly, they ask how one can characterize the agency of actors in these relational practices by considering the dimension characterized by the following two extremes: on the one hand, humans being connected to “a multitude of themes”, or on the other hand, “a single theme link[ing] a number of actors. Real situations are, of course, located somewhere between these two ideal types” (p. 459). Secondly, they visually explore the networks' structure and ask, for example, whether there are hubs - “obligatory passage points” - in the network, and if the edges are more or less “homogeneously distributed” (p. 459). They show the changing importance of human and non-human agencies in the GGC collaboration. The establishment of GGC fosters collaboration among a set of researchers, and coordinates three research efforts existing at the time. However, they argue that over time, non-human actors - such as specific diseases - play an increasingly important role in coordinating the biomedical collective.

The computational STS papers discussed in this section use statistical or visual heterogeneous network analysis to study how human and non-human actors coordinate large scale, collaborative, distributed scientific practices such as biomedical research. These studies inform my heterogeneous network analyses discussed in Chapter Eight. However, my heterogeneous network analyses differ in two main ways. Firstly, I use them to study how methods mediate geosocial research (on smaller scales) rather than how heterogeneous actors coordinate large scale research. Secondly, in contrast to studies which use heterogeneous network analyses as part of ethnographic field work, I use them to study aspects of geosocial research that I do not have detailed interview or observation data about. Next, I introduce network clustering (in the context of STS).

2.6.3.3 Community Detection - Identifying Geosocial Research Approaches

This section briefly introduces network clustering, also called community detection, which I use to identify the scientometric footprints of geosocial research approaches in Chapters Six through Eight and illustrate its use in STS studies. Community detection is a popular method to identify clusters of nodes (communities, for short) that are tightly connected to one another and sparsely connected to the rest of the network. Community detection can be computationally complex, and diverse algorithms identify communities differently. In addition, many algorithms are non-deterministic: they yield slightly different results every time they ran, even using the same parameters, and yield significantly different results as a function of input parameters.

Some methods divide networks into a pre-defined number of clusters. However, it is uncommon to know the number of relevant clusters *a priori*. Thus, many methods split networks into clusters inductively. This thesis uses two popular community detection methods: the 'fast and greedy' community detection method proposed by Clauset, Newman, and Moore (2004) and the Leiden community detection algorithm (Traag, Waltman, and van Eck, 2019). They both identify community structure which 'optimises' (computationally optimise the value of a function) the network's modularity. Divisions are optimal if "there are many edges within communities and only a few between them" (Clauset, Newman, and Moore, 2004, p. 1). I use these community detection algorithms because they are implemented in Python and R, and can cluster weighted networks.

The conceptual meaning of network clusters is unclear (cf. Cambrosio, Bourret, et al., 2014; Held, Laudel, and Glaser, 2020). STS scholars and sociologists of science use community detection in diverse ways, acknowledging the uncertainty of their meaning. As section 2.6.2 discussed, Cambrosio, Bourret, et al. (2014) and Munk (2019) interpret the outcomes of community detection in light of their ethnographic findings. Glaser and Laudel (2015) discuss the meaning of clusters with interviewees, using network visualisations as interview prompts. Elgaard Jensen et al. (2019) identify clusters in a network of influential papers about obesity research to explore types of obesogenic environments researchers study. They interpret clusters with the help of qualitative analysis of the papers they contain, informed by issue expert's domain knowledge.

As Munk's (2019) example discussed in section 2.6.2 showed - who identified thematic links among websites associated with New Nordic Food, instead of the geographic links he hypothesised - community detection can help generate 'surprising' findings. However, as Elgaard Jensen (2020) notes, it can also create the false impression that clusters are homogeneous, and over-emphasize their separation. In follow-up study to Elgaard Jensen et al. (2019), Elgaard Jensen (2020) explored terms that connected the clusters of a network of terms associated with obesity research to illustrate the ambiguity of differences highlighted by network clustering. Terms that connected clusters highlighted diverse types of relations among paper sets, such as reference to particular policy areas, shared institutional affiliations or shared use of "research devices" (p. 21) such as twin studies, census data or diverse obesity indices.

My network clustering in Chapters Seven and Eight is similar to the above studies. I identify the scientometric footprints of geosocial research approaches by clustering homogeneous citation network of geosocial paper. I interpret the findings in light of interview analysis and my participant observation; cluster multiple networks to highlight my findings' contingency; and explore similarities among approaches as well as differences. Network clustering, which identifies relative differences among nodes (e.g. papers) in a network helps identify geosocial research approaches because as section 2.3 outlined, I seek to identify approaches in terms of their relative differences and I do not presuppose the number of approaches I identify or their 'scale'.

2.6.3.4 Generating Surprise and Tracing Diverse Meanings with Digital STS

This section briefly situates my mixed methods approach in literature which reflects on the purposes of mixed methods social science. Theorists of mixed methods social science identified diverse ways methods can be combined. In their influential study, Greene, Carcelli, and Graham (1989) differentiate five main purposes of mixed methods studies: triangulation (corroborate findings), complementarity (elaborate or enhance findings), development (use results from one method to develop or inform the other), initiation (discover paradoxes or contradictions, develop new questions) and expansion (extend the breadth or range of analysis). They characterise the study designs associated with each purpose in terms of seven design characteristics: the extent to which methods differ in "form, assumptions, strengths, and limitations or biases" (p. 262); are intended to study different phenomena; have equal importance; study one or multiple case studies; and "are implemented within the same or different paradigm" (p. 264); independently or iteratively; and sequentially or concurrently. Bryman (2006) provides a finer grained categorisation comprising eighteen purposes, overlapping with but extending Greene, Carcelli, and Graham's categorisation.

Both classification schemes differentiate mixed methods study designs based on the extent to which methods are aligned. For example, Greene, Carcelli, and Graham (1989) argue that studies which mix methods for triangulation or complementarity use methods informed by the same paradigm to study the same or overlapping phenomena and align findings. In contrast, studies which mix methods for initiation or expansion do not align analytical units and (may) use methods informed by different paradigms. In some cases, "[t]o maximize the possibility of unlikely findings, mixing paradigms [...] is acceptable and even encouraged" (p. 269).

Mixed methods science studies may also serve different purposes. As section 2.6.3.3 discussed, community detection is non-deterministic, and as section 2.4 discussed, the outcomes of data analysis are situationally contingent. Many scholars who combine interviews, participant observation or STS and scientometrics acknowledge these. Some aim to solve such uncertainties by triangulating methods. For example, Glaser and Laudel (2015) and Held, Laudel, and Glaser (2020) seek ground truth through interviews, which they use to interpret the meaning of network clusters.

In contrast, many STS scholars and anthropologists embrace the inherent partiality of these methods and ask how - and if - they can help generate surprises, or insights about diverse actors' agencies and experiences. As section 2.6.2 discussed, to help generate surprising computational insights, Cambrosio, Bourret, et al. (2014) advocate the use of heterogeneous network analysis. They contrast this to the clustering of homogeneous networks, which they argue may yield "redundant illustration of well-known arrangements" (p. 21). In addition, as section 2.6.2 discussed, they stress the importance of diversifying data sources and creating multiple 'maps' (data visualisations) which may "destabilize conventional readings, generate a feeling of analytical strangeness, and record unexpected events" (p. 27). In addition, as outlined in section 2.6.2, Munk (2019) argues that computational pattern recognition methods can help generate surprises.

Other STS scholars use (participatory) methods to generate insights about diverse actors' agencies and experiences, assuming that values, experiences and issues are co-created by human and non-human actors in situated practices. Using Greene, Carcelli, and Graham's (1989) framework, in these cases, computational methods' use is informed by the 'STS paradigm'. For example, as 2.6.2 discussed, Cambrosio, Bourret, et al. (2014) and Venturini, Munk, and Jacomy (2019) advocate the use of heterogeneous network analyses to explore heterogeneous actors' agencies. Marres and de Rijcke (2020) discuss data visualisations with participants to study how diverse actors understand interdisciplinarity in AI ethics. As opposed to creating baselines or articulating fixed definitions, they seek to identify diverse meanings to de-essentialise understandings of 'interdisciplinarity', and engage with the field's diversity. Similarly, Anderson et al. (2009) discuss visualisations of participants' computer use to study their lived experiences and values associated with computer use. However, in contrast with Marres and de Rijcke (2020), Anderson et al. (2009) use purposefully ambiguous visualisations, arguing that these afford more opportunities to elicit participants' experiences than clearly annotated and self-explanatory visualisations.

As section 9.3.2 will discuss, I aimed to combine scientometrics, interviews and STS concepts for multiple purposes. Analyses in Chapters Six and Seven align scientometric and interview analytical units, use methods to complement each other and develop scientometrics in light of interviews. Other analyses in Chapters Four, Seven and Eight use interviews and scientometrics inductively, benefiting from methods' respective strengths, potentially seeking to produce surprising scientometric insights.

PART III

The third part of this chapter introduces literature related to seven themes I identified through interview analysis and scientometrics, discussed in Chapters Four through Eight: the relationship between computational data analysis and social science (section 2.7); the aesthetics of science and social media (section 2.8); the relationship between academic and non-academic practices (section 2.9); interdisciplinary collaboration (section 2.10); the role of exploration and experimentation in computational data analysis (section 2.11); reflexivity about data analysis process (section 2.12); and the role of local knowledge in mapping (section 2.13). I discuss the themes and associated literature in the order they appear in the thesis.

2.7 Relationship between Social Science and Computational Data Analysis

As section 2.1 argued computational data analysis is core to geosocial research. However, as section 2.2 argued, scholars disagree about whether geosocial research has the potential to facilitate exchange between computational disciplines and social science, and whether 'quantitative' and 'qualitative' geographic research traditions are best understood as distinct, contradictory or complementary. Informed by the STS literature discussed below which highlights the multifarious origins and situatedness of research methods, Chapter Four will empirically investigate my interviewees' views about the contradictory or complementary epistemologies and histories of computational data analysis and social scientific tradition of thought in the context of geosocial research.

STS and geography research have illustrated research methods' diverse origins. For example, tracing the history of the focus group method, Law, Ruppert, and Savage (2011) argue that a version of it "was created in the space between the academy and the state in US media research early in the Second World War" (p. 6). However, it soon disappeared from academic research and only re-appeared in the 1980s. Through these transitions, researchers re-theorised the data focus groups can provide. Earlier accounts assumed it provides information about attitudes, whereas in the 1980s, in line with social scientific research agendas, scholars argued they provide information "about how people negotiate and make positional arguments in contexts saturated by power relations (p. 6). Similarly, Dalton and Thatcher (2015) highlight the connections between geodemographics used for consumer research and the computational methods used for geosocial research.

In the context of social media research, Marres and Gerlitz (2016) call for empirically exploring the similarities and differences between how 'sociological research' and social media platforms - or the practices they mediate - use and enact (computational) methods in specific situations (cf. Lury and Wakeford, 2012). For example, they note that social network and co-word analysis methods platforms employ originate in social scientific inquiry. However, these methods can be performed in diverse ways. Instead of theoretically arguing that computational and social methods *are* or *are not* similar, they call for exploring affinities and differences in methodologies in practice.

As section 4.1 will discuss, my interviewees from diverse fields perceive computational data analysis and social research to be complementary, but sharing different institutional origins. They stated that a core challenge for geosocial research is combining these methodological traditions.

2.8 Aesthetics of Science and Social Media

Chapter Four will discuss participants' narratives about the beauty of geosocial research. Scholars have highlighted the importance of the aesthetics of scientific research and social media data. This section introduces literature that helps interpret interviewees experiences.

Aesthetic norms are inseparable from the epistemic claims scientists make. For example, Daston and Galison (1992) trace the history of changing epistemic virtues in science - historically specific ideals of the scientific method and the scientific self, shared by members of scientific communities that are at once aesthetic and epistemological. They contrast the ideals of 'objectivity', 'truth-to-nature' and 'trained judgment', and explore associated ideals of scientific images. Citing works by philosophers of science, McAllister (2002) notes two contrasting aesthetic ideals associated with science. He argues that "classical, formalist aesthetic" that values "unity, economy, symmetry, consistency" and order is predominantly associated with the mathematical sciences. Similarly, Hossenfelder (2018), a physicist reflecting on her experiences notes the central importance of aesthetic judgement in physics. She argues that Nobel Prize winner Leon Lederman's quote - "We believe that nature is best described in equations that are as simple, beautiful, compact and universal as possible" (p. 35) - still reflects a popular position among physicists, who argue for the accuracy of theories based on aesthetic criteria. McAllister (2002) contrasts this with aesthetics which values complexity, diversity and differentiation, primarily pursued in branches of biology and historical research.

Scholars have also noted the aesthetics of social media posts - mostly exploring the aesthetics of shared images. For example, Miller and Sinanan (2017) highlight the importance of images in social media and explore the situationally contingent - and thus diverse - aesthetics of Facebook posts through ethnographically studying posted images. Schreiber (2017) compares the aesthetics of images shared on different social media platforms. Like Miller and Sinanan (2017), she understands aesthetic norms situationally contingent - specific to the platforms and user (groups). Hochman and Manovich (2013) also focus on the visual aesthetics of social media: study photos posted on Instagram. However, in contrast with the previous studies which understand the aesthetics of social media contingent to their use - primarily through a distant analysis of images with computational data analysis methods - they argue that Instagram "enforce uniform appearances on its photos, thus creating a sense of atemporality and shared aesthetics" (p. 1).

Next, I discuss literature which highlights inextricable links of academic research with non-academic practices.

2.9 Academia - Industry

STS scholars have long argued that academic and non-academic practices are continuous and shape one another (e.g. Schönbauer, 2020; Gingras and Gemme, 2006). For example, Latour (1988) illustrates that scientific research is constituted by relational activities of collectives beyond the confines of academic institutions. He traces the scientific development and 'cultural' acceptance or uptake of pasteurisation. Latour (1988) argues that Pasteur's ideas gained popularity due to the alignment of the interests of these diverse 'societal' groups, including non-academic supporters such as the public hygiene movement and medical professionals, including military physicians and private practitioners. Like Latour, Galison (1997) highlights the inextricable links between academic and 'non-academic' practices. Through studying laboratory practices over the course of 20th century physics, he shows the way industrial-technological and scientific practices are co-constructed. Thus, he emphasizes the need to consider technical-industrial and academic practices on a par and highlight their interconnectedness, rather than positioning one as 'subordinate' to the other, for example, in accounts that position 'technology' as the application of research. In addition, as section 2.7 discussed, scholars highlighted the co-shaping of (social) research methods by academic, governmental and commercial practices. However, in relation to discussing computational social research, over a decade ago, Savage and Burrows (2007) argued that there was a growing gap between academic and non-academic social research. Through the examples of contemporary (social) network analysis at telecommunication companies and geodemographic research - which analysed large, privately owned datasets about situated and interpersonal interactions - they argued that 'commercial' and academic sociology were parallel research practices, the former "largely unknown to academic sociologists" (p. 887).

Funding arrangements shape the relationship between academic and non-academic practices. Recently, due to funding norms, many academic institutions have been actively trying to extend their commercially-relevant research practices (Lam, 2010) (c.f. Jackson, 2009). Financial pressures and job insecurity have been highlighted as a leading cause of stress among graduate students (e.g. El-Ghoroury et al., 2012; Teachout, 2004) and faculty alike (e.g. Reevy and Deason, 2014; Tytherleigh et al., 2005). Chapter Four explores how funding arrangements impact on interviewees' academic and non-academic activities, and how these, in turn jointly shape their geosocial research.

2.10 Interdisciplinarity and Scientific Collaboration

This section briefly outlines my working definition of the concepts 'institution' and 'interdisciplinarity' and discusses literature about them which informs my interview and scientometric analyses in Chapters Five and Six.

Chapter Six will discuss how my interviewees develop their geosocial data practices through dialogue with their institutions, such as research groups or university departments. Similar to Martin (2004) I define institutions as relatively enduring collective practices which nevertheless are in constant flux as they constrain and facilitate members' activities, rather than predetermining them, such as research groups and university departments. More specifically, Chapters Five and Six explore interviewees' experiences of interdisciplinarity and interdisciplinary collaboration at their institutions. Below I discuss literature about interdisciplinarity that I draw on when interpreting my interviewees' narratives about collaborating with scholars with different disciplinary backgrounds, or using methods pertaining to different disciplinary traditions. The scope of this literature review is limited: I do not intend to provide an exhaustive review of the history or definitions of interdisciplinarity. Rather, I discuss literature which focuses on interdisciplinary research practice.

Although academic disciplines are themselves heterogeneous (Becher, 1990), collaborative research by scholars from different disciplines has been extensively studied because specialised knowledge traditions are organising units of contemporary academia. Citing Peter Galison, Barry and Born (2013) argue that internal differences within disciplines, as well as differences among them facilitate their continuous enactment. Scholars of interdisciplinarity emphasize the historical contingency of current understanding of 'discipline'. Klein (1990) argues that contemporary connotations of disciplinary knowledge as 'specialised' result from the institutional changes in knowledge institutions in the nineteenth century. She notes that calls for integrated knowledge and education accompanied scientific development since the early 20$^{\text{th}}$ century (Klein, 1990). Interdisciplinarity, however, does not simply refer to holistic or integrated knowledge. Rather, it is a situated, normative term. Barry and Born (2013) argue that recently interdisciplinarity has become at once a "governmental demand" (p. 4) - aiming to make science accountable, for it solve 'grand challenges' and foster innovation in the knowledge economy - as well as "reflexive orientation within the academy and an object of knowledge" (p. 4). In addition, in line with Barry and Born (2013) I understand interdisciplinarity not as a fixed entity, but as a practice enacted in specific situations.

In my interview analysis, I draw on four main insights from literature which discusses collaboration among scholars from different disciplines: disciplines' propensity to relate to each other; the importance of disciplines for interdisciplinarity; trading zones; and the construction of scientific problems through encountering difference in how other disciplines formulate 'similar' (relatable) problems.

Firstly, Osborne (2013) argues that although disciplines are heterogeneous, their propensity to interact with other research fields in diverse ways is "part of the[ir] very 'disciplinarity'" (p. 91). He primarily discusses such differences in epistemic terms. For example, through the ethnographic method, anthropology encounters diverse practices, including scientific laboratories, businesses or other cultural settings. Osborne argues that instead of interdisciplinarity, such encounters perform the disciplinarity of anthropology. Economics, on the other hand, relates to other knowledge practices by translating issues - such as climate change - into economic terms by calculating economic costs.

Secondly, Osborne (2013), Maniglier (2019) and Strathern (2004) stress that interdisciplinarity - collaboration, communication and exchange among disciplines - requires disciplinarity, rather than the two being opposites. According to Osborne (2013), disciplines with strong epistemological profiles - "conceptual norms, research paradigms, procedures of formalisation" (p. 83) - can partake in interdisciplinary endeavours the easiest. For example, he argues that "unlike sociology, anthropology, with its basis in ethnographic fieldwork, is likely to thrive in an intensified interdisciplinary world. But that is because anthropology is more not less disciplinary than sociology." (p. 96). Osborne (2013) and Strathern (2004) argue that scholars who participate in interdisciplinary research become the proponents of their own disciplines. Collaborators are sought to contribute "'traditional' knowledge", knowledge which is assumed to be already in place. Interdisciplinary collaboration often prompts scholars to interrogate their own disciplinary identity. Maniglier (2019) argues that transdisciplinarity requires disciplinarity because it entails understanding differences in how other disciplines define terms, pose questions and create explanations, in contrast with one's own discipline. Experiencing, rather than erasing such differences makes possible to understand "what is at stake for each" knowledge practice. On this account, practicing inter- or transdisciplinarity "does not consist in an attempt to take seriously a certain number of real-life issues, [but] points to the introduction of comparative methods across the disciplines." (p. 18) For example, reflecting on a collaboration between computer scientists and sociologists who tried to trial a new technology 'in the wild' in a socio-economically marginalised community, Goulden et al. (2017) warns about the danger of 'going native': one discipline adopting the perspective and concerns of the other. They describe a project where sociologists adopted computer scientists' framing of the project as a technology development exercise. This, however, obscured the complexities of doing fieldwork with marginalised communities and reflecting on their values and concerns.

Thirdly, Galison (2011) argued that shared interest and training in computational data analysis techniques can result in trading zones - local collaboration and knowledge exchange - among scholars. He argues in the 1940s, the cluster of computational, statistical skills and complementary epistemological and ontological commitments associated with computer simulations prompted scholars from diverse disciplines to collaborate and exchange ideas. Physicists, mathematicians, engineers, chemists, statisticians, bomb builders (among others) - each with their own conception about the Monte Carlo simulation method - developed local coordinations ('trading zones'), which allowed them to work towards common goals amid epistemological differences. Later, a new sub-culture emerged around simulation science. Like the above scholars, Galison (2011) stresses that trading zones do not eliminate disciplinary differences. Rather, trading zones enable the possibility of collective practice across differences.

Finally, Maniglier (2019) argues that 'problems' (research questions or topics) are defined as part of scientific practice, rather than science addressing 'problems out there'. For example, "the Newtonian definition of mass as a quotient of force by acceleration defines it by its relation to the notion of force and speed, and conversely" (p. 10). He defines this as an ongoing structuration process. He also notes that problems are formulated through dialogues between ways of knowing, such as disciplines or research traditions. For example, disciplines define problems in light of how other disciplines articulate 'relatable' problems ('similar' or resonating problems expressed with the concepts of different knowledge traditions). For example, he argues that when "mathematicians realize that filmmakers do what they themselves do too, though in a radically untranslatable way, they experience [...] the very problem that they try to address" (p. 17). He argues that a problem can be encountered by understanding the difference between how two disciplines create problems. The problem is created in this very encounter between the two disciplines, the same way "disciplines can be articulated with one another in their divergence" (p. 16).

2.11 Computational Search for Data Patterns

As section 5.3 will discuss, many participants stressed that they try diverse data analysis methods as they search geosocial data patterns. This section introduces literature about three topics - exploratory data analysis (EDA), digital scholarship and experimental data practices - which I use to explore interviewees' practices.

Firstly, my participants' practice is similar to exploratory data analysis (EDA). Jebb, Parrigon, and Woo (2017) define EDA as an "overarching analytical attitude" (p. 267) which aims to identify patterns in the data or its structure. EDA, they argue is context specific. Scholars have named several guiding principles for EDA, such as flexibility, the willingness to find "both unexpected and regular phenomena" (p. 266), attentiveness to 'smaller' patterns, acceptance of the incompleteness findings. Simple calculations and data visualisation have been recommended for EDA. The authors argue that EDA has been conflated with 'data dredging' and 'p-hacking'. They differentiate the two depending on how scholars report their data exploration practices. They stress the importance of EDA for method and theory development, and to ensure that the data collected for research is analysed as much in detail as possible. They note, however, that whilst EDA aims to convey information, the findings identified with EDA are statistically less strong than findings identified using confirmatory data analysis. The 'p-hacking problem' is "limited to occasions where findings uncovered through data exploration are reported as if they had been specified in advance." (Jebb, Parrigon, and Woo, 2017, p. 269).

Secondly, my participants' practices resonate with Mackenzie and McNally (2013) who argue that digital techniques which travel across disciplines and are used to identify regularities or patterns multiply methods. They discuss the case of a data visualisation - a cluster-heatmap about proteomics research - which compares methods used to find proteins to explore why the number of proteins identified in the human blood keeps shifting. The heat map depicts the relationship between datasets used to locate proteins, the "laboratory techniques and experimental apparatus used, and the software and databases used to analyse the experimental data" (p. 76). They argue that in the heat map the diversity of methods overruns knowledge about proteins, and highlights the way methods travel across disciplines. For example, proteomics uses information retrieval methods when it seeks proteins as data sequences. By helping to compare the experiments that yielded the proteins, the heat map illustrates and embodies the way the digital multiplies methods. This proliferation of methods multiplies reality - in this case, proteins. In return, when scientists explore the proteins, they have to discuss the methods that yielded them.

Thirdly, I refer to my participants' practices as 'experimental'. Scholars have used the term 'experiment' in relation to 'big data' in three main ways: to refer to experimental study design, experimental method development and experiments in participation. Experimental study design refers to methods where the effects of an intervention are studied on a sample (e.g. research participants) to provide inferences about a population (e.g. a broader group who research participants belong to) (e.g. Kramer, Guillory, and

Hancock, 2014). Some argue that big data enable more and larger scale social scientific experiments than 'traditional' data (e.g. Grimmer, 2015).

Inspired by pragmatist philosophy, Marres and Gerlitz (2016) calls attention to experimentally developing digital social research methods to explore "what makes their deployment productive for social inquiry" (p.23). They develop methods to study the ebb and flow of climate change discourse on Twitter in a way that considers the discourses and the Twitter platform's affordances in tandem, by contrasting hashtag (co-)occurrence methods associated with diverse disciplinary traditions and the Twitter platform itself *in action*.

Finally, Lezaun, Marres, and Tironi (2017) explore how digital methods can help experiment with new forms of participation, where knowledge is created by diverse actors in multiple modalities. For example, they argue that STS-inspired efforts that use private and for-profit digital devices - such as social media analytics - for social research (thus re-purposing them for "more public oriented" (p. 209) inquiry), can be understood as examples of 'experiments in participation'. They use the concept 'experiment' as a heuristic to refer to studying and designing new participation and engagement practices. 'Experiments in participation' refers to experimental practices that open possibilities for creating diverse knowledge forms, often in a way that unsettles conceived definitions of 'participation'. For example, experiments in participation can re-define the goal of participation. For example, instead of devising a solution to a problem they can focus on co-construing possible futures and formulating new problems. Experiments in participation can also re-imagine the participants of experiments. To illustrate, some emphasize the role of everyday objects in mediating participation and everyday practices as settings for participation. They can also re-invent the way participation happens, possibly creating trading zones among traditions of 'participatory' experimentation. For instance, participation can be happen through artistic engagement or open source technology development instead of deliberation.

2.12 Reflexivity in Computational or Digital Research

This section discusses literature which informed my analysis of interviewees' reflexivity about how social media platforms and their analytical decisions shape the knowledge they create about spaces in Chapter Seven. I use the term 'reflexivity' similar to Lynch's (2000) notion of methodological reflexivity. I understand reflexivity broadly as the practice of reflecting on how knowledge is created as part of the research process.

Strathern (2004) argues that the value of disciplines is their ability to account for - or in my vocabulary, reflect on - how they create knowledge. Many computational social research practices use computational data analysis tools to study topics explored by social scientific disciplines. Scholars have highlighted the need to reflect on how computational methods and aggregate categories shape the 'social reality' they create (e.g. O'Neil, 2016; Hanna et al., 2020). I discuss two recent studies which explored the reflexivity practices of computational (social) researchers.

Borges Rey (2017) identifies three types of reflexivity associated with data journalists' knowledge practices: "traditional journalistic reflexivity" (assessing the structure of a news story and its relation to the readers), "computational reflexivity" (appreciating how computational analyses work and where data and tools can be found) and the "hybrid journo-coder mindset" (assessing the adequacy of computational tools for journalistic stories, as well as characteristics of contemporary journalism in light of the computational infrastructures that constitute it). The majority of Borges Rey's (2017) informants emphasize the importance of "traditional journalistic reflexivity" over that of "computational reflexivity", and only a few express "hybrid journo-coder mindset".

In contrast, Neff et al. (2017), who studied data scientists who work on 'data for social good' projects - which aim to use data science to tackle 'societal challenges' - argue that their participants more reflexive about their practices than critical data studies literature purports them. They note that data scientists often acknowledge - at least to some extent - that data are constructed and value laden, especially when data are used for purposes other than the reason for their collection. For example, their participants noted the difficulties of using electronic payment data - which excludes cash transitions, likely used by low-income users - to study access to services. In addition, they note that their participants acknowledge the mediated and interpreted nature of data and models: they actively altered the data models in light of information about the software platforms which afforded them, as well as the goals, preferences and expertise of the customer of their data products.

In sum, Borges Rey (2017) highlights the challenges associated with reflecting both on computational data analysis process and subject area, whereas Neff et al. (2017) emphasize the reflexive nature of data science. Chapter Seven empirically explores the reflexivity of my participants.

2.13 Mapping and Local Knowledge

As section 2.5 discussed, spaces are heterogeneous and co-constituted by diverse situated practices. Using local knowledge during mapping is often seen as a way to diversify the knowledge created through mapping, to ensure that maps capture local perspectives and local knowledge rather than those of the mappers (Pain, 2004). For example, Sunderland et al. (2012) used participatory mapping to study the situated determinants of health, by eliciting residents' local, sensory experiences using go-along interviews and community art projects. Landström et al. (2011) recounts the example of a collaborative flood modeling project. Scientists and local residents collaboratively developed models which drew on residents' experiential knowledge about local environmental problems and their management. The scientists - who thought popular modeling methods restricted the kinds of solutions that could be explored - were interested to co-produce flood models with local residents. Over the course of the collaboration, scientists changed modeling methods to explore the priorities and solutions proposed by local residents - the use of multiple upstream dams - for which their original modeling method could not account.

In contrast with the above examples, the urban data analysts Taylor and Richter (2015) studied work as part of New York City's data analytics team (MODA) and collaborate with other city departments less with the aim to diversify knowledge about places, but rather to validate models and disseminate their findings. For example, the MODA team worked with the city's fire inspectors to validate a predictive model about fire risk. Chapter Eight explores how my participants who study 'specific locations' - such as specific cities - use local knowledge as part of their geosocial research.

Altogether, this chapter reviewed literature that informs my research questions and mixed-methods study design. I outlined a brief overview of geosocial research in light of geography's history, and the divergence between STS and scientometrics that this project's mixed-methods approach discussed in the next chapter combines.

Chapter 3

Methodology

This chapter discusses the project's methodology. I conceptualise geosocial research as a case study which can help explore how different approaches to computational social research develop. I study geosocial research using interviews, participant observation and scientometrics, all informed by STS concepts. I conducted semi-structured interviews with 19 geosocial researchers, participant observation at a 10-day long geosocial research summer school, and scientometrics, including statistical and visual computational data analyses. In addition, I learned computer programming and computational data analysis as part of this project, which helped me better understand interviewees' experiences, and build trust and credibility with them during interviews.

The chapter is structured as follows. Section 3.1 discusses how I developed this case study about geosocial research, combining interview analysis and scientometrics informed by STS. It also reflects on my 'positioning' in relation to the geosocial research 'field site'. Section 3.2 outlines the data collection process, including the way I delineated the interview and scientometric fields. Section 3.3 presents my interview analysis method and scientometric data analysis infrastructure, including the software I used and the disciplinary classification method I developed that underpins most scientometric analyses. Section 3.4 discusses the details of all scientometric analyses used in Chapters 4-7 and the way I combined them with interviews. Section 3.5 outlines ethical considerations, and finally, section 3.6 reflects on the limitations of the project's methodological framework.

3.1 The Mixed Methods Case Study Approach

This section discusses how I created geosocial research as a case study and a 'field site' to explore computational social science using interviews, scientometrics as well as methodological insights from STS and ethnography. As section 3.1.1 explains, exploring geosocial research requires me to define it as a 'field' that I can study. Given the project's mixed methods approach, I discuss the way I constructed this 'field site' in light of STS and social network analysis literatures. Then, section 3.1.2 reflects on my accountabilities and positioning as an STS PhD student researching geosocial scholarship. Finally, section 3.1.3 presents the temporality of combining STS, interviews, participant observation and scientometrics as part of this project, yielding the mixed-methods case study.

3.1.1 Constructing the Field

Case studies are popular in STS and can serve various purposes, such as being ends in and of themselves, or serving as illustrations or "building blocks for theory" (Beaulieu, Scharnhorst, and Wouters, 2007, p. 2). Although some 'general' guidance is available for constructing social scientific case studies (e.g. Starke, 1978), STS research has demonstrated that the case study method is, to a large extent, discipline specific. STS has often used case studies to demonstrate the disunity of science, as well as the diversity of technology's origins, uses and users (Beaulieu, Scharnhorst, and Wouters, 2007). In line with this intellectual tradition, I study the case of geosocial research to contribute to existing efforts which seek to illustrate the diversity of 'data-intensive', computational or digital (social) science (cf. Cambrosio, Keating, and Mogoutov, 2004; Marres, 2017b; Beaulieu and Simakova, 2006). This line of research aims to counter efforts that reify definitions of and values associated with computational social research, such as definitions of 'computational social science' as a field which focuses on the statistical analysis and modelling of aggregate data (e.g. Lazer et al., 2009).

In line with ethnographic research, I reflect on my own role in creating geosocial research as a 'field site' - thus actively bringing it into existence - (cf. Beaulieu, 2010), and my 'positioning' as I write my account of geosocial research (cf. Parker-Jenkins, 2018). The 'field site' is anthropology's archetypal research setting, where the researcher can encounter those researched. I understand field demarcation as an active process:

through analytical decisions, I bring the field of geosocial research into existence (cf. Marcus, 1995; Beaulieu, 2010). Geosocial research is digitally mediated and distributed: scholars across the globe conduct geosocial research through computational data analysis with their computers. Marcus (1995) notes diverse strategies one can use to construct multi-sited spaces "through which the ethnographer traverses" (p. 105) when studying distributed practices, such as following people, things (such as commodities, gifts or art work), metaphors, stories, life histories or conflicts. Beaulieu (2010) argues that to investigate digital practices - and to study scientific practices which cannot be delineated with STS' classical 'laboratory' notion - the defining feature of ethnographic work becomes co-presence rather than co-location, which can mean digitally mediated and offline encounters alike. This brings attention to modes of co-presence. In her work with humanities scholars, Beaulieu (2010) attended lectures and participated in mailing lists, supported by a unified online presence (blog and personal website) she set up to communicate her ethnographic persona to achieve co-presence with her research participants. I could have encountered and delineated the field of geosocial research in many ways. I could have traced the journeys of geosocial data (cf. Bates, Lin, and Goodale, 2016), spent longer periods at institutions which conduct geosocial research (cf. Levin, 2014), traced the history of computational infrastructures (cf. Edwards, 2010; Bowker, 2005) or computational methods (cf. MacKenzie, 2017) geosocial scholars use. Each of these modes of encounter highlight diverse aspects of computational research.

Due to my interest in how diverse approaches to geosocial research develop and differ, and given that geosocial scholars spend a considerable amount of time conducting data analysis with their computer, I did not seek the 'field of geosocial research' as a physical space to visit - as the archetypal 'laboratory' field site in STS or the traditional anthropological field site. Rather, I conceptualised geosocial research as a computationally mediated research practice, where computational infrastructure co-constitute the 'field site'. To encounter this field, and given my interest in combining STS and scientometrics, I explored computational infrastructures my interviewees use - mostly Python and R programming, GIS and databases - as well as reading interviewees' papers and interviewing them. I interviewed scholars based in six countries on three continents of diverse seniority, ranging from PhD students to full professors. Some of the PhD students I interviewed worked alone - a requirement of PhD research at their institutions. Other interviewees conducted geosocial research as part of teams. Regardless of their affiliation, my interviewees spend considerable amount of time in front of their (often multiple) computer screens using a range of software that allow them to access and analyse geosocial data, highlighting the importance of exploring both their narratives and the computational infrastructures that constitute geosocial research.

Social network analysts have also reflected on approaches to delineating social networks - the archetypical units of social network analysis. Laumann, Marsden, and Prensky (1983) differentiate between 'realist' and 'nominalist' approaches to defining the boundaries of networks. In the first instance, the analyst assumes that "a social entity exists as a collectively shared subjective awareness of all, or at least most, of the actors who are members" (p. 65). 'Nominalist' approaches, on the other hand, pose network boundaries that are "analytically relative to the purposes of an investigator, and this network closure has no ontologically independent status." (p. 66). Even though, as the authors argue, many studies do not fall neatly into any of these categories, considering this distinction has methodological significance. Scholars can become part of my geosocial field if they use geosocial data for academic research. Similar to Cambrosio, Bourret, et al.'s (2014) suggestion to interrogate the sociological significance of network clusters, using mixed methods, this project reflects on the extent to which the 'field' can be understood as 'a community' or set of communities.

3.1.2 Concurrent Accountabilities and Subject Positioning

I developed this thesis, co-sponsored by the Horizon Digital Economy Centre for Doctoral Training (Horizon CDT) and Ordnance Survey. Next, I discuss how my professional history and relationship to sponsors shaped my research positionality, and my concurrent accountabilities to the STS and geography communities.

This project was funded by the Horizon CDT, a doctoral training program that sought students interested in digital technology from diverse disciplinary backgrounds - including arts, social sciences, natural sciences, computer science and engineering. Our program involved a first year of taught modules, with an emphasis on group work. Students were required to assemble a supervisory team with at least two scholars from two different departments within the university. After the first year, students joined the departments of their lead supervisor - in my case, the School of Sociology and Social Policy - and the Institute for Science and Society. Throughout my doctoral studies, to my knowledge, I was the only person to conduct computational data analysis among sociology and social policy PhD students within my School. At times fellow PhD students jokingly remarked on the 'weird' things they saw on my computer screen when I worked at my office - computer scripts that looked different from what they were used to seeing - highlighting the strangeness of computer programming not only to my research group, but to the student body of the School of Sociology and Social Policy in general. In the second, third and fourth year of my PhD studies, I assisted with the seminars of two quantitative research methods modules. These were the same modules that most of my colleagues who looked for teaching experience hoped to avoid. Similar to some of Balaban's (2018) interviewees, and in line with Osborne (2013), without a strong knowledge of either of the disciplines I was working with, I often felt homeless and wondered if I would fall between the cracks of institutions. I lacked opportunities to talk with scholars using similar methods that I might have had in a department with a more conducive methodological orientation.

In search for a community which shared my computational or scientometric research interests, in the summer of 2018 I conducted a research visit at the Department for Data and Network Science (DNDS) at Central European University (CEU). Meeting PhD students who also conduct computational social research was a formative experience that helped me feel confident in my interests and validate my journey learning to code. At the same time, the STS aspect of my work was largely alien to the DNDS community. In late 2018 I also met members of the Science and Evaluation Studies group at the Centre for Science and Technology Studies (CWTS) at Leiden University who expressed interest in my research. This gave me confidence in my interest to combine STS and science studies.

My experiences conducting research at the intersection of the interpretivist tradition of Social Studies of Science and scientometric computational data analysis were also shaped by my background. Prior to this research, I completed BA and MSc degrees in Psychology and conducted research in human factors, theoretical cognitive science and marketing strategy. I tangentially encountered Bruno Latour's work through exploring literature adjacent to my Masters thesis which focused on enactive, embodied, ecological and embedded 'cognition' frameworks (informed by my interest in the history of psychology (e.g. Pléh, 2010) and 'cultural' or 'contextual' approaches (e.g. Engeström, Miettinen, and Punamäki, 1999). Reading Bruno Latour's work sparked my interest in STS, but when I started my PhD I was a novice to social science research, STS, computational data analysis and scientometrics. Thus, I was not aware of the long-standing division between STS and scientometrics. However, my experiences reflect their institutional separation. I found it challenging to develop a skill set which combines computational data analysis and interpretative STS research. For the former, I used material available due to computations' commercial relevance and the open source software movement.

Given that my efforts to learn programming served two purposes - encountering the geosocial research field site and conducting scientometrics for this project - throughout the project there was a tension between the amount of time I spent engaging with geosocial research infrastructures compared to developing the mixed methods aspect of this thesis. For the former, I explored social media platforms' APIs, the QGIS software, independently, as well as by auditing a module about spatial computing at the University of Nottingham and attending a relevant summer school in the first year of my PhD studies. As my research progressed, I spent more and more time working on scientometrics, and less time exploring geosocial data infrastructures. The scientometric analyses this thesis presents result from iterative work. Mastering both the technical and conceptual aspects of scientometrics took considerable amount of time, also due to the relative lack of studies that combine STS and scientometrics which can provide templates. A significant body of computational STS literature was published in 2019-2020, towards the end of my research. At the same time, as section 3.3.1 will discuss, my programming skills and journey learning to code myself helped me establish trust and credibility with interviewees.

My concurrent accountabilities to the geography and STS communities (cf. Beaulieu, 2010) were a also a core part of my 'positioning' as a researcher. As an STS scholar, I concurrently seek to trace and facilitate diverse computational data practices, de-essentialise singular notions of 'computational knowledge', as well as empathise with all my interviewees - some of whom benefit from funding associated with discourse about the power of 'computational social science'. I am also part of the academic job market, and at the time of writing my dissertation, I felt that my personal interests in conducting mixed methods research resonated with funding opportunities. I attempted to balance these accountabilities towards the STS community and my interviewees by including findings in my thesis that I feel illustrate the nuances of my interviewees' perspectives and practices, showcase the diversity of geosocial research and (unrecognised) similarities among participants' work from diverse disciplines, whilst also highlighting interviewees' work to create differences that matter to them.

My accountabilities towards the geography community were also brought forth through my relationship with Ordnance Survey (OS). As a PhD candidate co-sponsored by the Horizon CDT and OS, I had to develop a project approved by both sponsors. Throughout my studies, I remained accountable to the geography community through OS, and negotiated OS' focus on geographic analyses and my thesis' STS focus. I completed a 3-month internship at OS (studying organisational change), attended events at the OS headquarter where PhD students showcased their work, and kept in touch with my OS based supervisor. When I presented my work for OS, I mainly discussed geography-relevant aspects. This provided opportunities to reflect on how my interviewees' research highlights aspects of spaces. Next, I discuss the iterative nature of my mixed-methods approach.

3.1.3 Temporality of Mixed Methods Analysis

To develop this mixed methods case study, I combined conceptual STS work, field research, interviews and scientometrics iteratively. In the first year of my PhD studies (2015-2016), in addition to completing the taught element of my PhD program, I started to review literature about geosocial research and identify the controversies outlined in section 2.2. I also audited two geography modules at the University of Nottingham - one about GIScience and one about theoretical human geography - and attended a geosocial research summer school in the summer 2016 as a participant. These experiences helped me develop tacit knowledge and familiarise myself with current debates within geography, and existing approaches to geosocial research.

In the second year of my PhD studies (2016-2017), I focused on reviewing STS literature and developing the first research proposal. In summer 2017, I conducted participant observation at a 10 day-long geosocial data analysis summer school.

In the third year of my PhD studies (2017-2018) I started to teach at the University of Nottingham, learn computer programming and explore scientometric data. I also conducted the research visit at CEU, and started to interview geosocial scholars. I completed the interviews in the fourth year of my PhD studies (2018-2019). That year, in addition to teaching, I also started interview analysis and further developed the scientometric code infrastructure. Between the autumn of 2017 and spring of 2020

I also had increasing informal caring responsibilities due to a close family member's terminal illness, which required me to conduct interviews and research whilst remaining internationally mobile.

Over the course of the 2019-2020 academic year I continued to iteratively develop the conceptual framework of the thesis, interview analysis and scientometrics. Iteration allowed me to develop the mixed methods approach of this project, such reflecting on the units of analysis identified via interviews ad scientometrics. STS studies demonstrated early on that scientific practice cuts across communities associated with scientific specialities (e.g. Knorr-Cetina, 1982) and reaches beyond the laboratories and communities (e.g. Latour, 1987). As section 3.6 will discuss, scientometric data affords a partial perspective on these relational practices and may obscure units which account for scientists' practices. Iterative data analysis allowed me to reflect on the relevance of scientometric units of analyses.

3.2 Data Collection and Field Delineation

This section discusses my data collection and field delineation methods. In the first instance, I delineated the scientometric and interview fields largely independently. I did not limit my scientometric analyses to the scientific output of my interviewees, their research groups, or the social media platforms that they collect geosocial data from. I also did not identify interviewees based on their position in the scientometric field, for example, by selecting scholars in the centre or periphery of the scientometric geosocial field, or scholars with more or less publications than average. Although these methods would ensure a direct link between the two fields, they also would limit my study to a specific subset of geosocial research. Instead, both scientometrically and with interviews I sought to capture geosocial research performed by scholars in diverse disciplines, using diverse social media data, whilst also seeking participants I could interview in person or with whom I felt there was enough common ground (e.g. shared cultural background, overlap in research interests) that I assumed I could establish trust and credibility via phone interviews. The relative ease of identifying social media platforms which afford geosocial data and the stability of their names (e.g. Twitter, Instagram, Panoramio etc.) helped ensure the relevance of both the interview and scientometric fields to geosocial research, without directly aligning them. I analytically account for the partial misalignment between the two fields in two ways. In some cases, I benefit from the misalignment by highlighting the partial perspective provided by each method. In other instances, I conduct scientometric analysis on a subset of the scientometric data, further delineating units of analysis identified through interviews.

3.2.1 Interviewing and Participant Observation

I identified interviewees or in other words, participants through literature review and snowball sampling. In the thesis, I use the term 'scholar' to refer to geosocial researchers I did not interview. I conducted 18 semi-structured interviews with 19 interviewees (I interviewed two participants, David and Daniel as part of one interview session). Interviews lasted between 50 min to 2.5 hours. I transcribed interviews in full, yielding 206520 words. I conducted 11 interviews in person and 7 interviews over the phone between June 2018 and May 2019. In addition, I conducted participant observation at a 10 day-long geosocial research summer school in 2017. Summer school participants included students from diverse disciplines and a few more senior scholars leading the event, interested in studying cities using social media data. Participants were divided into groups of 4-6. During the event, I wrote up field notes every evening (and sometimes during the day). My field notes documented steps of summer school participants' knowledge creation process, as well as my reflections.

I identified three main topics through participant observation. Firstly, I noted the importance of local knowledge for the geosocial research at the summer school. The summer school explicitly aimed to study a city 'remotely' using social media posts. An optional one-day field trip was scheduled in the second half of the 10 day-long event (after groups have already decided their research questions and methods) as an opportunity for break and team building. However, all groups had at least one participant who was familiar with the studied city. Although summer school organisers and participants often discussed the 'remote mapping' social media posts allowed, as section 8.3 will discuss, I observed that locals guided their groups through the development of research questions and interpretation of findings.

Secondly, I observed the diverse expertise required to facilitate interdisciplinary group work and potential differences in the rhythm of work. Fluent users of mapping software (QGIS) or data visualisation tools could build dialogues with participants with less technical knowledge by quickly producing results. However, tasks that required less readily available solutions - or when technical participants' enacted their interest to experiment with new methods - could 'slow down' the research process and leave scholars with non-technical focus without tasks for periods.

Thirdly, I noted the effort required to develop research questions suitable for geosocial research and how theory can guide this process. Each group had to present their project proposal after two days. The summer school organisers gave feedback on the presentations. Their feedback often focused on aiding groups to formulate new questions or alter their existing questions with respect to the affordances of social media data. They also guided participants to focus their attention to relationship between social media data and the built environment, informed by their theoretical framework.

I used my field notes mostly as background knowledge, rather than explicitly citing it (except for the insights about local knowledge). Fieldwork also informed my interviews by sensitising me to the explore the rhythm and role of collaborations, exploring how interviewees from different disciplines perceive the affordances of geosocial data. Next, I discuss my interview approach and interviewees.

Table 3.1 lists participants' pseudonyms, their disciplinary affiliation and academic seniority, as well as collaboration relationships among them ('Group'). The first letter of pseudonyms signal Groups: interviewees in a group share the first letter of their pseudonyms. Participants in Groups A-G have a predominantly social science background. Henry has extensive training in both social science and computer science, and interviewees in groups H-M have predominantly technical background. Anne and Luke work largely alone - required by their PhD programs - whilst others, such as Elias and Henry work collaboratively but are the only ones I interviewed from their academic 'group', because of their focus on geosocial data or availability and time constraints.

I conducted interviews informed by my interviewees practices in two ways. Firstly, similar to Laudel and Gläser's (2007) suggestion to interview scientists informed by their work, before interviews, I read interviewees' papers about geosocial research and customised the interview questions outlined in Appendix A in light of their research - especially about methodological choices interviewees made or collaboration arrangements they were part of. This helped me establish common ground with participants, and discuss the research process and their values associated with analytical choices and collaborations in more detail. Secondly, my computational skills and experience in geosocial research gained through the summer school shaped the interview process.

The extent to which the sociologist of science should have expertise in the knowledge practices they investigate has been debated since the inception of social studies of science (Collins, 2004). An early example of a still-influential ethnographic laboratory study by Latour and Woolgar ([1979] 1986) depicts the prototypical science ethnographer as completely alien to the specifics of the research area. However, Latour was employed as a part-time technician at the lab. Moreover, the authors conducted an extensive analysis of relevant literature, and conducted fieldwork for almost two years, which allowed them to acquire a good understanding of the research speciality. Following science studies ethnographies, such as Lépinay (2011), Levin (2014) and Nelson (2018) were conducted by analysts with (varying degrees of) training in the research fields they investigate (cf. Collins, 2004). Similarly, ethnomethodological participant observation posits that the researcher should have 'vulgar competency' in the observed practices (Lynch, 1993). However, also highlighting the importance of the 'anthropological strangeness' (Bowker, 2010), this research tradition stresses the importance of maintaining an open mind about how research is conducted in practice, being able to question assumptions, practices that scientists take for granted, and understand motives and affects that constitute scientific practice beyond scientists' narratives.

As section 3.1 discussed, I obtained competency in the computational social research through attending geosocial and GIS data analysis courses and through conducting scientometric data analysis. My computational expertise and my experiences learning to code helped me build trust and demonstrate credibility with interviewees, and study the technical aspects of geotagged social media data analysis. For example, Mike shared technical details of their work after I mentioned my computational work. Sharing details about my journey with learning to do programming also allowed me to gain insight into Chase's and Colin's computational work. In addition, I recruited interviewees, including Chase, Colin, Isaac, Kevin, Jane and Miles through connections I made at CEU DNDS. At the same time, I can maintain some level of 'anthropological strangeness' with the help of STS readings, because I am a novice computational analyst, and because scientometric data analysis considerably differs from geosocial research. Scientometric data providers offer standardised textual and numeric data about scientific practices, in contrast to geosocial data's multimodality (e.g. text, geotag, photos, hyperlinks, granular time stamp) and links to everyday practices.

Participant Pseudonym	(Disciplinary) Affiliation	Group
Anne	Interdisciplinary / Anthropology, Human Geography PhD Candidate	A
Ben	Human Geography Professor	B
Brian	Human Geography Postdoctoral Researcher	B
Bruno	Human Geography Postdoctoral Researcher	B
Chase	Economic Geography Scholar	C
Colin	Sociology Scholar	C
David	Sociology Postdoctoral Researcher	D
Daniel	Sociology Scholar	D
Elias	Cultural Studies Early Career Scholar	E
Frank	Urban Studies Early Career Scholar	F
Gary	Ecology PhD Candidate	G
Henry	Computational Social Science Scholar	H
Isaac	Applied Mathematics PhD student	I
Jane	Physics PhD Candidate	J
Josh	Physics Postdoctoral Researcher	J
Kevin	Physics Professor	K
Luke	Interdisciplinary / Computational Geography PhD Candidate	L
Miles	Applied Physics PhD Candidate	M
Mike	Computer Science Scholar	M

Table 3.1: Participants

3.2.2 Scientometric Field Delineation

This section discusses my scientometric field delineation method. I refer to geosocial research related publications matched my scientometric field delineation as *geosocial papers* or *citing papers*, and the papers they cite as *cited papers*. My field delineation yielded 2762 geosocial papers. I excluded all papers matched by my search published before 2008, yielding 2749 geosocial papers. I chose 2008 as a cutoff year because most of the social media platforms I included in my field delineation criteria were launched in 2005-2006. My publication year filter accounts for publication time lag.

Scientometrics have been used to explore diverse units of analyses. Some focus on modelling 'universal' characteristics of scientific work, such as the structure of scientific collaboration networks (Newman, 2001) and 'scientific impact' over time (Sinatra et al., 2016) across diverse disciplines. Other studies focus on specific disciplines (e.g. 'chemistry', 'sociology'), sub-disciplines or research areas, both of which "refer to larger, bureaucratized units of teaching and academic employment" or "research specialties" – "knowledge base generated collectively by a self-organizing, loosely coordinated community of researchers" (Velden and Lagoze, 2013, p. 2426). This project understands the analysis of geotagged social media data as a research specialty. Identifying the (scientometric) boundaries of sub-disciplines or research specialities can be challenging. If a study focuses on the activities of a well-defined group of scholars, literature published by them can be identified using their names as keywords. Analysis can also focus on specific journals (e.g. Wylie et al., 2018), or groups identified by publication based classification (Waltman and van Eck, 2012). However, the constant change in which journals may define specific fields, journals' wide scope and the scattered nature of specialised literature across journals may result in including loosely related articles, and the excluding relevant material (Aksnes, Olsen, and Seglen, 1999; Huang, Notten, and Rasters, 2011). Lexical search can help identify papers whose abstracts and titles match select keywords (e.g. Munoz-Ecija et al., 2013). Approaches can also be combined (Zitt, 2015).

I used a simple lexical search to identify geosocial research related publications by matching the names of social media platforms and geography related keywords outlined in table 3.2. Unlike specialist fields such as nanoscience and nanotechnology (Munoz-Ecija et al., 2013), geosocial research does not have specialist journals associated with it, and as outlined above, I did not wish to limit my scientometric field to papers of my interviewees. I downloaded geosocial research related publications from Clarivate's Web of Science (WoS), using the WoS database owned by the Centre for Science and Technology Studies (CWTS), Leiden University. I chose the WoS database because of its extensive coverage, although it is important to note that WoS over represents the the natural sciences and engineering, and under represents the social sciences, arts and humanities (Mongeon and Paul-Hus, 2016). I collected the data used in this thesis between the 13$^{\text{th}}$ of February 2020 and the 20$^{\text{th}}$ of May, 2020. I use data about citing papers published 2008-2019 inclusive. I downloaded all papers whose title, abstract or author keywords field matched the Boolean search expressions Table 3.2 schematically depicts. I selected all citing papers which contained any of the social media related search terms in the first row AND any of the the geography related keywords in the third row of Table 3.2. I identified the set of keywords iteratively based on information I gathered from reviewing literature about geosocial research and interviews. I omitted the search-terms ‘location’ and ‘space’ because they have general meaning beyond geography.

’”social media”’ ‘twitter’ ‘instagram’ ’flickr’ ’panoramio’ ’foursquare’ ’tripadvisor’ ’yelp’
AND
’city’ ‘geography’ ‘place’ ’geotagged’ ’urban’ ’spatial’ ’GIS’ ’GIScience’ ‘geeoreferenced’

Table 3.2: Field delineation searchterms

I used the following information about geosocial papers for my analyses:

1. WoS' unique publication identifier,
2. title,
3. abstract,
4. author keywords,
5. keywords plus keywords,
6. publication year,
7. source title (title of journal which published paper),
8. WoS 'subject category' name: WoS categorises *journals* into at least one of 254 Subject Categories Clarivate (2020) based on citation analysis and journals' titles (Zitt, 2015). These journal level Subject Categories are assigned to papers,
9. author name identifier,
10. WoS' unique publication identifier of cited references

I collected the following information about the cited references of the geosocial papers:

1. WoS' unique publication identifier of cited references,
2. source title identifier,
3. Subject Category name,
4. author name identifier.

3.3 Data Analysis Approach

This section discusses the project's data analysis approach, including thematic analysis and the scientometric data analysis infrastructure. Section 3.3.1 discusses my approach to thematic analysis, which combined inductive and deductive elements. Section 3.3.2 explains the scientometric data analysis infrastructure, including the software I used and my disciplinary categorisation method that underpins diverse scientometric analyses.

3.3.1 Thematic Analysis of Interviews

I analysed the interviews using thematic analysis (TA). Braun and Clarke (2012) defines TA as a method "for systematically identifying, organizing, and offering insight into patterns of meaning" or common themes across a data set, which are meaningful in relation to the research question (p. 57). The authors note that TA is flexible rather than prescriptive. It requires the researcher to make analytical choices which reflect their theoretical and epistemological orientation - for example, researchers can perform TA more inductively or deductively.

Similar to Braun and Clarke (2012) and Balaban (2018), my TA approach is both inductive and deductive. Both the data collection and data analysis had inductive aspects. The loose structure of the semi-structured interviews allowed participants to shape the direction of the interviews, which helped me collect data that reflected their experiences. I analysed data inductively because as section 2.3 explained, I sought theoretical concepts that helped identify themes which reflected interviewees' relational practices. As section 2.3 explained, I constructed the themes and analytical categories introduced in Part III of the literature review by iteratively coding interviews and reading STS literature.

At the same time, my data collection and data analysis were deductive. As Winthereik (2019) argues, the concepts STS scholars use become part of their "empirical-material" world, and they shape how the researcher "follows" actors of knowledge practices. The framework and concepts about scientific practice, digital scholarship, spaces and cartography outlined in sections 2.3 - 2.5 - and broader readings in STS and human geography sensitised my to potential topics of interest. They informed my data collection, interview questions and data analysis. Although the relational framework to scientific practice outlined in section 2.3 is relatively broad, as the Research Questions show, I focused my analysis on relational practices that I assumed help scholars differentiate approaches to geosocial research, I sought comparative knowledge about approaches and I kept reflecting on possibilities to study types of relational practices scientometrically. As a result, I identified diverse themes (relational practices), rather than exploring fewer themes or practices in depth. Alternatively, I could have focused my interview analysis around theories of identity work (focusing on self perceived identity rather than ANT inspired conceptualisation of 'performed identity' (Elgaard Jensen, 2017), more similar to my study's focus on relational practices) and definitions of interdisciplinary research

- both prominent topics discussed by my participants.

I analysed interviews iteratively. Firstly, transcribing them helped familiarise myself with their content. Secondly, I started to code the transcripts in NVivo, identifying themes of potential interest. While NVivo helped me develop initial themes, I stopped using NVivo after an initial, high level coding of the interviews. Rather, I developed the finer grained interview analysis and interpretations as I iteratively wrote and revised the thesis draft, and re-read relevant sections of the interviews in the process.

I identified themes in interviewees' narratives in four main ways. Firstly, I identified practices that scholars from the majority of research groups mentioned. For example, in Chapter Eight I will argue that interviewees from diverse disciplines use local knowledge when studying specific locations with geosocial data. I aim to describe the ways interviewees perform these practices in detail. For example, local knowledge helps interviewees in diverse ways: in some cases, it helps interviewees critically assess data patterns they find in light of local experiences. In other instances, local knowledge helps interviewees (quickly) interpret their computational findings.

Secondly, I compared interviewees' narratives who I assume develop distinct geosocial research approaches. For example, in Chapter Seven I will argue that technical participants reflect on their analytical decisions and data quality mainly in computational and demographic terms, whereas social scientist participants reflect on them in experiential and historic terms.

Thirdly, I identified themes that interviewees from a subset of the groups mention. For example, Chapter Six will discuss social scientist interviewees' efforts to differentiate their geosocial research from approaches informed by 'computational social science' and GIScience. I did not find comparable themes in the narratives of technical interviewees.

Finally, I identified 'outliers' - opinions that differ from the rest. For example, Josh - a physicist interviewee - stated that local knowledge about the country their geosocial research explored was not necessary for the success of their analysis. This contradicts other interviewees' narratives discussed in Chapter Eight, who highlight the importance of local knowledge for geosocial research. This difference in opinion illustrates that the modularity of computational analysis enables scholars to participate to sub-tasks without participating in the entire geosocial research process, highlighting the importance of interviewing multiple scholars from collaborative projects[1].

3.3.2 Scientometric Data Analysis Infrastructure

This project uses a range scientometric and network scientific techniques, such descriptive statistics, co-authorship, bibliographic coupling and co-word network analyses as well as community detection to explore geosocial research. This section discusses the software used for scientometric data analyses, and the way I develop disciplinary categorisation based on scientometrics.

I used a range of software. Although I initially downloaded scientometric data from Scopus through its Application Programming Interface (API), the data included in the thesis is based on CWTS' WoS database which I later received access to and which required less data cleaning. I accessed this using SQL. For the majority of the data analysis tasks, I used the Python 3.7.4 programming language – in particular, the packages pandas, numpy, re, networkx, igraph, seaborn and matplotlib – through the Anaconda version 3 distribution, using the Spyder version 3 integrated development environment (IDE). At the time of writing this thesis, Python is one of the most popular all-purpose programming languages used for scientific research. Packages are continuously developed and there is a growing user community. For the co-authorship network analysis section will present, I used the R 3.6.2 implementation of the igraph package (Csardi and Nepusz, 2006) using the RStudio Version 1.1.463 IDE.

I used the VOSviewer 1.6.15 software (van Eck and Waltman, 2010) for the term co-occurrence maps presented in Chapters Seven and Eight, and to identify noun phrases in the abstracts and titles of papers used for the analysis described in section 3.4.6. VOSviewer 1.6.15 affords analysing files obtained from major scientometric sites, including WoS. Importantly, it allows importing and exporting data formatted as tab delimited .txt files. I used this functionality to develop iterative workflows. I loaded scientometric data into VOSviewer to extract noun phrases from the titles and abstracts of papers, and saved the .txt files VOSviewer created. I further analysed these in Python - to select subsets of data and calculate metrices - and Excel to aid the thematic analysis described in section 3.4.6). Finally, I visualised the amended networks with VOSviewer.

[1]Due to space limitations, the thesis will not discuss this finding in detail.

Finally, I used Gephi 0.9.2 (Bastian, Heymann, and Jacomy, 2009), VOSviewer 1.6.15 (van Eck and Waltman, 2010) and igraph's R 3.6.2 implementation (Csardi and Nepusz, 2006) for network visualisation. I visualised the journal author-bibliographic coupling network discussed in section 6.3 with Gephi. I visualised the term co-occurrence networks in Chapters Seven and Eight with VOSviewer, and used the interactive, zoom-able maps the VOSviewer user interface affords for the visual network analyses. I used igraph to visualise the heterogeneous network constructed from author-keywords discussed in section 8.2.

3.3.2.1 VOSviewer Noun Phrase Co-occurrence Network Maps

This section discusses how I extracted noun phrases from the abstracts and titles of papers using VOSviewer 1.6.15 (van Eck and Waltman, 2010) for the analyses described in sections 3.4.6, 3.4.7, 3.4.8, 3.4.3 and 3.4.9, presented in Chapters Seven and Eight.

VOSviewer can extract noun phrases from the abstracts and titles of papers downloaded from scientometric databases. I use words 'noun phrase' and 'term' interchangeably to refer to noun phrases VOSviewer identifies. As the VOSviewer manual explains, "VOSviewer defines a noun phrase as a sequence of one or more consecutive words within a sentence such that the last word in the sequence is a noun and each of the other words is either a noun or an adjective. [...] [It] considers only the longest possible noun phrases that can be found in a sentence" (for more detail, see van Eck and Waltmann, 2020, p. 35 - 36).

VOSviewer also allows the user to filter the noun phrases it identifies based on the number of times they occur and their 'relevance score', before term-networks are created. The latter helps exclude 'generic' terms which occur in all papers in the dataset, such as 'study', 'paper', 'result' etc. VOSviewer calculates the relevance score by comparing the distribution of each noun phrase over all noun phrases with the "overall distribution of co-occurrences over noun phrases. The method assumes that for 'general' noun phrases the two distributions are similar. In contrast, terms with high relevance score co-occur with a subset of all noun phrases more often, thus, the two distributions differ. (van Eck and Waltman, n.d.). The VOSviewer clustering algorithm groups noun phrases which often co-occur together, identifying clusters in term-maps. The colors of nodes signal the clusters they belong to. VOSviewer creates weighted term co-occurrence networks: The weighted edges signal the number of times pairs of noun phrases co-occur. The VOSviewer user interface allows users to filter edges based on edge weight values.

3.3.2.2 Disciplinary Categorisation

The majority of scientometric analyses presented in Chapters 4-7 categorise geosocial papers according to disciplines. This section discusses my discipline categorisation method. I categorised papers into Broad Disciplinary Categories in two steps, based on the 254 WoS Subject Categories mentioned in section 3.2.2. Geosocial papers were assigned a total of 193 types of Subject Categories.

Firstly, I re-categorised the WoS Subject Categories into the 'Disciplinary Categories' depicted in Table 3.3. The column 'Search Terms' depicts the a search terms I used to identify Subject Categories relevant to each Disciplinary Category. I categorised geosocial papers into the Disciplinary Categories if their Subject Categories contained any of the search terms in the 'Search Terms' column. Using this categorisation scheme, each citing paper is assigned one or more Disciplinary Categories, depending on the number of Subject Categories associated with them. For example, a paper can be assigned both the 'SOCIAL SCIENCE' and 'COMPUTATIONAL' Disciplinary Categories if the journal that published them is assigned more than one Subject Categories, which contain both the search terms 'computer' (e.g. computer science) and 'urban' (e.g. urban planning). I developed the search terms iteratively, continuing the categorisation process until all but three citing papers were assigned to at least one Disciplinary Category. My categorisation excludes three papers which are only associated with the Subject Category 'art', because of the nature of my substring query. Using 'art' as a substring search term - which matches the 'art' Subject Category - would have also matched the 'artificial intelligence' Subject Category. This is a limitation of my discipline categorisation method. The biases introduced by WoS' selectivity likely have a bigger impact on my findings that the omission of these 3 paper which only account for 0.1% of all geosocial papers.

Disciplinary Category	Search Terms
COMPUTATIONAL	mathematics, engineering, computer, artificial, telecommunications, wireless, cybernetics, remote, statistics, physics, information, optics
SOCIAL SCIENCE	urban, hospitality, tourism, social, sociology, public, politic, area, education, linguistics, cultural, anthropology, communication, religion, criminology, ethnic, international, asian, folklore, archaeology, development studies, law
MULTIDISCIPLINARY, INTERDISCIPLINARY	multidisciplinary, interdisciplinary
GEOGRAPHY	geography
PHYSICAL GEOGRAPHY	"geography, physical"
HUMAN GEOGRAPHY	category developed using the 'GEOGRAPHY' and 'PHYSICAL GEOGRAPHY' categories: all geography which is NOT physical geography
HEALTH	dietetics, nutrition, gynecology, paediatrics, obstetrics, pharmacology, pharmacy, clinical, psychiatry, immunology, medicine, urology, genetics, infectious, health, oncology, nursing, surgery, psychology, anesthesiology, substance, gastroenterology, hepatology, hematology, cardiac, otorhinolaryngology, dermatology, respiratory, audiology, gerontology, veterinary, sport, anatomy, endocrinology, neurosciences, pathology, rehabilitation, medical informatics, rheumatology
BIOLOGY, ENVIRONMENT	Environmental, zoology, biology, biochemical, ecology, meteorolgy, marine, geochemistry, water, oceanography, biodiversity, forestry, green, food, horticulture, entomology
ARTS, HUMANITIES	history, literature, medieval, philosophy, literary, womens, theatre, music, film, television
ECONOMICS, BUSINESS, TRANSPORTATION	economics, business, management, fuels, transportation, ergonomics, architecture

Table 3.3: Disciplinary Categories and Search Terms

Secondly, using the Disciplinary Categories outlined in Table 3.3, I further categorised citing papers into Broad Disciplinary Categories in four ways - corresponding to Discipline Categorisation Methods 1-4 depicted in Table 3.4. These yield the Broad Disciplinary Categories Chapters 4-7 use. As Table 3.4 shows, the only difference between the four methods is the composition of the 'social' Broad Disciplinary Category. I created four alternatives for the 'social' Broad Disciplinary Category to ensure that my results hold constant when using different definitions of 'social science'. With Method 1, the 'social' category includes citing papers published in journals assigned to the 'SOCIAL SCIENCE' Disciplinary Category in the first step. With Method 2, the 'social' category includes citing papers published in journals assigned to the 'SOCIAL SCIENCE' and 'ARTS, HUMANITIES' Disciplinary Categories in the first step. With Method 3, the 'social' category includes citing papers published in journals assigned to the 'SOCIAL SCIENCE' and 'ECONOMICS, BUSINESS, TRANSPORTATION' Disciplinary Categories in the first step. Finally, with Method 4, the 'social' category includes citing papers published in journals assigned to the 'SOCIAL SCIENCE', 'ARTS, HUMANITIES' and 'ECONOMICS, BUSINESS, TRANSPORTATION' Disciplinary Categories in the first step. Thus, Method 1 uses the narrowest definition of 'social' Broad Disciplinary Category, and Method 4 uses the broadest definition of the 'social' Broad Disciplinary Category.

Discipline Categorisation Method	Broad Disciplinary Category	Description of Broad Disciplinary Category	Composition of disciplinary category
Method 1	'social'	Social science	'SOCIAL SCIENCE'
Method 2	'social'	Social science	'SOCIAL SCIENCE' + 'ARTS AND HUMANITIES'
Method 3	'social'	Social science	'SOCIAL SCIENCE' + 'ECONOMICS, BUSINESS, TRANSPORTATION'
Method 4	'social'	Social science	'SOCIAL SCIENCE' + 'ARTS AND HUMANITIES' + 'ECONOMICS, BUSINESS, TRANSPORTATION'
All	'computational'	Technical, computational sciences	'COMPUTATIONAL'
All	'multi-inter'	Multidisciplinary, interdisciplinary	'MULTIDISCIPLINARY INTERDISCIPLINARY'
All	'all_geo'	All geography	'GEOGRAPHY'
All	'phys_geo'	Physical geography	'PHYSICAL GEOGRAPHY'
All	'non_phys_geo'	Human geography	'HUMAN GEOGRAPHY'
All	'health'	Health sciences	'HEALTH'
All	'biol_env'	Biology and environmental sciences	'BIOLOGY, ENVIRONMENT'
All	'arts_humanities'	Arts and humanities	'ARTS, HUMANITIES'
All	'econ_bus_trans'	Economics, business and transportation research	'ECONOMICS, BUSINESS, TRANSPORTATION'
All	'comp_soc'	Technical and social sciences (intersection between the two)	'SOCIAL' + 'COMPUTATIONAL'
All	'only_social'	Only social science	'SOCIAL SCIENCE' + 'HUMAN GEOGRAPHY' – all other disciplinary groups
All	'only_computational'	Only technical or computational sciences	'COMPUTATIONAL' + 'PHYSICAL GEOGRAPHY' – all other disciplinary groups

Table 3.4: Broad Disciplinary Categories and Classification Methods

Finally, a note on term use: in Chapters Four through Nine, I use the term 'social science' to refer to interpretive branches of social science, such as (critical) human geography, STS, anthropology - cf. my interviewees with social science backgrounds. In addition, I use the term 'computational social science' in line with Lazer et al. (2009) to refer to emerging field comprising the analysis of 'social' digital traces by scholars with background in technical and natural scientific disciplines, such as physics, mathematics and engineering. Thus, my use of this term accounts for its politics (cf. Marres, 2017a).

3.4 Mixed Methods Analyses

This section discusses each scientometric analyses presented in Chapters 5-7 and their relationship to interview analysis. While Chapter Two discussed the mixed-methods analyses conceptually, this section discusses their methodological details. I present them approximately in the order they appear the thesis and as a function of the research questions they help explore. Figure 3.1 depicts each scientometric method with respect to the research questions they help study. The colors indicate the type of method, differentiating between statistical and visual network analyses, descriptive statistics as well as methods that combine descriptive statistics and thematic analysis.

Firstly, I discuss three methods that help study practices that help differentiate approaches to geosocial research (the First Research Question) used in Chapters Five, Six and Eight. Section 3.4.1 presents the co-authorship network analysis I use to study collaboration among geosocial scholars. Section 3.4.2 discusses the line graphs that explore the disciplinary composition of geosocial research and, when combined with interviews, geosocial scholars' belonging to diverse research communities. Section 3.4.3 explains two heterogeneous network analyses methods that allow me to study how computational methods mediate knowledge about spaces, a practice that shapes the development of geosocial research approaches.

Secondly, I discuss two methods - used in Chapters Six, Seven and Eight - to identify approaches to geosocial research to help explore the Second Research Question. Section 3.4.4 presents a temporal citation network analysis used in Chapter Six that helps me identify two approaches - social and technical - first hypothesised through interview analysis. Section 3.4.5 presents a static network clustering method I use to identify approaches to geosocial research inductively, using scientometrics, in Chapters Seven and Eight.

Thirdly, I discuss four methods I use to compare approaches to geosocial research to help study the Second Research Question. Section 3.4.5 discusses the bar charts I use to explore the disciplinary composition of network clusters which I hypothesise correspond to the scientometric footprint of geosocial research approaches. Section 3.4.6 presents the combination of term-frequency inverse-document frequency analysis and thematic analysis I use to compare the substantive foci of these two approaches in Chapter Seven. Section 3.4.7 discusses how I study similarities between the scientometric footprint of geosocial research approaches using the occurrence of shared terms. Section 3.4.8 presents a term co-occurrence network analysis (the 'modified ego networks') method I use to compare the ways social and technical approaches study spaces, and highlight the construction of my scientometric analyses.

Finally, section 3.4.9 discusses how I explore differences in geosocial research using term co-occurrence network analysis in Chapter Eight. As illustrated in the bottom right box of figure 3.1, this helps compare approaches to geosocial research in new ways - highlighting their diversity - and at the same time, illustrate the contingency of my scientometric findings on the data analysis infrastructure.

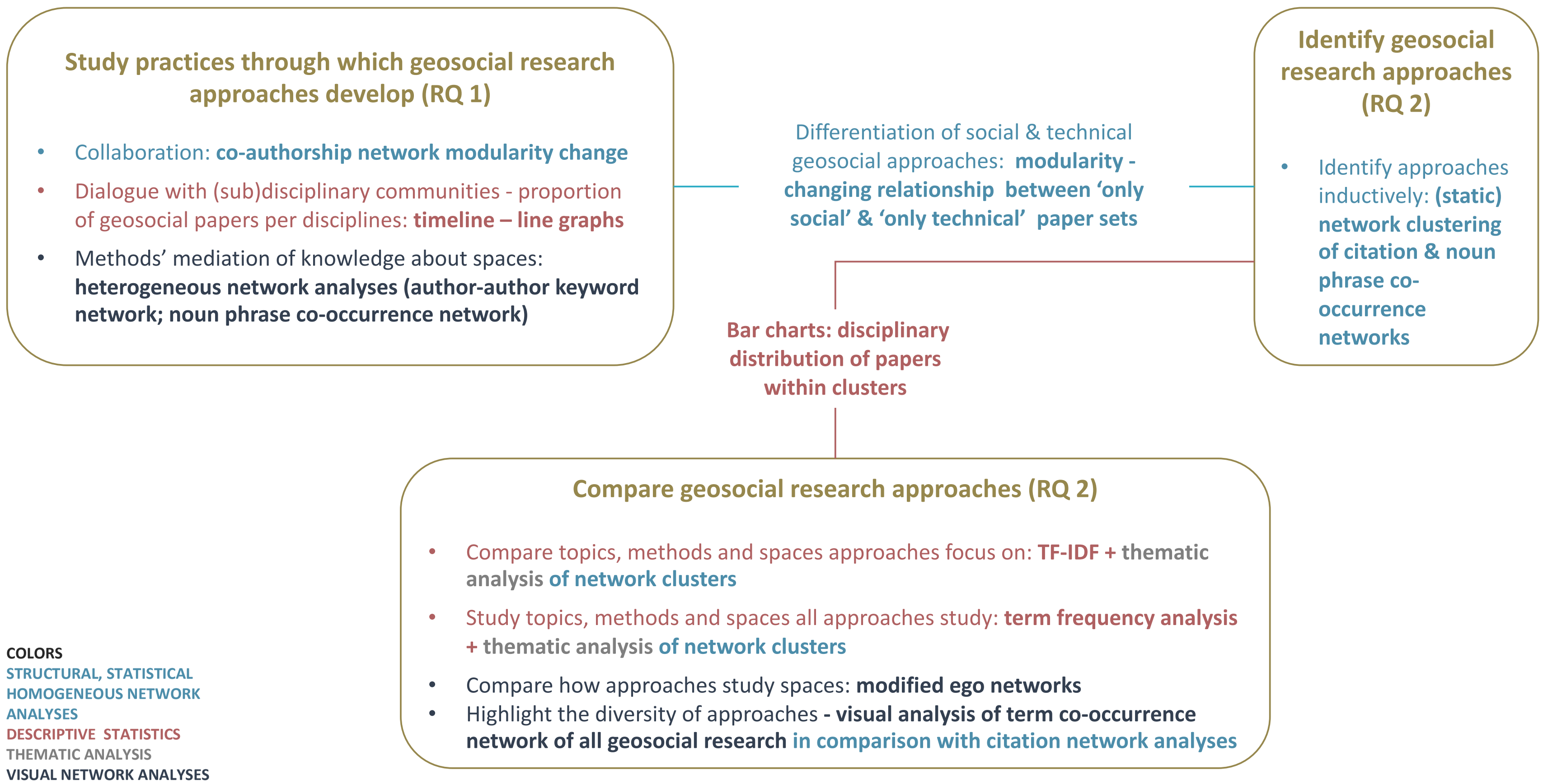


Figure 3.1: Scientometrics and Research Questions

3.4.1 Co-authorship Network Analysis - Tracing Homogeneous Associations

I analysed collaboration among geosocial scholars by combining interview analysis and co-authorship network analysis. Through interview analysis, I studied participants' motives for seeking collaborators, how they contribute to collaborations and the skill set of their ideal collaborators. Scientometrically, I analysed the co-authorship network of all geosocial papers. I created weighted co-authorship networks for four time periods - t1 = 2008 - 2013, t2 = 2008 - 2015, t3 = 2008 - 2017, t4 = 2008 - 2019 - where nodes are individual authors, and the edges among them is a function of the number of papers they co-authored. I first created a n X n matrix M where n = number of authors, and each value in the matrix depicts the number of papers each pair of authors co-authored. Then, I calculated the cosine similarity value for pairs of authors as explained in Equation (3.1) as an indicator of their co-authorship relationship (cf. Glaser and Laudel, 2015; for a discussion see Leydesdorff, 2008).

$$cosinesimilarity = \frac{U \cdot V}{||U||||V||} = \frac{\sum_{i=1}^{n} U_i V_i}{\sqrt{\sum_{i=1}^{n} U_i^2}\sqrt{\sum_{i=1}^{n} V_i^2}} \tag{3.1}$$

where U and V are column vectors of matrix M - each representing the co-author relations of an author -, and $\mathrm{U_i}$ and $\mathrm{V_i}$ are components of vectors U and V respectively. Cosine similarity values range 0-1 and express the cosine of the angle between column vectors in matrix M, thus quantifying the extent to which authors co-author papers in a normalised way. The cosine similarity values provided the weight of the co-authorship network edges.

I compared the modularity of the co-author network at each time point with the modularity of random Erdős-Rényi graphs with equal number of nodes and edges. Network modularity is a network structure metric which expresses the extent to which networks are cohesive or to the contrary, divided or 'modular'. Quantitative network attributes, such as modularity, transitivity and diameter depend on network size. Thus, unlike many descriptive statistical indices, such as median or variance, network attributes cannot be used to compare networks of different size. Thus, in the co-authorship analysis and the analyses discussed in section 3.4.4, I compared the modularity of scientometric networks with randomly generated networks with equal number of nodes and edges. The Erdős-Rényi graph model provides a method to generate random networks. There are two variants of it. In one version, edges among n nodes form randomly with probability

p. In the other version, m number of edges form randomly among n nodes. I used the second variant to generate random graphs whose number of nodes and edges equal to the number of nodes and edges of the co-author networks for each time frame. I used the second version of R igraph's implementation of the Erdős-Rényi random graph model (Csardi, n.d.[a]). I also controlled for edge weight by assigning the edge weight vector of the co-authorship networks to the edges of the random Erdős-Rényi random graphs.

In order to calculate the network's modularity score, I first detected communities - sets of nodes that are tightly connected with one another, and loosely connected to nodes in the rest of the network - within the co-authorship network. For this analysis I used R igraph's fast and greedy community detection algorithm (Nepusz and Csardi, n.d.) which implements Clauset, Newman, and Moore's (2004) community detection algorithm, suitable to detect communities in large networks. I then calculated the modularity score using R igraph's modularity function (Csardi, n.d.[b]) which calculates how modular a graph is given division of a graph into subgraphs (clusters). Networks with high modularity have dense connections among nodes within communities but sparse connections among communities. I then compared to the modularity score of each co-authorship network to that of 1000 randomly generated Erdős-Rényi graphs with equal number of nodes and edges, controlled for edge weight.

3.4.2 Disciplinary Timelines of Geosocial Research

I explored disciplinary contributions to geosocial research. Through interviews, I studied participants' perception about the popularity of geosocial research in their disciplines. Scientometrically, I studied the proportion of geosocial papers per Broad Disciplinary Categories. I created three line graphs (discussed in Chapter Six) which depict the percentage of geosocial papers associated with Broad Disciplinary Categories over time. I use the line graphs in two ways. The line graphs presented in figures 6.2 and 6.1 help reflect on interviewees' experiences about the popularity of geosocial research in their disciplines. Both interviewees and the line graphs suggest the relative increase in social scientific approaches to geosocial research. The line graph presented in figure 6.3 helps study similarities and differences between the composition of the ethnographic and scientometric fields. For example, I find that while the scientometric field includes health related geosocial research, none of my interviewees focused on health geosocial research. Below I discuss the line graph method in more detail.

All three line graphs presented in the thesis use the Discipline Categorisation Method 4 depicted in Table 3.4, - the broadest definition of the 'social' Broad Disciplinary Category. However, the observed trends hold using all Discipline Categorisation Methods. The three line graphs calculate the cumulative percentage of citing papers differently. I calculated cumulative percentage rather than raw percentage to signal the composition of the geosocial field at each time point, but the patterns I identified hold for raw percentages. Figure 6.1 - the first line graph presented in Chapter Six - depicts the yearly cumulative percentage of geosocial papers with respect to all papers published in the same journals in the same period, for each Broad Disciplinary Category between 2008 - 2019. In other words, it depicts the extent to which journals in each Broad Disciplinary Category 'specialise in' publishing geosocial research. Figure 6.2 - the second line graph presented in Chapter Six - shows the percentage of geosocial papers with respect to the sum of papers published in all journals listed in WoS whose Subject Categories match the string search outlined in Table 3.3 for each Broad Disciplinary Category between 2008 - 2019, regardless of whether they publish geosocial papers or not. For each Broad Disciplinary Category identified by Discipline Categorisation Method 4 depicted in Table 3.4, I searched for all journals in WoS using the search strings depicted by Table 3.3. I summed the number of papers they published 2008 - 2019, and used these values to normalise the count of geosocial papers for each Broad Disciplinary Category. In other words, it depicts changes in the popularity of geosocial research in each Broad Disciplinary Category over time. Finally, figure 6.3 - the third line graph presented in Chapter Six - depicts the yearly cumulative percentage of geosocial papers in each Broad Disciplinary Category between 2008 and 2019, with respect to all geosocial papers in the same period. In other words, it depicts changes in the proportion of geosocial papers Broad Disciplinary categories contribute over time.

3.4.3 Tracing Heterogeneous Associations

This section discusses two heterogeneous network analyses I used to trace connections among research methods and spaces, and in one of them, also geosocial scholars. As Chapter Seven will explain, approaches to geosocial research differ with respect to the methods they use and types of spaces they study. However, as section 5.1.1 will argue, interviewees from diverse disciplines are interested in using machine learning for geosocial research. Thus, in section 8.2 I will visually analyse heterogeneous networks, where nodes are methods, spaces and in one version, scholars, to explore how machine learning (ML) affords perspectives on spaces. To help interpret the network maps, and provide a comparative basis, I created comparable heterogeneous networks about a second method: social network analysis (SNA). Next, I discuss how I compare and create the 'machine learning' and 'social network analysis' heterogeneous networks.

To scientometrically explore the relationship between methods and the types of spaces geosocial researchers study, I compared heterogeneous maps using two sets of geosocial papers: those whose abstracts, titles or author keywords include the term 'machine learning', and those whose abstracts, titles or author keywords include the term 'social network analysis'. As section 8.2 will discuss in more detail, SNA provides a suitable comparative case because the two methods appeared in geosocial literature at similar dates (2014 and 2015) but originate in different scholarly traditions (computer science and social science), thus illustrating the methodological diversity of geosocial research.

I created two types of heterogeneous networks for both ML and SNA related geosocial papers. Firstly, I studied how ML and SNA enable new scholars to conduct geosocial research and how scholars position their research with respect to existing geosocial research. To this end, I created heterogeneous networks to study the association among author keywords - some of which refer to methods, spaces - and geosocial scholars using author keywords and author information. The shape of nodes of figures 8.12 and 8.13 presented in Chapter Eight depict whether they are authors or author keywords. These networks also depict temporal information: the color of nodes indicate whether a keyword or author appeared in my geosocial dataset before its association with 'machine learning' or 'social network analysis' respectively. I chose author keywords for this analysis for two main reasons. Authors choose author-keywords to describe and position their paper (cf. Whittaker, 1989, cited by Wen et al., 2017, p. 725). In addition, author-keywords allowed me to create networks where the proportion of authors and terms related to methods and spaces is of similar order of magnitude: geosocial papers in my sample approximately have 0-5 author keywords and 1-4 authors. This helps explore the relationship between human and non-human actors in one network.

I created the networks as follows. For both the ML and SNA paper sets, I selected the papers that have author keywords. I created heterogeneous networks whose nodes are authors or author keywords, and weighted edges indicate cosine similarity relations calculated as specified by Equation 3.1. In addition, I calculated the earliest occurrence of each author and author keyword. I created a binary attribute for each, which signals whether they were present in my dataset - the geosocial citing papers - before their appearance associated with 'machine learning' or 'social network analysis', or they first appeared in the data associated with one of these methods. The color codes on figures 8.12 and 8.13 indicate the value of this binary time variable.

Secondly, I created heterogeneous networks using noun phrases VOSviewer identified in the abstracts and titles of geosocial papers related to ML and SNA respectively, to study the association among spaces and methods in more detail. This extends the above analysis and places more emphasis on the relationship between methods and spaces, whilst omitting authors. As noted earlier, not all papers have author keywords, but the majority have titles and abstracts. Thus, this analysis includes all geosocial papers whose abstracts, titles or author keywords contain the terms ML or SNA. In addition, whilst author keywords signal how scholars position their research, noun phrases in the abstracts and titles of geosocial papers contain more information about the methods and spaces scholars study because abstracts and titles provide more details about papers. This analysis required me to categorise noun phrases as methods or space relevant. As section 2.6.3.2 explained, heterogeneous network analysis has, to date, been used to study biomedical research. Next, I note a few methodological differences between these uses.

In biomedicine, concepts in published scientific literature, or scientometric data themselves include categories which signal non-human actors, such as molecules and methods (e.g. Shi, Foster, and Evans, 2015). However, there is a less clear mapping between spaces and methods in geosocial research and noun phrases in published literature. Although method related terms are easier to identify, as section 3.4.6 will discuss, it can be difficult to ascertain if certain noun phrases refer to social media platforms as methods or technology mediated practices. Spaces, on the other hand, are difficult to map onto single noun phrases. Although some terms clearly relate to spaces, such as 'China', 'New York', 'space', 'city', 'neighbourhood', other concepts are difficult to categorise as space related or not, such as 'landmark', 'air pollution', 'disaster', 'protest'. As section 3.4.8 will argue, often, it is the collection of noun phrases that describe situated practices that constitute the spaces geosocial scholars study. To create heterogeneous networks, I categorised author keywords and noun phrases as methods and space related similar to the thematic analysis method section 3.4.6 will discuss. To identify space related noun phrases, I sought those that refer to 'types of environments', 'specific locations' and 'spatial scales'. In addition, I identified noun phrases that refer to situated events, such as 'disaster', 'protest' or 'bombing'. I categorised noun phrases using the .txt file VOSviewer outputs in Excel. Then, I manually created a variable called 'Score⟨category⟩' that allows controlling the color of nodes using the VOSviewer user interface. Thus, the colors in figures 8.14 and 8.15 in Chapter Eight depict whether a noun phrase is categorised as method relevant, space relevant or other.

3.4.4 Identifying 'Social' and 'Technical' Approaches to Geosocial Research using Interviews and Temporal Citation Network Analysis

Chapters Six and Seven study the relationship between 'social' and 'technical' approaches to geosocial research with mixed methods. As Chapter Six will explain, interviewees' narratives suggest that 'social' and 'technical' approaches to geosocial research increasingly differ. This section discusses a scientometric network analysis that helped me trace the separation between these approaches, and thus identify them.

I traced changes in the citation relationships among papers associated with the 'only social' and 'only computational' Broad Disciplinary categories. In this analysis I only used these two Broad Disciplinary Categories to help scientometrically trace approaches identified via thematic analysis. My interviewees have backgrounds in social sciences, human geography and computational sciences: the two Broad Disciplinary Categories include papers from these, and only these disciplines. In addition, as section 3.3.2.2 explained, these Broad Disciplinary Categories are mutually exclusive: a paper can be either in the 'only social' or 'only technical' Broad Disciplinary Category, but not in both. Thus, using these Broad Disciplinary Categories allowed me to identify two groups of papers whose connections I could study. As described below I conducted both static and temporal citation network analyses to study the relationship between 'only social' and 'only technical' papers.

I first created an author-bibliographic coupling network using citing papers in the 'only social' and 'only computational' Broad Disciplinary categories. The nodes of the author-bibliographic coupling network in this analysis are the journals that publish citing papers. I chose journals as the nodes of this network because as Wen et al. (2017) argues, journals are longer lived scholarly media, and studying their relationships may offer a perspective about how "disciplines and specializations" relate (p. 726). In addition, compared to networks whose nodes are papers, using journals as nodes reduced the network size, which reduced the computational power required for generating 1000 random counterfactual networks (described below).

I calculated weighted edges between pairs of journals as a function of the number of authors the geosocial papers they publish jointly cite. I calculated author-bibliographic coupling as opposed to bibliographic coupling - where edge weights would be a function of the number of jointly cited papers - because author-bibliographic relations provide a less conservative metric for this analysis. Given the novelty of geosocial research, I assume that two papers citing a paper by the same author can signal a shared knowledge base between the two papers, even if they do not cite the same paper.

I calculated author-bibliographic coupling in two ways. The two methods differ in how they normalise the raw author-bibliographic coupling measure (the raw number of jointly cited authors among pairs of journals). Firstly, I created graph *G1* using the cosine similarity method Equation (3.1) described. Secondly, I created graph *G2* using the normalisation method based on Waltman, Boyack, et al. (2020) (cf. Waltman and van Eck, 2012). To create a non-directed author bibliographic network, I calculated the normalised author-bibliographic coupling relationship $r_{ij} = r_{ji}$ between journals i and j by calculating the mean of the raw author-bibliographic coupling measure $b_{ij} = b_{ji}$ described by Equation 3.2

$$b_{ij} = b_{ji} = \sum_{k=1}^{N} a_{ki} a_{kj} (1 - \delta_{ij}) \tag{3.2}$$

where a_{ki} is an edge that indicates journal i citing author k; a_{kj} is an edge that indicates journal j citing author k - the product yielding a value bigger than 0 when journals i and j both cite the same author, or in other words, are author-bibliographically coupled; and δ_{ij} is the Kronecker delta function

$$\delta_{ij} = \begin{cases} 0 & \text{if } i \neq j \quad (3.3) \\ 1 & \text{if } i = j \quad (3.4) \end{cases}$$

The *(1-δ_{ij})* factor in Equation 3.2 ensures the exclusion of edges calculating author-bibliographic coupling between identical journals. Next, I normalised the raw author-bibliographic coupling values by dividing them by the sum relatedness of pairs of journals i and j with all journals, given by Equation (3.5)

$$r_{ij} = r_{ji} = (\frac{b_{ij}}{\sum_k b_{ik}} + \frac{b_{ji}}{\sum_k b_{jk}}) * \frac{1}{2} \quad (3.5)$$

where k stands for all journals (all nodes in the network). With this method, the relatedness of journals is calculated with respect to the sum of their co-citedness. Thus, tow journals can be strongly connected even if they share relatively little connections with other journals. Thus, the relatedness values of journals in different disciplines are normalised compared to their own total relatedness, ensuring that they are "of the same order of magnitude." (Waltman and van Eck, 2012, p. 699)

Inspired by Navon and Shwed (2012), to study whether citing papers published in the 'only social' and 'only technical' Broad Disciplinary Categories can be conceptualised as separate approaches, for both of the author-bibliographic coupling networks *G1* (number of edges = m) and *G2* (number of edges = m), I studied the modularity of their sub-graphs - *SG1* and *SG2* respectively - which omit edges among journals assigned to the two different Broad Disciplinary Categories (*inter-edges* for short; number of *inter-edges* = k). In other words, I studied the modularity of sub-graphs *SG1* and *SG2* with identical nodes to *G1* and *G2* respectively, omitting all edges among pairs of journals where one journal is categorised 'only social' and the other journal is categorised 'only technical'. The number of edges of *SG1* and *SG2* equal $n = m - k$.

As explained in section 3.4.1, modularity scores are contingent on network size. Thus, I compared the modularity of *SG1* and *SG2* with the modularities of 1000 randomly generated sub-graphs of *G1* and *G2* with edge count = n (equal the edge count of *SG1* and *SG2*). In other words, I generated 1000 random sub-graphs for both *G1* and *G2*, that omit the same number of edges as the number of *inter-edges*, but instead of omitting the *inter-edges* themselves, they omit n random edges. If the modularities of *SG1* and *SG2* are similar to the modularities of the 1000 randomly generated sub-graphs, the *inter-edges* do not play a special role in shaping network structure - they do not render *G1* and *G2* more or less connected. If the modularities of *SG1* and *SG2* are smaller (i.e. the networks without *inter-edges* are more homogeneously connected) than the modularities of the 1000 randomly generated sub-graphs, the *inter-edges* render *G1* and *G2* less connected. Finally, if the modularities of *SG1* and *SG2* are bigger than the modularities of the 1000 randomly generated sub-graphs, the *inter-edges* make *G1* and *G2* more homogeneously connected. In this analysis I calculated modularity scores using Python igraph package, using the Leiden algorithm (Traag, Waltman, and van Eck, 2019), suited to calculate modularity for weighted networks.

As table 3.5 depicts, for both *G1* and *G2* I generated the 1000 random sub-graphs in two main ways. On the one hand, I created 1000 random networks each of which omit k number of random edges from *G1* and *G2*. Secondly, I created 1000 random networks each of which omit k number of random edges from *G1* and *G2* whose edge weight distribution is similar to the edge weight distribution of the *G1* and *G2* respectively - by deleting random edges with weights similar to the *inter-edges*. I calculated both of these, for both *G1* and *G2* using all four Discipline Categorisation Methods. In addition, I completed the analyses which control for edge weights in two ways: with edge weights rounded to one decimal, and using edge weights rounded to two decimals. As the observed patterns similar for the different versions, Chapter Six presents the results of the versions highlighted in blue and red in Table 3.5: the networks generated with Discipline Categorisation Method 1, using the narrowest definition of the 'only social' Broad Disciplinary Category; and edge weights rounded to one decimal point.

Finally, I completed both static and temporal versions of the above analyses. The histograms in section 6.3 depict the outcomes of the static analyses. The line graphs in section 6.3 depict the the temporal analyses: they depict, for each year between 2010 and 2019, the modularities of *SG1* and *SG2* and the range of modularities of the randomly simulated subgraphs, using the 2.5^{th} percentile value as lowest bound, and the 97.5^{th} percentile value as the upper bound, yielding a form of 95% 'confidence interval'.

author-bibliographic coupling network	normalisation method	sub-graph without 'inter-edges'	simulated random sub-graphs for comparison
G1 number of edges = m number of inter-edges = k	cosine similarity	*SG1* number of edges = $m - k = n$	**1000 random sub-graphs of *G1***, number of edges = n (***version 1:* Discipline Categorisation Method 1** & *other versions:* Discipline Categorisation Methods 2 - 4)
G1 number of edges = m number of inter-edges = k	cosine similarity	*SG1* number of edges = $m - k = n$	**1000 random sub-graphs of *G1***, number of edges = n; edge weight distribution comparable to that of SG1 (***version 1*: Discipline Categorisation Method 1** & *other versions:* Discipline Categorisation Methods 2 - 4) **AND** edge weights controlled for (***version A:* edge weights rounded to one decimal point**; *version B*: edge weights rounded to two decimal points`)
G2 number of edges = m number of inter-edges = k	based on Waltmann et. al. (2020)	*SG2* number of edges = $m - k = n$	**1000 random sub-graphs of *G2***, number of edges = n (***version 1:* Discipline Categorisation Method 1** & *other versions:* Discipline Categorisation Methods 2 - 4)
G2 number of edges = m number of inter-edges = k	based on Waltmann et. al. (2020)	*SG2* number of edges = $m - k = n$	**1000 random sub-graphs of *G2***, number of edges = n; edge weight distribution comparable to that of SG2 (***version 1:* Discipline Categorisation Method 1** & *other versions:* Discipline Categorisation Methods 2 - 4) **AND** edge weights controlled for (***version A:* edge weights rounded to one decimal point** OR *version B*: edge weights rounded to two decimal points`)

Table 3.5: Simulation Analysis Method. The thesis includes versions describe in in blue and red fonts.

3.4.5 Identifying Approaches to Geosocial Research using Citation Network Clustering

This section presents the network clustering method I use in Chapters Seven and Eight to trace approaches to geosocial research. I created weighted author-bibliographic coupling networks. In contrast to networks *G1* and *G2* outlined above whose nodes were journals, the nodes of the networks for this analysis are citing papers. The weights of edges between citing paper i and citing paper j are calculated using the method outlined by Equations 3.2 - 3.5. I created two weighted author-bibliographic coupling networks. Firstly, I used all papers that form part of my scientometric field (network *G3*). Secondly, I used papers in the Broad Disciplinary Categories 'only social' and 'only technical' (network *G4*). I clustered both networks using the Leiden community detection algorithm. It is important to note that the outcome of community detection is non-deterministic: each run might yield slightly different results even using the same parameters. The analyses included in the thesis yielded five clusters with more than two papers for *G3*, and three clusters with more than two papers for *G4*. In addition, the Leiden algorithm identified two clusters in each network with only two papers that were not connected to the rest of the papers.

Using papers as the nodes allows me to perform paper-level analysis to study the clusters of the network. I study clusters' disciplinary composition (the distribution of papers pertaining to different Broad Disciplinary Categories). I characterised each cluster by the distribution of papers across 5 out of 13 Broad Disciplinary Categories: 'only social' (only social sciences), 'only computational' (only technical sciences), 'phys geo' (physical geography) and 'non phys geo' (human geography) and 'multi-inter' (multidisciplinary, interdisciplinary). For each cluster, I counted the distinct papers pertaining to each of these five Broad Disciplinary Categories: the analyses include the raw paper counts. In addition, as section 3.4.6 and 3.4.8 will explain, I explore the topics papers in each cluster study through comparing terms in their abstracts and titles.

As section 7.2 will explain, I interpret the meaning of network clusters in light of their disciplinary composition, as well as interview analysis and findings presented in Chapters Five and Six. Based on these, I will argue that network clusters depict the scientometric footprint of approaches to geosocial research. In both networks, I interpret the two largest clusters as the 'social' and 'technical' clusters, corresponding to 'social' and 'technical' approaches to geosocial research respectively.

3.4.6 Characterising Approaches to Geosocial Research using TF-IDF values

I explored differences among clusters of the networks *G3* and *G4* with more than two papers - to compare approaches to geosocial research - using the term frequency-inverse document frequency (TF-IDF) statistic of noun phrases in the abstracts and titles of papers in these clusters. I chose TF-IDF (Salton and Buckley, 1988) because it is comparative metric that helps study noun phrases' relative frequency in clusters. TF-IDF scores are calculated for each noun phrase within each cluster - the TF-IDF of the same noun phrase in different clusters can differ due to their different frequency in each cluster. Noun phrase n's TF-IDF in cluster c is high if n is frequent in c but not frequent in the other clusters. In other words, TF-IDF indicates the specificity of noun phrases with respect to clusters. This method helps me highlight differences among network clusters, and thus study the difference among approaches to geosocial research. However, TF-IDF also over-emphasises the difference among clusters: it identifies noun phrases frequent in each cluster that are infrequent in other clusters. Next, I discuss how I calculated TF-IDF values.

I extracted noun phrases from sets of papers pertaining to network clusters using VOSviewer 1.6.15. As section 3.3.2.1 discussed, VOSviewer affords filtering noun phrases based on their occurrence and relevance. I calculated TF-IDF for each noun phrase in each cluster which occurred at least 10 times in the abstracts and titles of papers. I chose noun phrases that occur at least 10 times to ensure that the terms I identify as characteristic of clusters occur in them with relatively high frequency. I did not filter the network based on relevance score to ensure that my analysis captures all terms that differentiates the clusters. Firstly, I calculated the term frequency of each noun phrase in each cluster as defined by Equation (3.6)

$$tf(n, c) = \frac{count\ of\ n\ in\ c}{total\ number\ of\ noun\ phrases\ in\ c} \tag{3.6}$$

where n stands for noun phrase, c stands for cluster, and the count of n equals to the number of times n occurs in the abstracts and titles of all papers in cluster c. Secondly, I calculated the document frequency of each noun phrase, on in other words, the number of times they occur across all clusters, as defined by Equation (3.7)

$$df(n) = total\ occurrence\ of\ noun\ phrase\ n\ in\ all\ clusters \tag{3.7}$$

Thirdly, I calculated the inverse-document frequency (IDF) of each noun phrase as defined by Equation (3.8)

$$idf(n) = \log \frac{N}{df(n) + 1} \tag{3.8}$$

where N is the total number of clusters the noun phrase n occurs in (in my case, this can be 1, 2 or 3 given that I studied three clusters of *G3*). IDF measures the 'informativeness' of each noun phrase. The IDF value of noun phrases that frequently occur in all clusters, such as stop words, is low, because the higher the document frequency of the noun phrase, the bigger the denominator is, and for frequent noun phrases across clusters, *df(n)* increases more steeply than *N*. The logarithm value is calculated because otherwise IDF values can be very large. Finally, I calculated the TF-IDF score as specified by Equation (3.9)

$$TF - IDF(n, c) = tf(n, c) * idf(n) \tag{3.9}$$

Next, I ordered noun phrases in each cluster according to their TF-IDF (high to low) and thematically analysed the top 30% of noun phrases in each cluster, to study differences in the substantive foci of papers across clusters. Through studying the top 30% based on TF-IDF values, I thematically compare noun phrases which differentiate clusters, de-emphasising the similarities among them. I categorised noun phrases into six thematic categories described below - collective practice or topic, method, media, space, actor and time -, informed by the conceptual framework outlined in Chapter Two. Next I provide working definitions for each of these categories.

I define collective practice as forms of participation or engagement. I aim to describe these in a way that avoids false binaries between nature and culture. Geosocial research explores diverse forms of participation, including commercial practices, collective opinion formation, change in power relations, leisure, health, emergency responses and ecological practices.

Under 'method' I list noun phrases related to research methods and data types. Under 'media' I list noun phrases related to media technologies. Although there is an overlap between 'method' and 'media' (e.g. 'twitter' refers to both the social media platform and tweet posts that scholars use as data), the 'media' category aims to capture the Janus faced nature of social media (cf. Marres and Moats, 2015): they can concurrently be studied in methodological terms and from the perspective of the collective practices they enable. I decided to include these as separate categories to explore differences in how approaches to geosocial research engage with these aspects of media technologies.

Using the category 'space' I list noun phrases that explicitly refer to types of spaces. As sections 3.4.7 and 3.4.8 will explain in more detail, this category is not sufficient to study how geosocial research approaches study spaces. As section 2.5 discussed, I conceptualise spaces are relational to practices. Noun phrases which express spatial units are only a fraction on those that describe the types of spaces approaches study. However, studying 'space' related noun phrases helps study space related concepts each approach favours. I differentiate four main types of spaces. 'Specific locations' refer to existing, named spaces, such as cities (e.g. London, Amsterdam) or countries (e.g. USA, China). 'Spatial scales' refer to terms that express spaces in terms of scales, such as 'city', 'country', 'spatio-temporal pattern', 'homogeneous area' etc. 'Types of environments' refer to terms that express spaces which do not have immediate scalar connotation, but capture spaces' diverse relations, such as 'ecosystem', 'landscape', 'environment' etc. and 'spatial pattern' refers to terms that express data patterns expressed in space (e.g. 'spatial distribution').

Under 'actor' I list noun phrases which refer to human (e.g. 'adolescent', 'user') or non-human actors (e.g. 'agent', 'product'). Although from an ANT perspective, actors and collective practices can be difficult to distinguish - practices create actors - and thus the 'collective practice' and 'actor' categories overlap, I included 'actor' as a separate category to illustrate the different subjectivities approaches highlight (e.g. 'race', 'respondent' vs. 'human' and 'social media user').

Finally, the 'time' category lists noun phrases which either explicitly mention temporal units (e.g. 'January') or highlight the temporality of practices (e.g. 'shift', 'history', 'urban dynamic').

3.4.7 Exploring Similarities Between Social and Technical Approaches using Term Occurrence

I explored similarities between 'social' and 'technical' approaches through thematically analysing noun phrases which occur in both the social and technical clusters of the network *G4* at least 10 times (for short, *intersection noun phrases*). I did not filter the network based on relevance score to ensure that my analysis captures all relatively frequent terms shared by the two clusters. As mentioned above, TF-IDF highlights differences among clusters - it characterises each in terms of how they differ from the other. However, as Chapter Six will explain, 'social' and 'technical' approaches differentiate and concurrently develop thanks to connections and exchange between them. Thus, I explore similarities between the approaches. As section 7.2 will explain, I chose the network *G4* for this analysis, comprising geosocial papers associated with the 'only social' and 'only technical' Broad Disciplinary Categories to align this analysis with the interviews. I focused on terms that occur at least 10 times to keep the list of all noun phrases constant between this analysis and the the TF-IDF analysis outlined in section 3.4.6. I found 278 intersection noun phrases. I thematically analysed them using the thematic categories outlined in section 3.4.6. However, I found that many of the intersection noun phrases were generic words that were difficult to interpret in themselves, such as 'interest', 'topics', 'challenge', 'content'. I could not categorise these and thus left them out of further analysis. Figure 7.8 in Chapter Seven depicts the 112 noun phrases (40% of all intersection noun phrases) that I could categorise, which form part of this analysis.

3.4.8 Modified Ego Networks - Further Comparing Social and Technical Approaches

I further compared social and technical geosocial research by exploring how they frame shared concepts, by studying the 'modified ego-networks' of a few, selected intersection noun phrases. By ordering noun phrases based on their (relative) frequencies, both TF-IDF presented in section 3.4.6 and term occurrence analysis presented in section 3.4.7 disregard their relations to one another, and their meanings in the context of social and technical geosocial research. Comparing how social and technical approaches frame intersection noun phrases helps compare them without contrasting them, in a way that accounts for both their similarities and differences.

I studied the 'modified ego-networks' of three intersection noun phrases - 'citizen', 'city' and 'network' - in the social and technical clusters of network *G4*. I refer to these as ego terms. I chose these three ego terms from the 112 intersection noun phrases that I could thematically categorise for two main reasons. Firstly, as section 7.2 will explain, they are relatively frequent in both clusters, and thus, their ego networks are relatively big (contain a number of nodes) and thus semantically rich. Secondly, each refers to a different thematic category - 'citizen' is an actor, 'city' is a space and 'network' refers to a method. I assumed that each of these would provide opportunities to explore how these approaches study and frame spaces.

I mapped the modified ego-network of the three ego terms using VOSviewer. I firstly created term co-occurrence networks which depict all noun phrases in the abstracts and titles of papers in each cluster that occur at least 10 times. I did not filter the network based on relevance score to ensure that my analysis captures all noun phrases the ego terms are connected to. The nodes of the term co-occurrence networks are noun phrases identified by VOSviewer as discussed in section 3.4.6. The weighted edges of the network indicate the number of times two noun phrases co-occur in abstracts and titles of papers. Next, I sub-set these term co-occurrence networks to obtain the 'modified ego-networks' of the ego terms.

Ego-networks consist of a 'focal node' - the ego -, all the nodes that the ego is connected to - the alters - the edges among the ego and the alters, and those among the alters. My 'modified ego networks' omit edges among alters for readability. However, I depicted the relationships among alters - without showing their edges - by keeping the x and y coordinates of each noun phrase fixed, as defined by VOSviewer in the first step of the analysis, when creating the term co-occurrence maps which depict all noun phrases. Thus, the position and color of alters indicate their connections. Thus, resulting 'modified ego networks' depict the ego terms, the alters, links between the ego terms and the alters, and the color and position of the alters indicating their relationship to each other. In addition, to further explore how approaches frame the ego terms, as section 7.2 will explain, I interactively explored the 'modified ego networks' by only visualising a subset of the ego terms' edges as a function of their weight, using the edge weight filter function through VOSviewer's user interface.

Finally, as section 7.2 will explain, I include screenshots of the data visualisations in the context of the VOSviewer user interface to illustrate the constructedness of my own data analysis. Each software I use - Python, R, VOSviewer and Gephi - afford options to export data visualisations in various file formats, including .jpeg, .png and .pdf. Using these methods, the analyst obtains 'visually optimised' data visualisations which omit most traces of the infrastructures that helped create them, and the analytical decisions that they embody. To highlight the constructedness of my data analyses, I create screenshots that illustrate the data analysis process. It is my hope that these screenshots visually remind the reader of how my data infrastructure's key role in shaping data analyses throughout the thesis.

3.4.9 Exploring Geosocial Research with Noun Phrase Co-occurrence Network

Finally, Chapter Eight presents the noun phrase co-occurrence map of all geosocial papers created with VOSviewer. Similar to Callon et al. (1983) I assume that the network of terms in scientific publications provides information about the way scientists position and create 'problems' or research questions. As section 8.1 will explain, this helps me further explore differences in geosocial research agendas and the diversity of approaches to geosocial research. I created VOSviewer 'term maps' similar to the method described in section 3.4.3. I included noun phrases that occur at least 10 times in the dataset - VOSviewer's default setting. I also used VOSviewer's default setting for relevance scores, to excludes 40% of the terms based on their relevance score. I chose the default settings for this analysis because based on my experience with VOSviewer, these provide easily interpretable maps when used with a few thousand papers. Experimenting with different settings for this analysis is beyond the scope of this thesis.

3.5 Ethics

This project was approved by the University of Nottingham's Ethics Committee. This section reflects on the main ethical considerations the project raises - balancing my accountabilites towards my interviewees and the STS community and pseudonymisng interview quotes - as well as the ethics of my data collection and scientometric analyses.

This project raises two main ethical considerations. Firstly, as discussed in section

3.1.2 negotiated my concurrent accountabilities towards my interviewees and the STS community by illustrating their research practices in as much detail as possible, showcasing both similarities among approaches and differences that matter to my participants. Secondly, I pseudonymised all interview quotes and removed all details that could signal participants' identities to ensure that my research participants cannot be identified. Anonymization is not a 'monolithic' concept or straightforward in practice. With regards to ethnographic studies, there does not seem to be full agreement about the need to anonymise findings. For example, Levin (2014) discusses the findings of her ethnography naming the specific field site (the Computational and Systems Medicine laboratory at Imperial College London) and Madsen (2015) names interviewees who are PIs and team leads. On the other hand, Nelson (2018) conducts an ethnography of behavioural genetics researchers' work who aim to model human behaviour such as addictions using animal models, and reports her findings "almost entirely anonymously" (p. 218). She argues that this approach is warranted, due to heightened public controversy about the research topic, the presence of animal rights activists near the research labs she visited. Offering anonymity also helped her diffuse researchers' angst about her gaining insight into controversial or 'problematic' aspects of the labs' work. Finally, research participants in her study were used to working with "de-identified" (p. 218) samples. Her offering anonymity made her "research techniques look more familiar and legitimate" (p. 218) to research participants. I felt it is important to pseudonymise my interview data and remove details that can help identify participants because the arguments I created reflect my analytical perspective as well as interviewees' narratives. Readers might interpret the quotes differently, and interview quotes can be taken out of context as the thesis manuscript is circulated.

Power difference between me and my participants and gatekeeping raised no ethical challenges. I studied the practices of researchers, at similar career stage or more senior than me. Thus, power differences do not negatively impact them. Participants volunteered their time and could freely decline participation in the study without formal or informal negative consequences.

Finally, I reflect on the ethics of the scientometric analyses. Scientometric analyses may need to be anonymised if they containe detailed author or institute level information (e.g. Glaser and Laudel, 2015) or could, relatively easily be used for research evaluation (for a discussion about the role of (sciento)metrics and indicators in research evaluation see

Wouters, 2014 and De Rijcke, Wouters, et al., 2016). Most of the scientometric analyses in Chapters 4-8 are aggregated on the level of network clusters or disciplines. They do not need to be anonymised because they do not contain personal information about research participants or research institutions. Exception are the heterogeneous networks described in section 3.4.3 which present author information. I did not anonymyse these network visualisations because they use public data - author keywords and the names of authors - and do not contain or use information that can easily be used to evaluate, assess or rank scholars. The maps enable visual, as opposed to quantitative or statistical analysis, thus, cannot easily serve as a basis for indicators or metrics or ranking. In addition, the information the maps contain do not directly relate to scholars' performance. They only depict a subset of scholars' papers and emphasise their thematic relations, providing no direct information about authors or their overall research output.

3.6 Limitations

This section concludes the methods chapter by reflecting on the limitations of the research design. I discuss four main limitations: limitations of the interviews, the field delineation, using co-authorship analysis to study collaboration and tracing relationship among approaches to geosocial research using citation analyses.

This project combines interviews, participant observation and scientometrics, in order to diversify the analytical perspectives on the geosocial research. This mixing of methods presupposes that neither method provides an accurate, or 'God's eye' view on geotagged social media data practices. Neither the interview, nor the scientometric fields capture all of geosocial research. Interviews are limiting because I could interview a small proportion of scientists who use geosocial data for academic research. In addition, interviews provide limited insight into participants' practices, which I partially mitigated by conducting informed interviews. Next, I discuss the limitations of my scientometric methods.

Academic publications and by extension, data associated with them can only provide a limited perspective on academic research. As Latour and Woolgar (1986) argue, scientific publications are artefacts that form part of research practice, but do not capture the diverse practices that constitute research, such as relational practices with actors outside scientific institutions. Hence, analyses that use bibliometric data can only take limited aspects of research practice into account.

Secondly, scientometric databases are selective. As section 3.2.2 discussed, WoS over represents papers from the natural sciences and humanities and under-represents the social sciences, arts and humanities (Mongeon and Paul-Hus, 2016). Thus, social scientific geosocial papers might be under represented in my scientometric field. I noticed that my scientometric field does not contain all of the papers of my interviewees. In addition, my choice of social media platforms and search terms limited the scope of the scientometric field. For example, my field delineation excluded research with the popular Chinese microblogging website Weibo. I conducted interviews with 19 geosocial scholars. I selected interviewees with diverse seniority and disciplinary background. However, as section 6.2.3 will explain, there is a main discrepancy between my interview and scientometric fields: even though my scientometric field contained health and ecology related geosocial research, none of my interviewees focused on these topics.

As section 3.4.1 explained, I studied collaboration among geosocial scholars with interviews and scientometric co-authorship analysis. Although, as I argued, scientometrics helps me study collaboration among more scholars over longer timeframes, co-authorship does not equal collaboration. Scholars may collaborate without co-authoring papers.

Finally, as sections 3.4.5 and 3.4.4 outlined, I differentiate between geosocial research approaches by combining citation network analysis and interviews. Inferring relationship between research areas using citation-based measures has significant limitations. The limitations of citation metrics and the ambiguity of citation behaviour have been discussed since citation analysis became widespread in the 1960s after the development of the Institute of Scientific Information (ISI, now owned by Clarivate Analytics) (cf. Bornmann and Hans-Dieter, 2008). Although numerous criticisms have been made against using citation-based measures to study and evaluate research, the limitations of citation analysis methods depend on the analytical purpose and the way one conceptualises citing behaviour (MacRoberts and MacRoberts, 1989). For example, for those who assume that citations signify informational influence, the variety of motives scholars have named that can motivate citations - such as citing research to confirms one's results; to dispute aspects of cited work; citing persuasive, well known texts; providing only perfunctory reference to cited work (e.g. Garfield, 1962; Cronin, 1982;

Bornmann and Hans-Dieter, 2008) - may pose challenges. The diversity of motives behind citing behaviour does not impact negatively the outcome of this study. Broadly speaking, this project uses citation analysis to explore thematic relationships between papers. I only assume that a citation signifies some level of familiarity with or influence of the cited work. This influence does not need to be 'informational influence', strictly defined. Nevertheless, the limitations of citations discussed below impact my study.

Firstly, the reference list might be incomplete and there might be systematic trends in it. This means, that the thematic clusters established based on citation-based measures may be a by-product of the data (citations), instead of reflecting real similarities and differences between them. According to MacRoberts and MacRoberts (1986) who compared the list of references with influences identified based on papers' text, most scholars do not cite the majority of their influences. In addition, there might be more or less systematic omissions in references. For example, the origin of earlier results may or may not be accurately cited. Secondly, references only indicate 'formal influence', and leave tacit skills - developed over time, and in a range of situations, often collaborating with others - as well as the day to day discussions, negotiations with fellow researchers, laboratory technicians ('shop talk') out of the picture (MacRoberts and MacRoberts, 1989). Science studies have repeatedly demonstrated that the narratives presented in published literature bear little resemblance to scientists' daily practices (e.g. Latour and Woolgar, 1986).

Interpreting scientometric findings in light of the interviews - which help study interviewees' motives and reflections - and using multiple types of scientometric analyses, including citation analyses, co-authorship and term co-occurrence maps helps offset the above limitations. I create the thesis' arguments in light of diverse data types. In some cases, I highlight findings which are constant across the methods, whereas in other cases I contrast findings from diverse methods to provide more nuanced arguments.

Altogether, this chapter outlined the thesis' methodology. It outlined my mixed-methods case study approach, scientometric and interview field delineation, approach to thematic analysis and discussed each computational method in detail. Finally, it discussed ethical considerations and methodological limitations. Next, Chapters Four Through Eight will discuss the thesis' empirical findings.

Chapter 4

Geosocial Research Across Institutions

This chapter highlights three practices in response to the First Research Question which explores how approaches to geosocial research develop. Firstly, as section 4.1 discuses, interviewees from all disciplines stated that geosocial research requires them to mix two research traditions which are often challenging to combine: computational data analysis and socio-spatial interpretation. In this sense, they work 'across' research traditions. I find that overcoming the challenge of combining computational data analysis and socio-spatial interpretation is a core aspect of interviewees' geosocial research, and essential for them to develop their geosocial research approaches. The rest of the thesis further explores interviewees' practices that motivate or help them to combine these knowledge traditions. This chapter highlights two such practices: participants' concurrent aesthetic appreciation of combining computational data analysis and socio-spatial interpretation (discussed in section 4.2), and the need to conduct non-academic and academic research in parallel - mainly due to financial pressures - which exposes them to computational data analysis methods (discussed in section 4.3).

4.1 Computational Analyses and Interpretation Diverge

All interviewees stated that a main challenge of geosocial research is combining computational data analysis and socio-spatial interpretation. The challenge stems from four main differences between the two: different communities teach them; they comprise different approaches to validating knowledge claims; scholars need prior experience in computational research to recognise infrastructure development - which is essential and time consuming - as a standalone scientific contribution; and time limitations impede any one person to keep up with developments in both traditions. Participants' views highlight the effort required to create scholarly narratives with geosocial data by combining computation and socio-spatial interpretation.

All interviewees - regardless of their disciplinary backgrounds - agreed that distinct institutions teach computational data analysis and socio-spatial interpretation, as the seven quotes below illustrate. In line with the divide between critical human geography and GIScience section 2.1 discussed, colleagues Ben and Bruno, respectively, stated that the geosocial research their team does is challenging because it requires familiarity with these two distinct geographic traditions of thought, while most scholars are primarily trained in one.

> *"...[this] falls right in between GIScience and human geography. It's something in between, something different. And it's not necessarily a comfortable space to fill ... because you have to be sort of familiar with both [...] And that can be hard to do." – Ben*

> *"[Brian] is a very unique guy because he knows social theory [...] but he's also a computer programmer and can access the data and scrape data and do all that stuff which is obviously important [...] And he also has statistical and spatial analysis and [...] visualization skills." – Bruno*

Henry, Colin, Daniel, Jane and Isaac - all quoted below - noted the separation between computational data analysis and socio-spatial interpretation in disciplinary communities beyond geography. According to Henry, who supervises graduate students from diverse social and computational disciplines, learning geosocial research requires knowledge about computational data analysis and socio-spatial interpretation, which bring complementary challenges depending on students' disciplinary backgrounds.

> *"[Students in the program with backgrounds in social science and computer science][...] interact daily [...] [they] help each other [...] these trends are fairly complementary [...] [and] when you're supervising them individually the aspects you have to work on are different."*

Colin and Daniel highlighted the separation between sociology and computational data analysis. Colin used online platforms to learn computational skills he did not learn as a sociologist.

> *"I have been doing classical sociological research with focus groups and surveys for 15-20 years, and I started to realise that I won't be able to sell this expertise in 20-30 years, with this whole big data trend... [...] [later] I used [online platforms] to learn data analysis skills. [...]"*

Daniel stated that scholars rarely have trained in both computational data analysis and social science, even though this combination is essential for geosocial research.

> *"first when I thought of computational social science, I thought of [...] somebody who has a mathematics background, who is a computer scientist, and who then went on to study sociology. [...] then as it turned out [David] actually could do both programming and sociology which is a rare combination."*

Jane and Isaac discussed the challenges for computational scholars to learn essential socio-spatial interpretation skills. Jane explained that reviewers often request her team to further develop the interpretative aspects of their papers. Her and her colleagues' backgrounds are in physics. She explained that the interpretative skills required for geosocial research - for example, discussing findings at greater length - is outside the training they received.

> *"Reviewers always ask us to work more on interpretation. [...] In retrospect, I also think that the discussions of the first few papers I published were pretty brief. [...] In the social sciences, there is a very different publication culture, that we don't know. That's not what we learned."*

As part of his PhD studies, Isaac eventually visited an interdisciplinary, social sciences-oriented institution where he learned about social theories which he uses to frame his research and interpret his findings. Whilst in his (technical) discipline the relationship between variables is primarily interpreted in terms of the mechanisms models afford, the sociologists and geographers he met introduced a novel interpretative approach, studying the relationship between variables using theories that cannot be reduced to the mechanisms created with data models.

> *"...maybe on my side [technical discipline] there is a lot of discussion about what's the relationship between variables [...] [we tend to come up with] models of how [patterns] might emerge [...] [but] it took the sociologists and geographers saying okay but what's the theory, [why does the phenomenon happen]... [is it] an agent to agent interaction [...] or social sanctioning? [...] also this leads naturally to the conclusions we draw from finding such a relationship."*

Towards the end of his studies at a computer science-focused institution, Isaac gave a talk discussing his geosocial research which included details about the above theoretical framing that the audience found novel and unfamiliar. This further illustrates the separation between socio-spatial interpretation and computational data analysis education.

> *"I talked about my [paper related to social media network] and I spent a few slides on theory that I learnt from the [social scientists] [...] after the talk [...] one guy gave the comment, 'that's the most theory I have ever seen!' ... like it it was really different for them."*

In sum, interviewees with backgrounds in geography, sociology, physics and mathematics stated that distinct disciplines teach computation and socio-spatial interpretation, which makes it challenging for one scholar to learn both.

The quotes below by David, Ben, Jane and Colin illustrate the second factor which maintains the difference between computation and socio-spatial interpretation: prior computational experience is necessary for recognising essential and time consuming computational tasks a valuable part of geosocial research, such as computational infrastructure development or querying large datasets. David, Ben and Jane discussed the prior experience required for valuing infrastructure development. David stated that many of his sociologist colleagues without experience in computing do not "value" his work of maintaining servers and databases.

> *"...stuff that I've been putting a lot of hours into, even just like maintaining a server, making sure the database runs [...] They cost a lot of time! But they are really not valued by most sociologists that you talk to..."*

Ben stated that his team's previous experience with computing sensitised them to some aspects of data infrastructure development - namely, the importance of storage space -

whilst they learned about other challenges, such as creating query-able databases on the job. Brian developed group B's data infrastructure as part of his PhD research. Ben claimed that the flexibility of the PhD program allowed Brian and the team to explore unforeseen challenges associated with data infrastructure development and make it a valued part of the research. Although Brian's PhD initially focused on the empirical analysis of geosocial data, it also grew to encompass infrastructure development.

> *"...not coming from a computer science but a social science background [...] we were not prepared for some of the challenges... There's the basic one that we knew about [...] storage space. [...] [We also] had to look into [...] how to make [data] usable [...] to [be able to] search it in a very timely way [...] that sort of PhD student experience can be quite liberating [because you can say] 'ah well, let's give it a try'. [...] [creating the data infrastructure] [...] became a big part of [Brian's] PhD, but his PhD was actually on sort of questions based on this data."*

Jane also argued that prior experience is essential for understanding the tasks required to build and maintain data infrastructure. Her colleagues who developed group J's data infrastructure built on their experiences in working with astrophysics databases:

> *"My colleagues very carefully planned the database structure, the data collection procedure: which computer they use for it and how they load the data into the database, how it will be search-able etc. [...] They used their experience with astrophysics databases."*

Ben's, Colin's and Jane's views below suggest that prior computational experience is essential to recognise the value of computational skills, for example to analyse large datasets. Ben stated that prior experience helps appreciate complementary skills: for social scientists to appreciate computational tasks as valuable contributions and vice versa.

> *"...I think it's important to have sort of an understanding or at least sort of a working knowledge of one another, or an interest because otherwise it's very easy to say, if you're say a critical human geographer [that] data is just [...] too reduction[istic], 'I'm not going to pay attention to it' [...] And [...] essentially the same on the reverse."*

Colin discussed plans to hire a new team member who has experience in working with large and remotely stored data. He contrasted his experience working with csv files and data hosted locally with analysing big data hosted remotely. His computational knowledge allowed him to recognise the limits of his expertise and the challenges associated with accessing and manipulating remotely stored large datasets.

> *"But the [other, larger data], that's going to be very good, but we need to have the skill to work with it. That one we won't have on our servers... Accessing it will be a lot more strict, both legally and technically. Then it doesn't work anymore that the social scientist manipulates csv files and databases on their computer. We'll need an IT person or similar on our team."*

Jane also stated that working with big data efficiently requires prior experience. According to her, querying large data efficiently requires tacit knowledge which is difficult to learn alone by tutorials. Working with mentors with experience with large datasets was essential for her to learn big data techniques.

> *"...with these [large datasets] you have to learn to write very efficient SQL Query Plans. And I think you can only learn that by doing and having a more senior mentor. You can't really learn this from books."*

Altogether, David's, Ben's, Colin's and Jane's narratives highlight that prior computational experience is necessary for conducting, and recognising essential and time consuming computational tasks as valuable scientific contributions. Since, as outlined above, computation and socio-spatial interpretation is taught by different disciplinary institutions, scholars without computational training may lack the experience required for recognising essential tasks as valued part of geosocial research.

The quotes below by Jane, Kevin and Daniel explain a third way in which computation and socio-spatial interpretation differ: they take different approaches to validating knowledge claims, including the way they construct evidence and arguments. Kevin contrasted the mathematical skills he thinks computational arguments require with knowing previous literature for constructing social scientific arguments - and stresses the importance to bridge knowledge gaps for geosocial research.

> *"Tolerance also plays a big role. [...] natural scientists trained in mathematics have to accept that not everyone had to learn as much mathematics as we did [and] social scientists should not get immediately outraged if it turns out that there are big gaps in our knowledge,*

> *for example, if we don't know famous scholars they cite. Somehow this needs to come together."*

Like Kevin, Daniel stated that ideally, computational social research combines diverging approaches to establishing knowledge claims: mathematics or computation, and skills curated by the interpretive social sciences, including theorising and reflecting on how the researcher's assumptions shape their work.

> *"[But] I think that what characterises a lot of computational social science is a total ignorance of social science, even a disdain for it. Often it disavows any sort of theory, interpretation, reflexivity, [...] that sets social science apart from the natural sciences..."*

Like Daniel, Jane stated that geosocial research - in contrast to physics - requires interpreting findings by discussing them in longer narratives and contextualising them with theories.

> *"We find it difficult to write an essay-like discussion, or contextualise our findings with theories. We usually do the measurement, write up the results and discuss their contribution in a few sentences. In the social sciences, there is a very different publication culture that we don't know."*

Altogether, the above quoted interviewees contrasted computational reasoning - which uses mathematics to create evidence and presents findings briefly - with socio-spatial interpretation, which validates findings by relating them to established theories and longer narratives reflecting about the researcher's position.

Finally, David, Anne, Ben and Henry emphasized the challenge of finding time to keep up to date with advances in both computation and socio-spatial interpretation. David's earlier quote emphasized the time required for database maintenance which is not recognised by many of his sociologist colleagues. Anne, Ben and Henry stated that developing expertise in both computation and socio-spatial interpretation would require more time than available for any one scholar.

> *"It's really hard to stay on top of the software and the data ... if you're going to be focusing on data processing, computer languages, [...] that leaves less time for you to read critical theories [...] it's not impossible to have both, but you tend to specialize more in one versus another." – Ben*

> *"...the day has 24 hours so I can't keep up on all the latest theoretical developments on cities and all the latest literature debates around you know where cutting edge social context." – Henry*

> *"...it was clear to me very early on, that I would never be an expert coder. While I did write my own code in Python in order to filter data from the API, I realised that writing my own code in order to run queries on that data would take me longer than it would be worth it." – Anne*

In sum, David, Anne, Ben and Henry emphasize the time required to develop and maintain expertise in either computation or socio-spatial interpretation, which makes it challenging for scholars to develop learn both.

Altogether, this section illustrated a core challenge interviewees from all disciplines experience when conducting geosocial research: combining computational data analysis and socio-spatial interpretation. This chapter further explores practices which help interviewees succeed in conducting geosocial research at the intersection of these research traditions. Section 4.2 will argue that interviewees are motivated to combine them because find their combination aesthetically pleasing. Section 4.3 will argue that the industry-relevance of computational skills prompt or motivate scholars to conduct geosocial research.

4.2 The Aesthetics of Geosocial Research Across Computation and Social Science

Participants from diverse disciplines claimed they find using geosocial data for research, or combining computational data analysis and socio-spatial interpretation for geosocial research aesthetically pleasing. In particular, many interviewees highlighted aesthetic values associated with aspects of geosocial research they have less training in. Interviewees with technical backgrounds emphasized the beauty of conducting social research and social scientists emphasized the beauty of geosocial data or computational data analysis methods. I argue that this aesthetic appreciation underpins interviewees' efforts to combine computational data analysis and socio-spatial interpretation in the face of the above challenges, and thus develop their approach to geosocial research.

Firstly, two physicists, Josh and Jane stressed the beauty of creating scientific explanations using (models with) few assumptions and variables whilst studying 'real life' practices, such as cities. They noted that this joint interest motivates them to pursue geosocial research as opposed to other physics research topics.

> *"I'm really interested in what I can say about human behaviour. [...] not only on an abstract level. [...] At the same time, I studied physics because I find its extremely simplified way of seeing the world really beautiful. [...] [beautiful is, for example,] with a few equations [or] few variables, creat[ing] a universal model. [...] I want to continue this research with social media data to stay in between..." – Jane*

> *[one of the most interesting aspects of this research] is that I can study topics close to reality, or practical applications. [...] Physicists often do theoretical research [...] [research is mathematically interesting] when you can show a relationship using minimal amount of starting assumptions [...] what you can calculate on paper [...] I think lots of people with background in maths and physics would agree, but we have to admit that sometimes the most useful models are not the ones that are mathematically most interesting." – Josh*

Secondly, several social scientist participants noted the aesthetic value of computational data analysis methods. Like the physicists above, Brian, who has background in human geography, noted the beauty of the 'elegance' of simple models. At the same time, he noted he finds capturing nuances - afforded by ethnography - essential, and argued for team based geosocial research that combines both.

> *"...sometimes working with a physicist can be great because they can come up with a really efficient, elegant way to quantitatively make sense of a certain dataset but then also by its very nature that is so abstracting that sometimes we miss out on really important qualitative nuances. So that's where it's great to work with let's say an ethnographer [...] So you can have that kind of tension and contrast within the same team."*

David's quote suggests that he appreciates the aesthetics of computational research methods, even when he is not interested in the research questions these methods are used to explore. For example, his interests - shaped by his sociological training - oppose the research questions explored with experimental 'computational social science' studies, which he can aesthetically appreciate nevertheless.

> *"I guess I have a sensibility that's shaped by kind of classical sociology, like 'verstehen' rather than explanation and, yeah, I don't have a lot of patience for these long, rigid, experimental designs [...] They're very impressive, I mean I can definitely appreciate the labour [....] and [...] the whole aesthetic of it, but I don't get really excited about those kinds of findings."*

Chase, Colin and Bruno - interviewees with background in human geography and sociology - noted the aesthetic value of computational data visualisations. They find it attractive when data visualisations communicate complicated arguments. They believe that data visualisations help capture readers' attention and communicate findings.

> *"...being able to do interactive web based visualizations or whatever, that's what drives attention and interest in this stuff even when the questions aren't actually interesting. [...] Nobody is going to read the thousand words. They just want to look at the pretty interactive map or whatever." – Bruno*

> *"...data visualisation is really important. At the time scientists didn't really invest effort into it - I understood its importance through my market research experiences. Put it simply, making something that looks good is half way to success." – Colin*

> *"...it's better to do research more quickly, for a bigger audience, sexier things [...] more beautifully... [using] good data visualisations. [...] when they can communicate these ideas as part of a simpler story with the visuals, that really sells it. [...] It's important to have something visually pleasing when you look at it, and so you can discover what's interesting in what you can see." – Chase*

In addition, Brian finds attractive the puzzle posed by the very challenge of combining social theory with empirical research designs with social media data.

> *"I think social media data initially grabbed my attention because [...] connecting theory with an empirical research design with social media*

> *data that's an incredible puzzle. [...] [it also] allows us maybe to create insights into how social processes work that maybe are not as easy to gain from more conventional data sources."*

Finally, Anne, Ben, David and Frank - interviewees with background in anthropology, geography, sociology and urban studies respectively - noted the aesthetics of social media posts themselves. Anne likes that they express brief, everyday passing experiences - 'everyday chatter'.

> *"...probably you're not going to learn anything really profound about the history of a place, but I think the joy is in the un-profound and like the ridiculousness of what people [post] about, and how they kind of enact their identity day to day in relation to location. I love... [...] kind of the chatter element [...] When you know your feed is being updated second by second there's a sense that you just say whatever is on the top of your head..."*

Like Anne, Ben noted the conversational nature of geosocial data: he stated that geosocial research helped them create research which differs from the academic conversations they are used to - 'fun', shorter research pieces which engage the public.

> *"Another motivation was to share this research with non-academic audiences [...] questions that would be of interest to the general public. [...] [do research that is] fun and like less serious [...] sort of tongue in cheek. [...] This was before there was a lot of [...] 'click bait'..."*

David and Frank stated that the aesthetics of the pictures associated with the geosocial data they collect help capture the audience's attention.

> *"The pictures with the social media posts help us engage policy makers. It's beautiful, they catch people's eye." – Frank*

> *"...I think that we had something to show for that was also very important, even if the results were not conclusive [....] Maybe also the power of images [...] [it] really helped convince people that there might be something interesting." – David*

Altogether, interviewees' quotes suggest that aesthetic appreciation of both social research and computational data analysis or geosocial data underpins their motivation to conduct geosocial research across computational data analysis and socio-spatial interpretation. Next, I explore how non-academic work prompts or motivates scholars to conduct geosocial research.

4.3 Geosocial Research Across Academia and Industry

As section 4.1 discussed, interviewees feel that geosocial research requires them to learn skills distinct from their home disciplines' intellectual traditions. As the quotes below show, many of my participants learned computational skills they use for geosocial research through non-academic activities, which they often sought to cope with financial pressures or sometimes due to a longstanding interest in computing. Thus, I argue that interviewees' concurrent academic and non-academic work results in a more diverse the pool of geosocial researchers.

Firstly, Jane, Miles and David learned computational data analysis skills as part of their non-academic employment. Jane explained that she learned technical skills she uses for geosocial research through non-academic work, which she could not learn during her academic studies because most colleagues with relevant technical skills had left her department.

> *"During my PhD I worked part time at a [company that] [my department] has a really good relationship with. [...] [there I learned] to code better [...] I use a lot of these skills in my academic work. [...] At the university there weren't many people who did similar research to me, because most of the people who worked with this data had already finished their PhDs and left."*

Miles, who sought non-academic work for financial reasons learned social media data analysis as part of the commercial research he conducted with geotagged social media data. After receiving advice from senior researchers at the company, he learned data scraping and computational map-making largely on his own, using online material and asking friends for advice.

> *"I decided to do computational data analysis related research because that's where most of the job opportunities are these days. [...] At the beginning I received some advice from senior researchers at the*

> *company. [...] But I figured most of the things out on my own."*

David's geosocial research also benefits from computational skills he learned outside of academia through his longstanding interest in open source computing. He was interested in Daniel's geosocial data project because it allowed him to re-position himself as a social researcher who uses computational methods. After pursuing social scientific and computational research interests separately for years, geosocial research enabled him to combine and apply these interests in academic research.

> *"I have this relatively long history [...] of working with Python on and off [due to my interest in UNIX], kind of just building toy things [...] Part of what really appealed to me about [the geosocial research] position [with Daniel] was the ability to kind of reinvent myself [... as] somebody who does the computational stuff kind of as a more serious part of my research work."*

Secondly, in addition to developing computational skills, non-academic research prompted the interviewees quoted below to explore geosocial research. For Anne and Luke, such non-academic research was initiated through their departments' non-academic funding: they partnered with a non-academic institutions with an interest in geosocial research for their PhDs. A collaborative project with his partner organisation - which mapped geosocial data at the granularity of a publicly accessible building - prompted Luke to explore the accuracy of geosocial data in his (academic) geosocial research.

> *"[At that point my industry partner] got more involved as well [...] Part of that was analysing [a publicly accessible building] [...] so I dwelled really deeply into the location coordinates [and discovered that the social media platform] gets the resolution wrong. So a lot of my [geosocial research] methodology has been trying to fix that."*

Henry, who also works at a department which provides on PhD training in collaboration with non-academic institutions stated that the industry relevance of computational skills used for geosocial research are beneficial for knowledge diversity in urban research. He thinks the commercial relevance prompts scholars with diverse disciplinary backgrounds to study cities - including those who do not seek academic employment. He believes that this helps diversify research about cities, even considering research which 're-invents the wheel'.

> *"I think most of [our PhD students] will [...] go to industry, which I think in itself is a victory or a success of the program or [...] it says something about the relevance of these methods [...] 10-20 years ago, if you wanted to do a PhD in studying cities you probably wanted to end up as a university professor. [...] [the recent change] I think that's a good thing. it's going to add more variety [...] there's also a lot of wheel reinvention and [...] misunderstandings [...] but I think in the big picture it also adds more perspectives..."*

Colin, who has been working in the private sector besides his academic job since finishing his MA degree, learned computational data analysis and visualisation - that he used for geosocial research - as part of his commercial work. In addition, him and Chase started to work with geosocial data in the first instance, because their non-academic professional network allowed them to gain access to a geosocial dataset.

> *"Due to financial reasons [...] since I finished my MA I have always worked in the private sector too. [...] I got most of my practical skills like data visualisation from market and other social research projects. [...] [One of the companies] I did market research for is [the technology company that in part owned the social media platform we studied]. [...] this is how we started to consider getting access to the [social media data]."*

Like the above interviewees, non-academic research - which he pursued for financial reasons - prompted Elias to pursue geosocial research in the first place. He also learned relevant computational skills through this non-academic work. He adapted his academic research to include his non-academic research: although he first felt that computational research differed from his academic work and later came to see connections between the two, altering his academic research focus.

> *"[For financial reasons] throughout my entire PhD I was working for different NGOs... [And around that time] a lot of people were doing these kind of network analysis of big data from [social media]. So I gave it a shot and started to learn some of the techniques [...] For a long time [...] I've felt like I've got these like two sides of my research and now I'm realising that I'm only really doing one thing [...] I'm working on ways to try to integrate [them]."*

Henry's non-academic research also prompted him to pursue geosocial research. He participated in non-academic discussions about computational data analysis - due to his interest in open source computing - for several years before he realised the relevance of computing and new digital data for urban research. As with David, geosocial research allowed him to combine his interest in computational data analysis with urban research, his original research focus.

> *"I've been a interested in open source development, software and computational methods [...] somewhere in the early 2010s, the word Big Data started becoming a thing. Not in academia and definitely not in urban studies but more in the tech world. [...] And then at some point I realized that it was going to change how we look at cities."*

Finally, Bruno needed to take a part-time job, unrelated to his PhD research, that left him with less time for his academic research. Conducting geosocial research for his PhD enabled him to work on his research flexibly in terms of hours and location. He completed his doctoral research in collaboration with Ben and his team, who he knew from a previous project.

> *"...needing to take on another part time job to be able to pay my bills and stuff. I knew that I couldn't go do this in-depth ethnographic project in a place I didn't live without having some research funding. Whereas the social media project I could do from my couch, basically in my spare time. And that's how I finished my dissertation."*

Altogether, the interviewees quoted in this section use computational data analysis skills they learned through non-academic research - or due to their departments' non-academic funding - for geosocial research. I argue that conducting non-academic and academic research concurrently diversifies the pool of computational analysts because it prompts researchers with diverse backgrounds to learn computational data analysis skills - or may even expose them to research with geotagged social media.

4.4 Conclusion

This chapter discussed three practices through which interviewees develop their approaches to geosocial research, answering the First Research Question. Firstly, I argued that for my interviewees, combining computational data analysis and socio-spatial interpretation is a core aspect, and challenge associated with geosocial research. In contrast to Marres and Gerlitz (2016) who highlight the historical links between social and computational research methods that comprise social media platforms and Mayhew (2011) who highlights epistemic similarities and genealogical links between the 'quantitative' and 'qualitative' geographic research traditions, I found that my interviewees perceive computational data analysis and socio-spatial interpretation as complementary but distinct. In line with Dumit and Nafus (2018), they stated that distinct disciplinary institutions teach computation or socio-spatial interpretation, which makes it challenging for one scholar to learn both. My participants also stated that recognising computational tasks as a valued part of research requires prior experience, and that computational data analysis and socio-spatial interpretation comprise different approaches to validating knowledge claims, which can be difficult to master and combine. Computational data analysis favours mathematical reasoning, while socio-spatial interpretation requires longer narratives and familiarity with social theorists. Finally, they emphasized that both computational data analysis and socio-spatial interpretation take time to master. Given the growing rate of scientific publications that (geosocial) scholars need to keep up to date with (e.g. Bornmann and Mutz, 2015), such time constraints will likely keep affecting their work. Next, the chapter highlighted two practices that motivate or enable interviewees to combine these research practices: their aesthetic interests and concurrent academic and non-academic research.

I argued that interviewees' aesthetic appreciation of methods required for geosocial research which are complementary to their background motivate them to combine computational data analysis and socio-spatial interpretation. Physicists emphasized that they find both the simplicity of modeling and applying such models to studying societal phenomena beautiful. Social scientists noted their aesthetic appreciation of either computational data analysis methods - such as data modeling and data visualisation - or the aesthetic appeal of social media posts themselves.

Several interviewees' aesthetic interests combine elements of both "classical, formalist aesthetic" that values "unity, economy, symmetry, consistency" and order, and a valuing of "diversity, differentiation, complexity", contrasted by (McAllister, 2002, p. 9). Three interviewees - two physicists (cf. Hossenfelder, 2018) and a social scientist - noted the beauty of values similar to classical aesthetics. However, for the physicists Josh and Jane, this aesthetic appreciation is linked to their interest in empirically studying situated practices and human behaviour - such as cities. For Brian, a human geographer, it is linked to his aesthetic appreciation of detailed explanations and human geography theories which highlight historicity and situatedness. The latter are more similar to the second type of aesthetic noted by McAllister (2002) that values "diversity, differentiation, complexity", (p. 9), which he contrasts with the above classical, formalist aesthetic. This suggests that geosocial research helps interviewees enact diverse aesthetics as part of one research project, and that these sensibilities motivate them to overcome the challenge of combining computational data analysis and socio-spatial interpretation. In addition, several social scientist interviewees noted the beauty of data visualisations. They claimed that data visualisations are relatively rare in their home disciplines, and their interest in exploring data visualisation methods. Finally, several social scientist interviewees claimed they aesthetically appreciate social media posts (cf. Schreiber, 2017) - including the beauty of photos and pictures posted on social media (cf. Miller and Sinanan, 2017) which, they state, can also help engage fellow scholars or the general public. I argue that this aesthetic appreciation motivates social scientist participants to explore computational methods that allows them to analyse social media posts.

Finally, I argued that the non-academic research interviewees pursue parallel to their academic careers results in a more diverse set of geosocial scholars in terms of their disciplinary background. Many sought non-academic research to cope with financial pressures and job insecurity. As argued in section 2.9, the latter have been highlighted as an important cause for stress among graduate students and faculty alike. In addition, a number of interviewees work at academic institutions that depend on funding received from (or in partnership with) non-academic institutions. The commercial relevance of geosocial data research may fit well with institutional arrangements fostered by universities' attempts to develop graduates with industry-relevant competencies and research practices (c.f. Jackson, 2009) as discussed in section 2.9. Some participants pursue non-academic research parallel to their academic careers due to longstanding interests in computing - such as the open source software movement (cf. Kelty (2008)).

Non-academic research prompted all interviewees quoted in section 4.3 to learn computational data analysis skills, and some first worked with geosocial data through their non-academic work. Such, non-academic research altered their academic research focus. For example, non-academic research or collaboration prompted Luke, Anne, Colin, Henry and Elias to explore the affordances of computational data analysis and geosocial data for academic research. Thus, in contrast to the state of the field Savage and Burrows (2007) described over a decade ago - when academic and non-academic social research developed largely independently from one another - my interviewees' narratives suggest that academic and non-academic social research shape one another. Participants' use of computational methods they learned outside of academia for geosocial research also illustrates the 'thick provenance' of social research methods - their origins within and outside of academic research (cf. Law, Ruppert, and Savage, 2011). Altogether, my interviewees' experiences suggest that rising financial insecurity and job precarity in academia discussed in section 2.9 foster a mingling of academic and non-academic social research.

Chapters Five and Six will further discuss how interviewees overcome the challenge of combining computational data analysis and socio-spatial interpretation. Chapter Five will argue that collaborating with scholars with complementary skills and experimenting with computational data analysis methods help interviewees combine these and develop diverse geosocial research approaches. Chapter Six will discuss interviewees' efforts to create academic homes for their geosocial research which allows them to develop their own approaches to geosocial research.

Chapter 5

Collaboration Among Geosocial Researchers

Through exploring *how* interviewees combine computational data analysis and social scientific research as they develop their approaches to geosocial research, this chapter addresses all three research questions.

In response to the First Research Question - which explores how scholars develop geosocial research approaches - it discusses three practices: collaboration with scholars with complementary skills, setting up their own 'geosocial laboratories' and experimenting with computational data analysis methods as they search for geosocial data patterns. These help interviewees overcome the challenge of combining computational data analysis and socio-spatial interpretation discussed in Chapter Four, and are respectively discussed in sections.

I discuss collaboration practices in two phases. Section 5.1 argues that interviewees seek collaborations with scholars with complementary skills pairing social science and computational analysis. Then, section 5.2 draws on interviews and scientometrics to argue that such collaboration enables the development of diverse geosocial research approaches. I argue that participants seek collaborators with complementary skills with whom they share theoretical or methodological common ground. Through such collaborations, interviewees set up diverse geosocial laboratories - infrastructures to access and analyse geosocial data - on their own terms, and develop distinct approaches to geosocial research. The modularity of the co-authorship network reflects the importance of multiple small scale collaborations. Section 5.3 argues that experimenting with data analysis methods in search of patterns also helps interviewees combine computational data analysis and socio-spatial interpretation and thus develop their own approaches to geosocial research.

Regarding the Second Research Question - which explores differences among approaches to geosocial research - section 5.3 argues that the temporality or rhythm of combining computational data analysis and socio-spatial interpretation differentiates geosocial research approaches. While several scholars with technical backgrounds narrate the combination of these research traditions modularly or sequentially, the majority of social scientists state they perform them iteratively. This signals a difference in how interviewees create valid knowledge claims. In the former case, valid data patterns are identified mainly based on computational criteria. In the latter case patterns are identified using a combination of computational and social scientific criteria.

Finally, addressing the Third Research Question - which explores the project's mixed methods strategy - this chapter combines interview analysis and scientometrics to trace the intensity and meaning of relations among homogeneous actors: collaboration among geosocial scholars. Using interview analysis, it discusses *why* and *how* my participants collaborate with fellow scholars. Co-authorship network analysis helps evidence *the extent* to which geosocial scholars collaborate.

5.1 Importance of Collaboration

This section argues that the majority of my interviewees seek collaborators with complementary skills. Some consider collaboration desirable for geosocial research, and many find it essential. As section 4.1 argued, my participants stated that geosocial research requires them to combine socio-spatial interpretation and computational data analysis. However, they emphasize the challenge of mastering both methodological traditions. As this section argues, they seek collaboration with colleagues with complementary skills to help combine these research traditions. As section 5.1.1 discusses, social scientist interviewees seek technical collaborators in order to help conduct or optimise computational data analyses. As section 5.1.2 argues, participants who perform computational data analysis seek social science collaborators to help identify 'relevant' or 'novel' research questions and help interpret computational findings.

5.1.1 Seeking Computational Collaborators

Many participants with social science training emphasized the importance of collaborating with scholars with computational expertise - a view held both by both interviewees with little computational training and those who actively conduct computational research. In the quotes below, they stated that technical scholars help perform diverse computational tasks, including helping with specific methods and optimising computational analyses.

Firstly, Bruno and Daniel, who do not perform computational data analysis, seek collaborators to help them with diverse computational data analysis tasks. Bruno, as section 5.3 will discuss, contributes social theory and performs exploratory GIS analysis. He seeks computational collaborators to help conduct computational, statistical analyses and (interactive) data visualisations.

> *"[my collaborator] can access, scrape data and do all that stuff which is obviously important whether you're dealing with social media data. [He can also perform] statistical and spatial analysis, and [...] [he also] has [...] visualization skills that increasingly people care about..."*

Likewise, Daniel seeks computational collaborators to help him analyse social media data. He noted that he was interested in using social media data to study urban practices, but was not sure how. He thus sought a collaborator with programming skills.

> *"I developed a strong interest in [big data, complexity methods] [...] I also became interested in social media [and its constitutive role in urban practices]. [...] [such research with social media data is] something that I would like to do, but I don't really know how. [...] I was very happy to see that [David] applied [to the job], partly because he had the programming skills..."*

Secondly, Elias, Gary, Brian, Luke have backgrounds in social science and act as computational collaborators with social scientists colleagues, as well as Ben who has computational skills, are interested in collaborating with computational analysts who specialise in machine learning.

> *"If it's about [a specific line of research], I need support with doing the machine learning I'd have to do." – Elias*

> *"[For one of the projects], machine learning would be a cool method. And this would be someone who is better, has a better background in the methodological questions." – Gary*

> *"Secondary interest would be computer scientists because they would be able to help me refine my machine learning approach." – Luke*

> *"...some sort of machine learning, deep neural networks, artificial intelligence kind of thing. But I will say I don't feel qualified to sort of say which one would be [...] I have sort of feelers out with various people who are coming from a computer science or statistics background..." – Ben*

Thirdly, Henry and Brian - who, as Chapter Four discussed, both have computational expertise - seek computational collaborators to help optimise computational speed. Henry stated that specialists in specific computational techniques have detailed, relevant knowledge about hardware for this:

> *"I look for very very domain specific people who know a lot about very technical aspects, so things like computer vision or hardware, and people who can solve the problem I have from a CPU so it can actually finish computing before I die."*

Brian noted that computational collaborators are essential because social media research challenges are difficult to resolve through singular disciplinary approaches. For example, he said that experts in computational data analysis can help with data management and to improve computational data analysis by providing "elegant" and "efficient" analyses.

> *"...there are all kinds of technical challenges to even [...] accessing the data and storing it and making it kind of searchable and accessible. And to do that all from one kind of disciplinary home – that's quite difficult... sometimes working with a physicist can be great because they can come up with a really like efficient, elegant way to quantitatively make sense of a certain dataset..."*

Altogether, many interviewees with social science backgrounds seek computational collaborators to help perform computational data analysis, assist with specific computational methods, or optimise computational speed. Next, I discuss why participants with computational background seek social scientist collaborators.

5.1.2 Seeking Social Collaborators

My interviewees who conduct computational data analysis agreed that collaborating with social scientists is beneficial to geosocial research for two main reasons: to help develop research questions and interpret geosocial data analysis.

As illustrated below, interviewees' noted three ways social scientists inform research question development. Firstly, Kevin stated that they help develop research questions which address gaps in the social science literature. He contrasted this with physicists' research without social science collaborators, which he states often re-invent existing knowledge in ways that discourage social scientists from future collaborations.

> *"Collaboration [with social scientists] is essential, because several times, physicists who arrogantly think they are smarter than everyone re-invent things that are well-known in social science or economics [...] This doesn't help collaboration and makes the community look bad..."*

Secondly, Josh and Jane believe that collaborating with social scientists helps formulate research questions 'relevant' for urban research. Josh noted that whilst he values models based on mathematical ideas, he observes that relevant questions for urban science do not necessarily coincide with models which he finds mathematically interesting. He stated his collaborators help choose relevant variables:

> *[Social science collaborators] mostly help identify relevant questions. [...] it might not be clear to me what's worth modeling, which question is truly important [and] relevant for the future of cities. [...] what should we optimise [models] for, and which contextual factors should we consider the effect of in an urban scenario [...] I like models which use interesting mathematical ideas. [...] But sometimes the most useful models are not the ones I find most interesting."*

Like Josh, Jane believes that social science scholars can ask questions which better reflect the empirical phenomena (urban processes):

> *"I'm very interested in collaborating with [social scientists] at my next research group, because [...] I think they will ask different questions*

> *from the data, because they know a lot more about the context of the phenomena I study."*

Thirdly, Mike and Henry stated that social scientists help apply computational methods to new research topics. Mike stated that having social science collaborators can bring invigorating new ideas:

> *"...for example, the summer before we collaborated with [a] very brilliant [social science] student [...] and, you know, the type of breadth of thinking, and [...] freshness of ideas was amazing [...] it's really complementary to what we do, and you know, enriches the understanding of what we do as well."*

Henry claimed that social scientist collaborators help apply computational methods to new research areas which could benefit from computational exploration.

> *"[I] also [look for] people who are really interested in substantive questions [...] that can benefit from the analysis I do, but are not necessarily fields where you would expect people to have a big computational background."*

So far the section showed interviewees' preference for collaborating with social scientists to co-develop research questions. In addition, Isaac and Jane stated that social scientists help interpret the results of data analysis. Jane noted that social scientists' knowledge about spaces they study help them interpret the computational findings.

> *"I'm very interested in collaborating with people [social scientists] at my next research group, because [...] they see completely different context, they interpret the data differently than me. That will be very interesting, and I think that's been missing all along."*

Isaac emphasized social scientists' theoretical knowledge helps re-frame data analysis results. His quote discussed in section 4.1 contrasts computational data analysis' focus on 'computable' mechanisms with social scientists' questions about the nature of social practices. To explore the latter, social scientists invoked theoretical knowledge that did not form part of the model to re-interpret the results of his data analysis.

Altogether, this section showed that the majority of interviewees seek collaborators with complementary expertise. Next, section 5.2 will argue that interviewees seek collaborators with complementary skills with whom they also share methodological or theoretical common ground. These collaborations develop diverse approaches to geosocial research.

5.2 Collaborative Problematisation

This section argues that interviewees seek collaborators with complementary skills with whom they share methodological or theoretical common ground. As section 5.2.1 argues, many participants who seek social scientific collaborators state the importance of methodological common ground. Section 5.2.2 discusses social scientists' preferences to work with computational analysts with whom they share theoretical common ground. In addition, the 'strength' of the common ground interviewees seek differs in both cases. At one end, interviewees seek collaborators with expertise in the theories or methods they also use. At the other extreme, interviewees seek collaborators interested in the theories or methods they use without detailed knowledge about them, or collaborators who perform computational methods differently. I argue that through seeking stronger and looser common ground with collaborators, interviewees develop distinct approaches to geosocial research. Section 5.2.3 juxtaposes interview analysis with co-authorship network analysis. I argue that the collaboration landscape of geosocial research remains 'modular' because through collaboration, scholars develop diverse approaches to geosocial research and set up distinct 'geosocial laboratories'.

5.2.1 Common Ground with Social Scientists: Methodological Principles

This section argues that participants with computational background seek social scientist collaborators with knowledge about computational data analysis methods. However, as illustrated below with quotes, I argue that the 'strength' of the methodological common ground participants seek differs in two ways: depending on their ideal collaborators' skill level, and computational interviewees' interest in collaborating given different approaches to specific methods. I argue that through seeking 'stronger' and 'looser' common ground with collaborators, computational interviewees develop diverse approaches to geosocial research.

Kevin and Josh seek social scientist collaborators trained in the computational methods they use and who perform these methods similarly to them. I refer to this as 'strong methodological common ground.' Through such collaboration, they advance geosocial research in line with existing research traditions. Kevin seeks social scientist collaborators trained in mathematics which is the foundation of the methods he uses, hoping that they share assumptions about data analysis. He stated that shared mathematical knowledge helps communication because it facilitates consensus:

> *"I look for open minded people with whom we are on the same wave length. [...] And it's also very helpful if the collaborator has [...] a strong background in mathematics, because then it's easier for us to communicate with them, because they will think similarly to us about a lot of things by default."*

In line with Kevin's quote above, Chase - who collaborates with Kevin - suggests that Kevin prefers collaborators who perform computational methods similar to his. During their joint research, Chase performed a computational analysis method popular both in physics and geography. As part of the collaboration, Chase understood that scholars in these disciplines perform the analysis differently. Kevin deemed Chase's approach incorrect based on mathematical principles and advised Chase to alter his method. However, as Chase explained, the physics approach could not capture the empirical phenomena he wished study. To better suit the empirical question, he altered the modeling approach, but found that the physics research community - including Kevin - was not interested:

> *"[the modeling method] is part of the standard repertoire in both disciplines [physics and geography]. But they do it differently. I realised this when [Kevin] told me that I did the modeling wrong. He explained the correct way to do it and its mathematical assumptions. I read up about it and I did it his way. But there are things the physics model can't do. [...] We later modified the model so we could put in extra variables, [...] but so far it looks like the physicists [and Kevin] are not too interested."*

Confirming Chase's account, Kevin narrated his contribution to interdisciplinary team work in terms of correcting technical details based on his mathematical, methodological expertise:

> *"For example, on one occasion my collaborators used the wrong formulae for community detection and I corrected it. This is how these*

> *interdisciplinary collaborations work, we add our knowledge based on our expertise."*

Josh stated he collaborates with urban studies scholars who conduct modeling similar to him, and that his expertise in modeling serves as common ground with collaborators. As his quote in section 5.2.1 showed, he noted a difference between his and his collaborators' understanding of the context and applicability of the models, rather than differences in modeling approaches.

> *"I get the impression that a significant body of sociological research area aims to model phenomena [...] I became part of an interdisciplinary research group, where I continue to focus on data analysis, and I try to build connections with others using my background in modeling with urban planners who also do modeling."*

Altogether, Josh and Kevin seek social scientist collaborators trained in the computational methods they use - which I referred to as strong methodological common ground. In contrast, as evidenced below, Mike, Jane and Isaac are open to collaborate with scholars with whom they share looser methodological common ground: either collaborators less-skilled in computational data analysis, or who perform computational analyses differently.

Mike seeks social scientist collaborators who understand the affordances of algorithmic data analysis 'at scale', but are not necessarily skilled data analysts. I describe this a form of 'loose methodological common ground.' He stated that such expertise is essential and sufficient to help apply computational methods to new topics. Through such common ground, his ideal collaborators can help create computationally adequate hypotheses in dialogue with theories from the social and behavioural sciences.

> *"it would be perfect if this person [the collaborator] had a sort of computational understanding [...] [otherwise] it's a bit of a struggle [... for example, if they understood] the things you can do with the data, the process of hypothesis building [...] [if they had] the ability to think about the data a bit at scale, not just because OK the guy can directly do the stuff, but also because it impacts the way of thinking about hypotheses around the problems that you have. [...] even just trivial things like understanding the data format of things, or have a rough algorithmic understanding of how you would parse some data*

> *[...] or even the things you could measure [with] algorithms. [...] People who never had this type of experience [...] the difference is striking!"*

Like Mike, Jane and Isaac are open to collaborating with social scientists less skilled in computational methods. They are interested in exploring different ways specific data analysis methods can be conducted. Thus, I understand the methodological common ground they seek loose. Through these collaborations, they apply computational data analysis methods to new topics. Jane noted that her new social scientist collaborators hired her because of her expertise in parallel computing. Her skills enable social scientist collaborators' geosocial research. Jane stresses that her collaborators' (less advanced) computational knowledge allows her to discuss technical details of her work.

> *"My future employer [Chase] [...] probably want[s] to hire me so I can establish parallel computing and data analysis processes quickly. [...] I'm very interested in collaborating with people at my next research group because they are social scientists who have a lot of technical knowledge too, so I can discuss the technical details with them as well..."*

In addition, as the next two quotes suggest, Jane is interested in reflecting on methodological differences. She noted her interest in understanding differences between her and her collaborators' approaches to regression analysis. She contrasted social scientific regression models, which include multiple variables, with her physics approach, which involves linear equations. She hopes her next collaboration can help her better understand the social science approach.

> *"I would like to get more experience in traditional social scientific modeling, for example, those big regression models. In a lot of situations when we [physicists] create a system of linear equations, social scientists run a regression model. I have never worked with those, and I find it very difficult to interpret them."*

When discussing what she considers good science, Jane noted her disciplinary training in physics influences her preferences:

> *"I think model building differentiates between simple data science and 'real' scientific research. Or it might be the case that I think this because this is what physicists learn..."*

Isaac also actively explores disciplinary methodological differences and seeks collaborators who are interested in his methods but conduct or interpret them differently. As his quote presented in section 4.1 showed, he noted that whilst his discipline focuses on variables' relationships, the social scientists he exchanged ideas with as part of a research visit approached the phenomena - and his results - in light of theories about social mechanisms, which in turn shaped his interpretations. However, loose common ground collaborations can be challenging, for example, because of uncertainty about the work's reception:

> *"I anticipated that they [collaborators during a research visit] would reject a lot the representativity of the data, they didn't do that so much, they just said be aware, I said of course... maybe they are used to seeing these things now so and they know they know it's hard to fight it..."*

Isaac claimed that his loose methodological collaborations with social scientists helps apply computational methods to study topics previously under-explored with computational methods. He visited a social science oriented interdisciplinary institution and later got hired by a computer science department. He credited his employment in computer science to the theoretical knowledge he obtained collaborating with social scientists, and stated that he has the most theoretical knowledge in his new group:

> *"During the job interview I spent a few slides on theory that I learnt from [social science collaborators] and after the talk one person gave the comment, 'that's the most theory I have ever seen'. Like it it was really different for them. [...] I think that they think that I can contribute to their group a lot because I have experience in turning circumstantial things, relating these kind of abstract concepts that are poorly quantifiable in a still relatively convincing way."*

Finally, both Jane and Isaac stated they seek collaboration with technical scholars and social scientists, each of whom offer unique contributions:

> *"when I [did my research visit], I really enjoyed engaging with [sociologists] and that part of academia, and probably [in my next job at a] Computer Science Department, I'll continue to miss that. But had I taken a job ... with the sociologists, perhaps I would miss the physicists. So I think if you are in an interdisciplinary thing, whatever is missing at the moment or on the project is what you miss in general." – Isaac*

> *"I really look forward to collaborating with the social scientists in my next job [...] I have missed that all along [my PhD]. At the same time I can imagine collaborating with a traditional physicist who does modeling and equations, they could complement my work from the other direction." – Jane*

Altogether, by alternating collaborations with technical scholars and social scientists with whom they share loose common ground, Jane and Isaac develop geosocial research in diverse directions. This section argued that through seeking social collaborators with whom they share stronger and looser methodological common ground, computational analysts develop diverse approaches to geosocial research. Next, I discuss how social scientists' ideal computational collaborators share theoretical common ground.

5.2.2 Common Ground with Computational Analysts: Social Theory or Computational Methods

This section argues that social scientist participants seek computational collaborators with whom they share theoretical common ground. As with computational analysts, the strength of common ground they seek differs. Firstly, I discuss interviewees' experiences seeking computational collaborators with strong theoretical common ground for projects to counter 'computational social science' or GIScience. Secondly, I discuss interviewees' narratives of the importance of looser theoretical common ground for methodologically-focused collaborations. Thirdly, with Brian's and Chase's narratives, I argue that alternating between collaborations using stronger and looser theoretical common ground helps advance geosocial research in diverse directions.

David and Bruno seek computational analyst collaborators with strong theoretical common ground for projects that explore 'good' uses of geosocial data. Bruno, whose geosocial research aims to counter GIScience-inspired approaches to geosocial research, states that his ideal computational collaborator is trained in social theories he uses. In his team, he specialises in social theory, and his ideal collaborator shares some of his knowledge.

> *"...our research largely tried to talk back to the more technical, computer science-dominated versions or even just GIScience dominated versions of this research [...] if I could clone [Brian] so there is lots of him, so that he has more time, it would be [the ideal collabora-*

> *tor]. [Brian] is a very unique guy because he has social theory, he has substantive knowledge about the field [...] but he's also a computer programmer and can [do data analysis]"*

Bruno's computational collaborator is Brian. Like Bruno, Brian has had a long-standing interest in social theory and countering GIScience-inspired geosocial research:

> *"...in a sense we're trying to push back against is this kind of adoption by let's say folks who come from engineering or computer sciences [...] now studying cities and makes very kind of grand claims [without] having read [...] any previous work that has been done in the last century."*

Like Bruno, Daniel's ideal collaborator is a computational analyst trained in the social theories he uses. As his quote in section 4.1 showed, he claimed that his shared theoretical interests with David is core to their collaboration. Like Bruno and Brian, Daniel and David share an interest in critiquing 'computational social science' through their geosocial research:

> *"then I realised, most of these people [computational social scientists], actually, couldn't care less about the social science part. [...] [but] we do need to do verstehen, we do need to situate, we do need to contextualise, and actually these new data help us do it" – Daniel*
>
> *"...everybody was so excited by this research [at a computational social science conference], and I couldn't help feeling like 'so what'? [laughing] this doesn't answer any of my questions around, like, what the digital means, or [...] what this does to our interactions with one another. I guess I have a sensibility that's shaped by kind of classical sociology, like 'verstehen'..." – David*

They contrasted their approach with other computational social research informed by 'methodological individualism'. Daniel stated that he once decided against hiring a scholar - a strong candidate with computational and sociological training - who thought of geosocial data in terms of individual users. He noted that using geosocial data to study relationships and their change over time is also a new challenge for him. Thus, he seeks collaborators like David to co-design methods in line with his theoretical commitments.

> *"we had a great candidate [who] said 'I would get all the information about the individual [social media] users' [... then] we asked 'what*

> *if we can't get all the information about the individual [social media] users? what if you only have the patterns? [...] then she was like 'well, you're asking me to step outside my own paradigm'... and yeah, I mean it's something that I want to do, but I don't really know how. [... I realised] whooa! Actually people are really trained to think about their data in terms of a matrix... [and...] individual attributes [...] [In contrast, looking at relationships] is a hugely important point of our work..."*

Altogether, Bruno and Daniel seek computational collaborators with whom they share strong theoretical common ground for projects that wish to counter 'computational social science'. Next, I discuss social scientists' experiences who seek loose theoretical common ground with computational collaborators.

As section 5.1.1 argued, several social scientists seek computational collaborators with shared interest in specific computational data analysis methods. Elias', Ben's, Brian's and Chase's narratives demonstrate the importance of looser theoretical common ground in collaborations with computational scholars rooted in shared methodological interests. Elias and Ben seek computational collaborators skilled in machine learning. Although their ideal collaborators are not necessarily trained in social science, they emphasized the importance of interest in social scientific questions and frameworks. Elias, who wishes to apply machine learning to a research topic in his discipline, seeks collaborators with computational expertise and theoretical open-mindedness sensitive to (but not necessarily trained in) social science:

> *"[for this project I seek] an expert in computer vision who can do a lot of things that I can't. [...] But [...] that kind of ideal colleague would have to be someone who is [...] not necessarily a sociologist but someone who can appreciate and understand the sort of complexities of that discipline and also build some connections [...] someone who's a bit ecumenical in their in their outlook. I think it's something we all have to be ready to do."*

Ben also emphasized the importance of shared interest in research questions when discussing his interest in collaborating with machine learning experts (see also section 5.1.1). He stressed that computational collaborators should share an interest in social scientific topics and the lack of such common interest hinders collaboration.

> *"the trick with that is always how to make it interesting for the collaborator [...] the kinds of questions I'm interested in are not necessarily interesting to a statistician who is more interested in what method can you use. [... they may not be interested in the questions I ask] because they're social science questions."*

Brian and Chase reflected on collaborating with computational scholars with whom they share methodological interests but whose theoretical interests differ. Their experiences show that such collaboration, if successful, can help knowledge exchange among scholars from different backgrounds. However, ethical and epistemological differences can pose challenges, highlighting the importance of loose theoretical common ground. Brian collaborated with computational scholars on a predictive algorithm to study urban change. The algorithm had commercial applications, which his collaborators saw as a benefit and opportunity to monetise. In contrast, Brian hoped to apply the algorithm to influence urban processes contrary to such 'market logic'. He stated he may encounter such potentially-hindering differences when attempting collaboration with computational scholars.

> *"[collaboration with technical scholars] is a big hit or miss [...] sometimes we just speak such a different language and maybe our goals are so far apart that it's difficult to come together [...] [In one project] I'm very much interested in neighborhood change, issues like gentrification, but [my collaborators] always went back to real estate price increases and that was [...] awkward [... because] they saw as a really useful application [...] if you could give this to a real estate developer or a property speculator [...] I'm more interested in maybe the same algorithm or the same method, but actually to make sure that we put limitations on them. [...] So there's a bit of a tension there"*

Like Brian, Chase collaborates with computational scholars with whom he shares methodological interests, but lacks theoretical common ground. He seeks to expand his computational data analysis skills. As section 5.2.1 discussed, he trained himself and his team in computational data analysis methods to facilitate successful collaboration with physicists. In contrast to Brian who agreed with computational collaborators about the modeling approach but disagreed about their method's applications, as section 5.2.1 discussed, Chase found that the modeling approach advocated by his computational collaborator - which he was interested to learn - could not account for the empirical phenomenon he studied. In particular, as the quote below shows, he claimed the physics modeling method was not suitable to investigate the relationship among multiple variables. Thus, he altered the model, but as discussed in section 5.2.1, he then found that physics scholars were not interested in it.

> *"But there are things the physics model can't do... [...] Those types of models only work in one dimension, for example, distance. If you have other variables, like similarity or a network proximity measure, you can't use them to model the relationship between these. [...] In this case we modified the model so we could put in extra variables..."*

Altogether, Brian and Chase collaborate with technical scholars with whom they share interest in data analysis methods amid theoretical differences. Their experiences show that such collaboration can help knowledge exchange, but the lack of theoretical common ground can pose challenges.

Finally, Chase's, Brian's and Elias' journeys show that alternating between collaborations with looser and stronger theoretical common ground also helps diversify approaches to geosocial research. As discussed above, Brian seeks collaborators with diverse disciplinary backgrounds, including Bruno, with whom he shares theoretical common ground, and physicists and engineers with whom he shares methodological common ground. Like Brian, Chase collaborates with scholars from his disciplinary background, as well as with physicists with whom he shares methodological common ground. Alternating between such collaborations allow Chase and Brian to advance geosocial research in line with their disciplinary background as well as the research agenda of technical scholars.

Like Chase and Brian, Elias seeks both computational and social scientist collaborators. His quote above discussed his search for machine learning experts. As the quote below shows, he also acts as a computational collaborator with social scientists with whom he shares theoretical and disciplinary common ground, in a research field he is familiar with but is not an expert:

> *"I don't really care if they [my collaborator] are an expert in cultural studies or data science because I can do a bit of both. [... my ideal collaborator is always] driven by a particular kind of research agenda. [For example, on another project] I work with a [social scientist] and*

> *I can support [them] with data science methods. That's a great way of doing collaboration for me as well. But we both know the field. [They are] just much better trained at writing in a way."*

Altogether, I argued that social scientist interviewees develop diverse approaches to geosocial research by seeking technical collaborators who share stronger or looser common ground with them. Next, I conclude the section by juxtaposing the above findings with co-authorship network analysis.

5.2.3 Co-authorship Patterns and Geosocial Laboratories

Based on interviewees' quotes, this section argued that the stronger and looser common ground participants seek with collaborators with complementary skill sets results in the development of diverse approaches to geosocial research. The co-authorship analysis depicted by figure 5.1 - which shows that the collaboration landscape of geosocial research remains modular - supports this argument. As section 3.4.1 explained, figure 5.1 compares the modularity of the co-authorship network (red line) with statistically comparable networks (1000 random graphs with equal number of nodes and edges) at four time points. The blue lines show the minimum and maximum of the modularity distribution of the 1000 simulated random graphs. The figure shows that the modularity of the co-authorship network remains considerably higher than the modularity of the random graphs. Over time, the difference between the modularity of the empirical and randomly generated networks slightly increases. This suggests that over time, the modularity of the co-authorship increases, even as its absolute value remains the stable. Interviewees' narratives and the co-authorship network analysis show that scholars' collaborations remains modular - geosocial research does not become an integrated, coordinated research field. Through collaboration, interviewees advance geosocial research in line with diverse disciplinary research traditions. Next, I show the modular nature of collaboration from a second perspective - participants' narratives about setting up their geosocial laboratories.

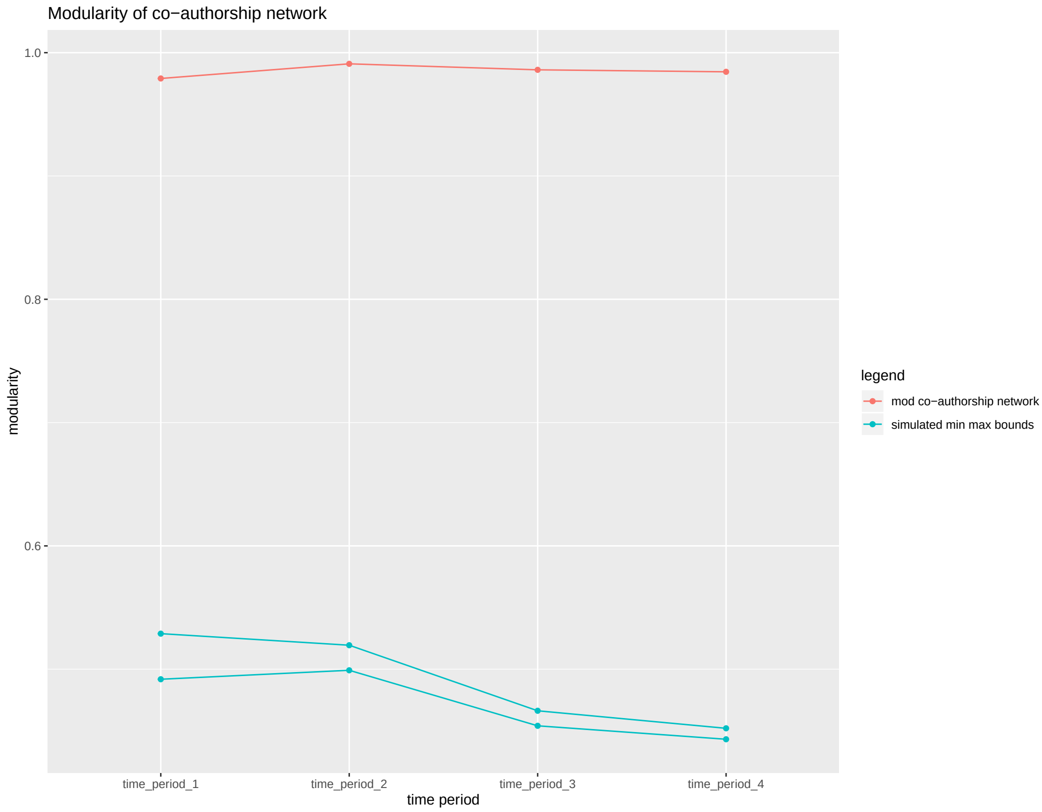


Figure 5.1: Modularity of the co-author network compared to the modularity of 1000 simulated random graphs with equal number of nodes, edges and edge weight distribution over time

Interviewees with varying levels of knowledge about computational infrastructures set up their own 'geosocial laboratories'. Most of them, including participants with relatively little computational training, succeed in accessing and analysing geosocial data on their own or in collaboration with colleagues with whom they share epistemological and methodological common ground. This helps them develop their approach to geosocial research, without having to negotiate methodological or theoretical differences. As the quotes below illustrate, participants use diverse methodologies to access and analyse geosocial data. Some, such as Anne and Elias, access data as .txt or .csv files. Others, - including groups B, D and J - build databases, and groups C and K purchase a proprietary database of social media posts.

Anne and Elias accessed geosocial traces in standalone files. Anne downloaded them as .txt files through the social media platform's API, and uses proprietary software for data analysis. The software allows her to analyse the data she obtained through the API without further data processing (for example, it does not require changing the file format).

> *"There was definitely a time where I [thought] I need to get somebody who can write me code so that I can analyse all this data really quickly. But I'm glad I didn't commit to it in the end, because that's precisely the opposite of why I started the project in the first place. [...] I chose to use [this data analysis software], because it allowed me to upload the data in its entirety, without [having] to do like any cleaning of the data, I didn't have to do any parsing, or any processing of any kind. I could upload the raw text files obtained through the API straight to [the data analysis software].*

Like Anne, Elias explained that he downloaded geosocial data from the social media platform's API, which he analyses with a combination of programming and custom-made software:

> *"I basically grabbed all of the [social media posts] that I could, and then I put it all into a network graph. And basically, then I used a layout algorithm [...] in [network analysis software]"*

Whilst Anne and Elias worked with data files they downloaded through APIs, other groups built or bought databases which facilitate data sharing and collaborative research. For example, interviewees from three groups - B, D and J - discussed building their own databases. Participants affiliated with group J (who were not interviewed for this project) used their experience with astronomical databases to build a social media database. They optimised the database for ad hoc queries and statistical analyses they were familiar with from previous research. Jane stated that her colleagues who collected the database prior to her involvement in the project had extensive related experience.

> *"...a data collection like this needs to be carefully planned in advance. You have to think through how to structure the database, how you'll collect the data: which computers you'll use, how you load the data into the database and how you make it searchable. My colleagues who collected the data had a lot of experience with databases from prior work with astronomical databases."*

Like group J, interviewees in groups D and B store geosocial data using databases. Their experiences highlight the effort required to build databases. As section 4.1 discussed, David collected data for group D, but he feels that the time he invests in creating and maintaining the database is not valued by his disciplinary peers. Similar to David in group D, as section 4.1 discussed, Brian had computational expertise to collect the geosocial data which group B used. Ben states that Brian's PhD's flexibility allowed them to build the database even though they had relatively little experience.

Altogether, interviewees with varying computational skills successfully obtain and analyse geosocial data on their own or in collaboration with scholars with similar research interests. As section 5.3 discusses, setting up their own laboratories allows interviewees to develop geosocial research in line with their own methodological and epistemological interests through experimenting with data analysis methods. Next, I discuss how experimentally exploring data patterns helps interviewees successfully combine computational data analysis and socio-spatial interpretation.

5.3 Variable Experimentation

This section argues that the 'experimental' use of diverse data analysis methods to identify patterns helps interviewees develop their approaches to geosocial research, and that the rhythm of such experimentation differentiates approaches. I use the term 'variable experimentation' to refer to two aspects of participants' geosocial research. Firstly, interviewees from research groups A, B, C, D, J, K, L and M emphasized their 'experimental' use of diverse data analysis methods to explore possible research questions and 'variables' of interest, and thus identify patterns. Secondly, I argue that interviewees' experimentation is 'variable': their narratives differ about the temporality of combining computational data analysis and socio-spatial interpretation (sequentially versus iteratively) as they experimentally search for patterns. I argue that these also signal different approaches to validating data claims. Below I illustrate participants' variable experimentation by first discussing the narratives of those who understand the combination sequentially, followed by narratives about iterative computational and interpretative experimental data analysis.

Firstly, as the quotes below show, Miles, Luke, Josh and Jane experiment with various computational data analysis methods in search of patterns. Then, I illustrate their preference to combine computational data analysis and socio-spatial interpretation sequentially - using social scientific knowledge either to inform research questions and variables of interest or interpret results. I argue that they validate data patterns based on computational and statistical criteria, excluding social scientific considerations from that aspect of the process.

Experimental data exploration is central to Miles', Jane's, Luke's and Josh's geosocial research: they recount experimenting with various data analysis methods to identify valid patterns with geosocial data.

> *"The biggest challenge was the workload: in 3 months we had to develop a publishable project - we had to find the right direction - so it was pretty intense. We tried a lot of methods..." – Miles*
>
> *"[my supervisor] said we should look for a topic that will yield publishable results. Already when we started this project he knew that we would get results because they had already done previous calculations with [Josh]." – Jane*
>
> *"My [geosocial research] originally had a stronger focus on event detection [...] [but] [I changed methods] [...] [Later,] I realized that cleaning the data was actually a lot more complex [...] So a lot of my work has been [about] advanced or more complex cleaning methods" – Luke*
>
> *"We didn't have a very precisely stated goal when we started, rather, we explored in general what [types of] analyses we can do, how can our data yield interesting results." – Josh*

On one occasion, Miles' group considered a more complex calculation, for which they contacted another scholar who published about the topic but didn't reply, and the idea was discarded. This shows that experimental exploration of variables and research methods can benefit from input from collaborators.

> *"Another feature, which would have been pretty complicated to calculate, would also have used network relations. I found a relevant paper and e-mailed the author. I wanted to ask them if it would even be worth to try this calculation with our data. But they did not reply, and we just discarded the idea."*

Altogether, Miles, Luke, Josh and Jane recounted experimenting with diverse computational data analysis methods in search for patterns. As the quotes below show, they stated that they use social science knowledge either as an input to inform the research question, or after having completed computational data analysis, to help interpret the results. As section 5.1.2 discussed, Josh and Jane seek social scientist collaborators to help identify research questions and variables to include in models. Like them, Miles stated that social scientific theories can inform variables models include:

> *"Usually when we do predictive modeling, we look for variables which can explain our predictions. We can choose variables based on intuition, theories or previous research."*

Like Josh, Jane and Miles, Luke's narrative suggests that he uses socio-spatial interpretation to identify research questions. He stated that he seeks collaborators who can help apply his method to new topics.

> *"I would really like to work with a [social scientist] [...] to apply my methodology to a field that I'm unfamiliar with."*

In addition, Miles claimed that social scientific knowledge can help interpret data analysis results.

> *"...we develop our interpretation when we write up the paper. I mean, the trends in the data are clear. But then we look for good examples, or maybe link them to findings in previous papers or even news stories..."*

As discussed so far, Miles, Josh, Jane and Luke experiment with diverse methods to identify data patterns, and use social scientific knowledge either to help identify variables and research questions, or to interpret data analysis results. Using the quotes below, I argue that once they identified variables and research questions, they identify patterns with geosocial data based on statistical, computational criteria alone: they validate patterns without explicitly considering social scientific knowledge.

Josh stated he finds social media research more complicated than physics projects in the sense that social media phenomena can be studied from multiple valid perspectives. It is challenging to control or repeat measurements, and account for all possible causes for variation. Given these constraints, he stressed that searching for statistically significant patterns is a good approach to analysing social media data.

> *"Compared to physics, where you can do the same experiment a hundred times, and you can decide which model is good [with geosocial]*

> *research there are a lot more ways to approach topics... Often times it seems very complicated. For example, we can't claim to explain why people post certain things, because it depends a lot on individuals. But it's a good question to ask what statistically significant patterns we can find in the data [...] [but] it's a lot more difficult [than physics] because we can't do controlled experiments..."*

Jane explained that in one project, her team identified valid patterns based on the outcomes of quantitative data analyses, whilst being initially agnostic about what they would find. They included the outcomes of all the statistical analyses in their resulting paper. Once the analyses were complete, Jane gathered knowledge about the country by reading and talking to colleagues to write up the results.

> *"...for that project [...] we correlated social media data with any data we could access for the same spatial units and time window. [...] we included all [statistical patterns we identified] in our paper. [... then] I had to read a lot and talk to people who lived in the country to write up the results."*

Miles stated that his team also validates patterns using statistical and computational criteria. He recounted two reasons for the failure of data analysis methods they experiment with, both invoking statistical criteria. Firstly, he claimed that sometimes results were 'too noisy' - meaning his team could not discern statistically clear patterns. Once they identified patterns, they sought recurrent patterns which they could identify computationally.

> *"We tried to analyse the data on several temporal scales, and decided to use [the one] which yielded the best results [...] meaning that this yielded the clearest statistical trends. [...] And then we were interested if there were patterns in the trends. [...] In the end we identified [x] clusters."*

Secondly, confounding processes obscured some results they obtained, which prompted the team to use different variables. Several statistical methods assume the 'independence' of variables. When independence is not guaranteed, attempting to disentangle the effect of specific variables from that of confounding variables is a core challenge in statistical research.

> *"We realised that even though our methods counted for likes, we still found that [entities] that got more likes were more similar [based on*

> *the feature]. So the number of likes confounded our feature, so we abandoned it."*

Finally, Luke hopes to apply his data analysis method to diverse research topics. He claimed that the algorithms he created can be used to study diverse topics, without changing them. Like Josh's, Jane's and Miles', his method identifies patterns based purely on computational and statistical criteria.

> *"...my [research] is a methodological contribution. So if you have a topic of interest you can apply my method to it, and it'll spit out relevant [social media] data and maps. So it's very versatile and very dedicated to the cleaning method [...] [it yields] more reliable [results]..."*

Altogether, Jane, Josh, Miles and Luke stated they combine computational data analysis and socio-spatial interpretation sequentially, and they validate patterns based purely on statistical and computational criteria. In this sense, their practice is similar to Kevin's advice to Chase to conduct computational analysis in line with mathematical validation criteria independent of the characteristics of the socio-spatial phenomenon (discussed in section 5.2.1). Next, I discuss interviewees' quotes who narrate combining computational data analysis and socio-spatial interpretation iteratively.

Interviewees from groups A - D narrated the relationship between computational data analysis and socio-spatial interpretation iteratively. In contrast to the interviewees quoted above who seek patterns based on statistical and computational criteria, many participants quoted below state that they experimentally identify patterns using a combination of computational, statistical and socio-spatial interpretative criteria.

Anne explained that she experimentally explored geosocial data using various methods afforded by the data analysis software, and informed by her social scientific frameworks. She iteratively used statistical data analysis and fine-grained analysis of social media posts' content.

> *"the [software] allowed me to kind of zoom in and out of the data, so I could run like quick frequency queries, or make some statistical analyses [...] but at the same time I could drill down and like zoom in...."*

Like Anne, interviewees from groups B, C and D state that they experimentally explored geosocial data and sought patterns by iterating computational data analysis and socio-spatial interpretation. In addition, they claim that they modified computational data analysis in light of socio-spatial interpretative knowledge. Bruno explained that initially, he explored data with GIS to look for 'stories' and later Brian, his colleague with computational expertise, performed statistical analyses. They claimed that they repeat such exploratory and statistical analyses iteratively to identify valid patterns.

> *"...we work in an iterative way, usually bouncing back and forth [between] statistical, kind of big picture view [...] I use GIS for a lot my research. I do a lot of the exploratory stuff, dig into the nitty-gritty to try to find the interesting stories. And then usually the way we do things, I do that and [Brian] will dig in, create a new, kind of more robust analysis, I'll go back and I'll dig in more to the [GIS stories] type thing."*

In addition, Bruno stated that they modified computational data analysis methods using their local knowledge of the places they study. He claimed that their local knowledge shapes not only their research questions and interpretation - as with Jane - but also their methods. Thus, their criteria for identifying valid data patterns combined computational and social scientific criteria.

> *"we like to [study] places where we live, places that we're familiar with... [We look for projects] where we can not just run the numbers on [social media] data [... but] connect what we see in the data to stuff that we know that has nothing to do with what's in the [social media] data, and allowing that [...] local, experiential knowledge [...] to shape our questions, our methods, and the ways we interpret..."*

Like the interviewees above, Chase explained that they experimented with various data analysis methods in search for patterns with the data.

> *"...we kept working for months, trying various approaches... That was pretty tough. When we finally had something, I contacted [Kevin, the natural scientist] and asked if he was interested in having a chat."*

As section 5.2.2 explained, like Bruno, Chase states that they modified a computational data analysis technique in light of social theory. They modified a method borrowed from physics, developed to study the relationship between two variables, to include more variables to better reflect the empirical phenomenon. Like group B, they used a combination of computational and social scientific criteria to devise their method to identify valid data patterns.

Like Bruno and Chase, David and Daniel claimed that they experimentally identify patterns by combining computational and social scientific criteria. They state that they articulated, and evaluated their assumptions and expectations - informed by social theory - in light of the outcomes of exploratory data analysis. In addition, as section 5.2.2 explained, they claimed they developed computational data analysis methods informed by social theories which focus on relationships and their temporality.

> *"...[we were] really just exploring it [geosocial data] with anything we could think of. [...] Sometimes we would search for specific keywords [...] we [also] try to be quite upfront that we had certain ideas about this data and then allowed ourselves to also be surprised..."*

In sum, interviewees from groups A - D state that they experimentally explored geosocial data by iterating computational data analysis methods and socio-spatial interpretation. Moreover, participants from groups B - D state that they modified or created computational data analysis methods informed by social scientific theories. Thus, they claimed that they identify data patterns using a combination of computational, statistical and social scientific criteria.

Altogether, this section argued that experimentally using data analysis methods to identify patterns with geosocial data is central to interviewees' developoment of their geosocial research approaches. I also argued that approaches to geosocial research differ in terms of whether interviewees combine computational data analysis and socio-spatial interpretation sequentially or iteratively, and whether they validate claims based on purely statistical criteria or also considering social scientific knowledge.

5.4 Conclusion

Through exploring *how* interviewees combine computational data analysis and social scientific research as they develop their approaches to geosocial research, this chapter addressed all three research questions.

In response to the First Research question - which explores how geosocial research approaches develop - I highlighted three practices: collaboration with scholars with complementary skills, setting up 'geosocial laboratories' and experimenting with computational data analysis methods in search of geosocial data patterns. Through combining co-authorship network and interview analyses, I argued that collaborations between scholars with complementary skills helps them develop diverse approaches to geosocial research. Interviewees seek collaborators with complementary skills with whom they also share common ground. Computational analysts seek social scientist collaborators who share methodological common ground with them, and social scientists seek computational collaborators with whom they share theoretical common ground. Social scientist interviewees who collaborate with computational scholars who do not share their theoretical commitments recount challenges, such as the need to negotiate ethical differences and the inadequacy of computational methods recommended by collaborators to study research questions.

Maniglier (2019) proposed that scientific research creates 'problems' (research questions) through relating diverse concepts pertaining to specific research traditions (cf. Callon et al., 1983), also in dialogue with other research traditions' approaches to raising questions. Drawing on this framework, I argue that through seeking common ground, my participants seek collaborators with complementary skills with whom they can co-create research questions.

As section 5.1.2 argued, computational scientist participants seek social scientist collaborators to help identify 'relevant' research questions. As section 5.3 argued, computational research methods guide the way interviewees with computational backgrounds create knowledge (cf. Bateson, 1972). Thus, I argue that collectively identifying research questions with social scientists necessitates a shared understanding of the affordances of computational methods with computational scholars. In Maniglier's (2019) terminology, shared methodological common ground helps to link concepts during the problem creation process.

As section 5.1.1 argued, social scientist interviewees seek computational collaborators to help perform computational data analysis. Section 5.3 argued that during geosocial research, both computational methods and social theory inform the way social scientist participants validate geosocial research findings. I assume that social theory also shapes the questions social scientists ask. The example of actor-network theory, which Latour (2005b) differentiates from Durkheim-ian sociology based on its explanatory concepts and units of analysis, supports the assumption that social theory guides the types of explanations and questions social scientists can pose. Thus, drawing on Maniglier's (2019) framework, collaboratively identifying research questions with computational scholars necessitates shared understanding of theoretical concepts that they link to create research questions.

However, the strength of common ground interviewees seek differs. Seeking collaborators with strong common ground enables interviewees to develop geosocial research in line with existing research traditions. Collaborating with social scientists who share strong methodological common ground with them helps physicists Kevin and Josh develop existing approaches to geosocial research, such as urban modeling. Social scientists Bruno and Daniel seek computational collaborators with strong theoretical common ground - trained in the same social science theories they use. Their geosocial research critiques computational social science and GIScience informed by social science frameworks. In addition, Brian's and Chase's narratives highlighted ethical and epistemological challenges scholars who collaborate without shared theoretical interests can face further highlighting the importance of theoretical common ground with computational collaborators.

At the same time, participants who collaborate through looser common ground exchange knowledge with scholars from complementary disciplines and use this knowledge to apply computational methods in new ways. Based on Elias', Brian's and Chase's examples, I argued that collaborating with computational scholars with whom they share loose theoretical common ground, and acting as computational collaborators to social scientists who share strong theoretical common ground with them helps them develop diverse approaches to geosocial research. Similarly, Jane's and Isaac's movement across research groups is enabled by loose methodological ground with social scientists collaborators, and stronger methodological common ground with computational collaborators. The skills they bring to their respective teams are core to the these teams' success in geosocial research. Isaac contributes social scientific theories he learned through collaborating with social scientists to geosocial research at a computer science department. Jane contributes computational knowledge she learned through her collaboration with experienced physicists to group C's work, where most of the others have background in social science.

Using the notion 'experiments in participation' Lezaun, Marres, and Tironi (2017) introduced in section 2.11 as a lens through which to interpret collaboration practices, collaborating amid loose common ground can be seen as an experiment to include scholars in one's geosocial research who might alter its direction in unexpected ways. The loose common ground facilitates these experiments in participation by providing opportunities for local coordination. However, as Brian's and Ben's narratives suggest, the success of these experiments to yield geosocial research is uncertain.

I also interpret the varying types of common ground interviewees seek drawing on Maniglier's (2019) argument that transdisciplinarity requires introducing comparative methods among disciplines, and that scientific questions are raised as scholars experience how different knowledge traditions ask questions (discussed in section 2.10). Using these insights, my interviewees' aim to collaborate with scholars with complementary expertise can be understood as efforts to introduce comparative moments between disciplinary research practices, which allow them to (better) define their geosocial research questions. However, participants who seek stronger and looser common ground experience differences between ways of knowing in settings with different stakes. Participants who collaborate with loose common ground experience such differences whilst they also have to interpersonally negotiate analytical decisions with those who practice different ways of knowing. Collaborators who seek stronger common ground either experience such differences to a lesser extent or without the need to negotiate such differences interpersonally.

Interviewees' narratives can also be interpreted through Osborne's (2013) insight that disciplines have distinct propensities to relate to one another. I find distinct ways computational and social scientific disciplines relate to one another during geosocial research. On the one hand, computational interviewees' search for methodological common ground with social scientific collaborators - and the capacity of computational methods to provide common ground between computational and social scientist interviewees even amid theoretical differences - illustrate that computational disciplines relate to other fields through the capacity of mathematical, statistical and computational data analysis methods to travel across disciplines (cf. Mackenzie and McNally, 2013; Knuuttila and Loettgers, 2014; Osborne, 2013). On the other hand, social scientist interviewees' search for strong theoretical common ground with their collaborators when attempting to differentiate their geosocial research from computational social science or GIScience signals the fundamental role of theoretical assumptions - which inform scholars' decisions about explanatory units in the interpretative sciences - for social sciences to interface with other research areas.

However, while Osborne (2013) primarily notes epistemological reasons for such differences, my analysis highlights the role of researchers' personal interests. Physicists Kevin and Josh seek social scientist collaborators who conduct computational data analysis similar to theirs. In contrast, Jane, who is also a physicist, seeks to explore disciplinary differences in approaches to computational data analysis. Similarly, there are differences between Brian and Bruno, both of whom have backgrounds in human geography. Even though both are interested in developing approaches that counter GIScience inspired geosocial research, Brian is more interested in understanding disciplinary differences in methods.

Finally, as their narratives about collaboration show, many interviewees develop exchange with scholars with complementary skills akin to 'trading zones' (cf. Galison, 2011). However, unlike 'simulation science' after World War II discussed by Galison (2011), geosocial research does not develop into a separate research field. Rather, the collaboration and epistemic landscape of geosocial research remain dispersed. I observe differences among the practices of the scholars Galison (2011) studied and those of my interviewees. The scholars Galison (2011) studied came together in a series of meetings to discuss the potentials of a new computational method - Monte Carlo simulations. In contrast, my interviewees share interests in the skill to conduct computational data analysis with geosocial data, rather than a specific computational data analysis method. In addition, my interviewees do not meet through meetings or shared research agendas. Rather, a shared set of computational skills allows interviewees to 'set up distinct laboratories', which further aid the development of diverse geosocial research approaches.

Secondly, in response to the First Research Question, I argued that experimentally searching for patterns helps interviewees develop their approach to geosocial research. Many interviewees emphasized their 'experimental' use of diverse data analysis methods to explore possible research questions and variables of interest, and thus identify patterns. This practice is similar to the exploratory data analysis Jebb, Parrigon, and

Woo (2017) discuss and supports Mackenzie and McNally's (2013) argument that the quest for patterns with digital data multiplies methods.

In response to the Second Research Question, which asks how geosocial approaches differ, this chapter argued that interviewees' preference for combining computational data analysis and socio-spatial interpretation in either sequential-modular or iterative fashion - which also signals difference about the way they create 'valid' findings - differentiates geosocial research approaches. I argued that several participants with technical background combined the two research traditions in a sequential fashion. Social science informed their research questions and their interpretations of data patterns. However, they identified and validated data patterns based on computational and statistical criteria alone. As section 2.4 discussed, I consider data analysis and interpretation inextricably linked. Thus, even when interviewees use statistical or computational criteria to establish valid findings, I do not argue that data analysis speaks for itself and does not require interpretation. Rather, I argue that they identify data patterns in terms of computational and statistical criteria, without actively considering social scientific knowledge at that stage of the research process.

In contrast, several participants with social science backgrounds stated that they iterate computational data analysis and socio-spatial interpretation, and participants from groups B, C and D claim they create and modify data analysis methods in light of social scientific theories. Their narratives suggest the criteria they use to validate patterns result from combining computational data analysis and social science. Similar to Marres and Gerlitz (2016), they perform or modify computational data analysis methods if their research questions and theoretical framework require it. While this section illustrated these differences in how interviewees conduct geosocial research, Chapter Six will provide further evidence that social and technical geosocial research can be understood as separate approaches.

In response to Research Question 3.1 - which asks how methods mixing help study the differentiation of and differences among geosocial research approaches - I argued that combining interviews and network analysis helped study collaboration (a relational practice among homogeneous actors (geosocial scholars) through which they develop geosocial research approaches) from complementary perspectives. Similar to relationalist social network analysis, through interviews, I studied *why* and *how* participants collaborate and co-authorship network analysis helped me study *the extent* to which geosocial scholars collaborate. Network analysis also helped study collaboration over time among many geosocial researchers. Combining methods shaped my argument. Both interviews and the co-authorship network's high modularity over time suggested that through small scale collaborations among scholars with complementary skills, scholars set up their own geosocial laboratories and develop diverse approaches to geosocial research in parallel. While network analysis did not alter my interview analysis, it shifted the narrative in subtle ways. The high modularity of the co-author network over time prompted me to reflect on the importance of the concurrent development of small collaborations or geosocial laboratories, and the lack of 'coherence' of the geosocial research 'field' further illustrated in subsequent chapters.

Finally, in response to Research Question 3.2 - which explores how to evaluate the affordances of computational methods 'for STS' - I argue that methods need to be evaluated in their interpretative context. Thus, I discuss how my use of a structural, homogeneous network analysis method (the modularity of the co-authorship network) - often critiqued by STS scholars for obscuring the dynamism of research practices and heterogeneous agencies which enable it - hinged on the interpretative context. Interviewees' narratives suggested that geosocial research mainly develops through numerous small scale collaborations. I also found that these are enabled by fewer collaborative acts among these cohesive groups by interviewees who collaborate with researchers with complementary skills who share loose theoretical or methodological common ground with them. Collaboration amid looser common ground helps scholars learn skills they later contribute to the collaborations with scholars with whom they share stronger common ground or vice versa. In other words, the interviews suggested that collaboration arrangements - and in particular, the relative size of collaborations with respect to all geosocial research and the number of collaboration relations among groups, captured by network modularity - provide information about the differentiation of geosocial research approaches. In this context, I could interpret changes in the modularity of the co-authorship network with respect to the development of geosocial research approaches. Changes in network modularity helped me study the extent to which geosocial scholars develop multiple small scale collaborations in parallel, or whether, to the contrary, geosocial research becomes an integrated, larger scale collaborative research practice.

I compared changes in the modularity of the co-author network at four time periods. STS highlights that the identity of actors and collectives that comprise research practices change over time. Cambrosio, Bourret, et al. (2014) note the danger of structural, temporal network analyses which assume that networks and nodes capture identical entities over time. I argue that the temporal analysis of the co-authorship network did not require me to assume the identity of the network or nodes over time. Comparing the modularity of the network to statistically comparable graphs at four time points helped study the extent to which large scale collaborations emerge in geosocial research, without assuming that scholars (the nodes of the network), the nature of collaboration among them (network edges), or geosocial research (whose scientometric traces the collection of nodes capture) remain identical over time. I assume that the co-authorship networks at each time point capture the scientometric traces of collections of geosocial research and researcher subjectivities that are different, but are shaped by earlier geosocial research.

Chapter Six will further explore how interviewees develop their distinct approaches to geosocial research by exploring how they find institutional homes in university departments or disciplinary communities.

Chapter 6

Making Academic Homes for Geosocial Research

Through exploring how interviewees make institutional homes for their geosocial research, this chapter explores all three research questions.

Addressing the First Research Question, it discusses three practices that help my interviewees develop approaches to geosocial research by creating institutional homes for themselves and their geosocial research. Sections 6.1 argues that interviewees imagine geosocial research in light of their institutions' research foci - mainly in disciplinary terms. In addition, section 6.2 argues that social scientists create homes by changing affiliations in light of a 'mis-fit' they experience with their (preferred) departments, and are actively differentiating their research from technical geosocial scholarship.

In response to the Second Research Question, through combining interview analysis and scientometrics, section 6.3 argues that social and technical geosocial research increasingly differ. Interviews and line graphs show the increase of social geosocial research, and citation network analysis shows a decrease of citation links among geosocial papers published in social scientific and computational journals over time.

The chapter's mixed methods approach helps address the Third Research Question. I combine two types of scientometric methods - structural network analysis and descriptive statistics - with interviews, to explore the differentiation of social and technical research approaches.

Supported by Chapter Five's findings, I argue that geosocial research cannot be understood as a coordinated research community that interviewees identify with. Rather, it is a collection of research approaches.

6.1 Imagining Geosocial Research along Disciplinary Lines

In order to make homes for their geosocial research, interviewees situate their geosocial research with respect to their institutions' perceived research agendas. I argue that this highlights how interviewees primarily relate their geosocial research to their respective (sub-)disciplinary communities rather than to other geosocial scholars, suggesting that geosocial research is not a coordinated research community. I identified narratives about two kinds of institutions that interviewees state welcome their geosocial research. Some participants claim their institutions seek to foster collaboration among scholars from different disciplinary backgrounds, and consider their geosocial research to result from the merging of diverse disciplinary knowledges. Others perceive their institutions to have strong disciplinary identities, claiming their geosocial research helps these institutions extend their computational research capabilities. However, in both cases, geosocial research is to a large extent advanced by relating and re-imagining the use of computational methods in light of interviewees' disciplinary heritage, thus developing geosocial research in diverse directions.

Brian, Chase, Colin, Jane and Anne state that they find homes for their geosocial research because their institutions are interdisciplinary: they foster collaboration among scholars from diverse disciplines. I argue that two factors - publication requirements and anchoring their contribution in light of their existing knowledge - facilitate developing their geosocial research along disciplinary lines in their institutions that they perceive as interdisciplinary.

Brian and Chase state that publishing in journals associated with their home discipline is key to their research at interdisciplinary institutions. Brian explained that as he collaborates with scholars from diverse disciplines, each of them lead publications in journals associated with their own disciplines due to research evaluation requirements. Thus, collaborating with his colleagues helps Brian further develop computational tools for his own discipline.

> *"Right now I'm the only geographer in my [...] unit which has humanities, arts, and social sciences. [...] but on a daily basis I work with people from architecture engineering, computer science, physics. [...] generally we still publish in our home kind of discipline because that's ultimately, you know, where your community is and also in terms of like career progression, how you will be judged [...] [usually] each person kind of takes the lead and pitching on in their kind of home discipline..."*

Like Brian, Chase publishes in disciplinary journals associated with his home discipline, which he stated is interested in his group's research. In addition, he attempts to publish in a journal associated with another, more computationally-oriented discipline which he seeks to align himself with through geosocial research.

> *"It looks like people from [my original] discipline really like [our geosocial research] [...] [But] I really would like to belong to [a more computationally oriented disciplinary] community, at least get into the periphery [...] But [scholars from the more computational discipline] have not reacted yet. We uploaded the paper to [a preprint portal], but it hasn't been tweet-ed yet, whereas they have tweeted previous papers."*

In sum, Brian and Chase stated that publishing in journals associated with their home discipline is key to their research at interdisciplinary institutions.

Next, I show that Colin, Chase, Jane and Anne frame their contributions along disciplinary lines. Colin discussed his role in interdisciplinary collaborations through 'anchoring' his approach in his disciplinary heritage of sociology. He believes this helps him to better interpret the results of data models than computational scholars. He also expects that in collaborations with computational scientists, he will have to adapt to computational scholars rather than the other way around.

> *"and probably [our relationships with IT people or computer scientists] will be an asymmetric relationship and I'll have to adapt to them. [...] I think our [sociologists'] role can be asking smart questions, and perhaps we interpret the results better as well..."*

Jane also stated she contributes to interdisciplinary collaborations based on her disciplinary heritage in physics and computational social science. As section 5.2.1 described, she believes that Chase's team hires her because of her ability to contribute computational data analysis skills.

Finally, Anne explained the difficulties she encountered when her disciplinary identity was questioned as she entered into an interdisciplinary PhD program. Many of her colleagues at her multidisciplinary department worked with approaches more aligned with computational social science. She became uncertain about her work when her disciplinary identity was questioned, and tried to imagine her contribution along disciplinary lines:

> *"I came in [to the PhD program] with a clear idea, you know, that we need to push against [computational social science], but, you know, as [...] more and more people are going 'well I'm using code and JSON files and Python and algorithms', you start to feel a bit kind of in the shade of all this big, spiffy technology. Yeah, I did lose my way a bit..."*

In sum, the above interviewees who consider their institutions interdisciplinary they anchor their research in their disciplinary heritage.

In contrast to the above interviewees, David, Elias, Gary and Isaac stated that they are making their homes in 'disciplinary' departments that seek to expand their computational research capabilities. Like the interviewees quoted above, they frame their geosocial research in disciplinary terms.

Gary credited his employment to his department's wish to extend their computational data analysis capabilities.

> *"It's a geography department, and they hired me partly because they know that I can work with social media data. My research focus will be broader than that though..."*

As sections 4.3 and 4.1 discussed, David stated that his research with geosocial data allowed him to re-negotiate his professional identity and change institutional affiliation. Like Gary, he credited his employment to his department seeking faculty members with combined knowledge of interpretive social sciences and computational data analysis, who could expand the department's computational data analysis capabilities.

> *"My new position, that [social science] department hired me because they want to expand their capacity to work with digital data and bring in computational approaches."*

Elias also credited his job to his disciplinary department's desire to expand its computational research. He believes that his combined expertise in computational data analysis and the department's disciplinary focus helps him obtain the job.

> *"these skills [computational analysis] are quite hot and universities are looking for people who can teach it. So now I've got a new faculty position in the [names country] because I do this kind of stuff."*

As section 5.2.1 discussed, Isaac likewise credited his employment to his computational skills and his disciplinary department's desired methodological expansion to study social phenomena that are challenging to quantify.

In sum, this section argued that interviewees develop their geosocial research in light of their disciplinary research heritages. Next, I discuss social scientist interviewees' efforts to create academic homes for their geosocial research.

6.2 Making Homes for Social Scientific Geosocial Research

This section discusses two additional practices interviewees with social science background conduct to make home for their geosocial research: as section 6.2.1 discusses, they change their affiliations because of a mis-fit between their geosocial research and home institutions. As section 6.2.2 discusses, they actively differentiate their research from technical geosocial scholarship. Finally, section 6.2.3 reflects on social scientists' experiences about these moments of transition in light of scientometric analyses which show that the proportion of geosocial research with respect to social science grows at a quicker rate than the proportion of geosocial research with respect to computational sciences. Altogether, my interviewees' quotes and the scientometric analyses suggest that social scientists are currently, actively developing geosocial research approaches in dialogue with their sub-disciplinary communities and through coping with institutional constraints.

6.2.1 Changing Affiliations

This section discusses how most interviewees with social science background felt a mismatch between their geosocial research and departments, and made academic homes for themselves by changing their affiliation. Seven interviewees with social science background from seven research groups (A - E, G, H) felt a discrepancy between their geosocial research and the epistemological or methodological norms of their 'home institutions' - university departments or disciplinary research communities. In contrast, most interviewees with backgrounds in the computational sciences felt their geosocial research fit well with their home institutions, and successfully completed and published their geosocial research with institutional support. Reflecting on epistemological and methodological differences in light of institutional constraints is itself a social scientific research topic. I thus cannot make the claim that this practice is unique to social scientists who may just be more articulate about these factors. However, articulating these aspects of the work means enacting this reality, and my interviewees' narratives highlight struggles to find home for their geosocial research, and fit computational data analysis with their institutions' norms. The need for social scientists to change their affiliation to find home for their geosocial research suggests that they are renegotiating computational methods associated with geosocial research with respect to existing norms in their disciplines.

As the quotes below show, social scientist participants re-negotiated their professional identities by changing their affiliations with varying levels of willingness and agency for three main reasons: the misfit between institutions' disciplinary boundaries and geosocial research, the will to re-position themselves or their institutions' willful rejection of geosocial research.

Firstly, three interviewees - Henry, Brian and David - stated that they changed their affiliation because they felt that the disciplinary norms of their prior institutions could not accommodate their geosocial research. Henry chose an interdisciplinary institution where he can supervise students with both social scientific and technical backgrounds. This suits his interest and expertise in combining computational and urban research. However, he stated that colleagues' expectations are often based on disciplinary research lines, which is at odds with his mixed background, which he finds challenging.

> *"I don't see myself as a pure social scientist anymore and I don't mean it [...] in a good and a bad way. [...] I know more technical stuff [...]*

> *But [...] I can't keep up on all the latest theoretical developments [...] And that's [...] hard... [... because] you're sort of expected to know that just because your PhD says [a social scientific discipline] even though it was 10 years ago and you spent those 10 years writing code and talking to computer scientists..."*

Like Henry, Brian feels that disciplinary divisions cannot accommodate his geosocial research. Thus, he sought employment at an institution where - in contrast with his previous home institution - disciplinary divisions do not map onto departmental units.

> *"I sought employment at [names institution] specifically to join a university that was new and didn't have traditional departments. [...This is] important especially in this subfield [where] we have to come together around some of these data sets. [...] But it's not an easy process...."*

Finally, David also feels that existing disciplinary boundaries do not readily accommodate his geosocial research. As his quote in section 4.1 illustrated, he feels that 'technical' tasks associated with data management and data analysis that are essential for his research are not valued by his social scientific peers. Identifying with digital sociology, a new research area, helps him create an academic home for himself in a community he feels values both computational work and social research.

> *"I had a lot of misgivings [...] for a long time [...] about the term digital sociology, because I felt like 'do we really need yet another specialisation?' [...] but more recently I started using that label also to refer to what I'm doing, just because I feel like there does need to be a space [where] experimentation and playfulness even [laughing] is valorised and rewarded [...] It would be nice to have a little bit of community that values [technical tasks]."*

In sum, Henry, Brian and David re-negotiate their institutional affiliation because they feel that institutional divisions based on existing disciplines cannot accommodate their geosocial research.

Secondly, as section 4.3 discussed, David and Colin used geosocial research to willfully change their disciplinary affiliation - either to integrate their academic and non-academic research, or out of the belief that geosocial research can help them obtain jobs. Like them, Chase uses geosocial research to alter his institutional affiliation. Like David, he explains that he tries to affiliate himself with disciplines with a stronger focus on computational methods.

> *"I really would like to belong to the [computational discipline] community, at least get into the periphery, and link it to [my original discipline]. [...] [For now] I'm much more embedded in [the community of my original discipline]..."*

I sum, David, Chase and Colin wilfully change their affiliation and seek departments that help them pursue their computational research interests.

Thirdly, three interviewees - Anne, Bruno and Brian - stated they were prompted to change institutional affiliations because colleagues in their original home institutions reacted negatively to their geosocial research. All three felt their research was alien to the geography departments with which they were originally affiliated. Anne, who already had dual institutional affiliations, stopped attending geography meetings and physically moved office to another supervisor's department.

> *"In the end [the conflict] affected me so much that I stopped going to geography departmental meetings. A lot of the questions I would get were things like 'what does this have to do with geography?' and 'Why are you based in this department?' [...] [Eventually] I changed department."*

Similarly, Bruno explained that his 'home' geography department was particularly hostile toward his geosocial data research. He believes the department allowed him to defend his PhD thesis primarily because of the credibility granted by his extant publications that used geosocial data.

> *"[my committee] told me to my face that this was not real geography [...] I think had I not been so far ahead [with publications], and had a job lined up already, they would have been happy to make me completely rewrite my dissertation or to have failed me or whatever else."*

Brian travelled internationally to another university to pursue his research interests with geosocial data. He explained that the geography department with which he was affiliated when he started his PhD studies was not interested in geosocial research.

> *"At the time [when I started my PhD] in [my previous university] [...] I don't think people really saw the value of this type of data. There was a lot of kind of thinking that the digital or virtual and 'real' world were two separate things."*

In sum, Anne, Bruno and Brian felt the geography departments with which they were affiliated excluded their geosocial research, prompting them to change affiliations.

Altogether, this section showed that the majority of interviewees with social science backgrounds experienced an uneasy fit between their affiliations and geosocial research, and proactively sought new academic homes for their geosocial data research by changing their affiliation. Unlike the social scientists quoted above, most computational researchers did not express a mismatch between geosocial research and their home institutions. Instead, as section 5.2.1 argued, those who changed their affiliations - like Jane and Isaac - did so as a result of the collaborations they developed with social researchers as part of their geosocial research. This suggests that social scientists are currently, actively re-configure the relationship between their disciplinary heritage and geosocial research, and thus develop geosocial research in new directions.

6.2.2 Differentiating Social Geosocial Research

Social scientists from five groups (A - E) narrated their geosocial research in terms of its differences from technical geosocial research, such as 'computational social science' and 'GIScience'. As this section discusses, four interviewees - Bruno, Brian, David and Colin - demarcated their geosocial research from computational social science or GIScience in terms of the richer interpretive capability afforded by their knowledge beyond computational modelling, such as local knowledge and conceptual frameworks in interpretive social science. As section 7.1 will discuss, six interviewees with social science background demarcated their work in terms of their reflexivity. Chapter Seven will discuss interviewees' demarcation in terms of reflexivity because it will argue that a key difference between approaches to geosocial research is the type of reflexivity associated with them. In contrast, narratives about differentiating geosocial research were not prominent in my discussions with interviewees with technical background. This suggests that some social scientists are actively developing geosocial research agendas that they consider different from technical approaches. Next, I illustrate how Bruno, Brian, David and Colin differentiate their social geosocial research from technical approaches.

As section 5.2.2 discussed, Brian and Bruno stated that their geosocial research seeks to challenge technical geosocial research mainly informed by computer science and GIScience. Bruno claimed that their work differs from technical approaches because in contrast with the the latter, their research is informed by social scientific literature and local knowledge.

> *"I think that one of the things that we've always felt made our work different [from GIScience] is that we were much more embedded in the places we research and that there was a lot more like the non-[social media] data stuff that went into the projects. [...] We we've shied away from doing a project about some random city."*

Similarly, David demarcated his and Daniel's research from computational social science in terms of their familiarity with the place they study, as well as research frameworks they draw on beyond computational analysis. He argued that computational social scientists would have had difficulties analysing a subset of data pertaining to a neighbourhood that did not yield evident (statistical) patterns, which they were able to analyse due to their disciplinary background and familiarity with the neighbourhood.

> *"I think it's illustrative of the differences between [our] approach [and computational social science], because I think many people who have a different background would have great difficulties making sense of something like the [data pertaining to a specific place that does not afford an apparent pattern], and we're completely comfortable analysing it because of our [background in a specific sociological tradition of thought] and our familiarity with the neighbourhood."*

Finally, Colin claimed that social scientists can provide more detailed interpretation than scholars with computational training and help formulate research questions, drawing on social scientific theories.

> *"I think the role of the sociologist could be introducing these worries into the research process, and come up with ideas and hypotheses that a data-driven engineer or data scientists cannot come up with. And [...] do more in depth interpretations. Because for these you need these doubting, bit ideologically inspired social science mindset."*

In sum, interviewees with social scientific backgrounds from groups A-E delineated their work from computational social science through emphasizing their reliance on knowledge beyond computational data analysis, such as local knowledge and social science knowledge, and as section 7.1 will show, in terms of their reflexivity. Next, I reflect on participants' efforts to differentiate their research from technical approaches in light of scientometrics.

6.2.3 Rise of Social Geosocial Research

Through combining scientometrics with the interviewee analysis discussed in sections 6.2.1 - 6.2.2 and below, I argue that through social scientists' efforts to differentiate their approaches, social geosocial research is currently on the rise. Firstly, I discuss interview quotes which suggest that although social scientific geosocial research is growing, it is not widely recognised yet. Secondly, I discuss line graphs which suggest that the proportion of social geosocial research published by sub-set of social scientific journals increases at a faster rate than technical geosocial research. Finally, I discuss interviews which further illustrate the divides within social science between journals which do and do not publish geosocial research.

As shown below, Anne and Ben feel that although social scientific geosocial research is gaining momentum, it is still not widely recognised by colleagues in their broader disciplinary communities. Their experiences suggest that change is currently happening: disciplinary sub-communities publish more and more social geosocial research but it is not (yet) widely known. Ben claimed that based on the feedback he gets from scholars at events, geosocial research still feels like a niche even as he hopes it should be more widely accepted and feels that it is growing.

> *"I guess it's starting to feel like it's not sort of as fringe as it once was. [...] but I'm sort of surprised that there aren't more people within geography or the social sciences more broadly [who do this type of research] [...] sometimes I feel like I've known a secret for a while that shouldn't be a secret. I'll give a talk and things I assume have been sort of settled ground for like five years, people come up afterwards say 'oh well that was really interesting', like 'oh, I had no idea that it was possible to do something like this with social media'. I'm like 'Really? We've been talking about this forever..."*

Like Ben, Anne believes that there is a relative paucity of research that uses digital data similar to her research, which is informed by social scientific theories and methods, in contrast to technical scholarship which aims to uncover statistical patterns in large datasets. Colleagues she met at conferences and at her home department associated her geosocial research with computer science, far afield from her social scientific home disciplines.

> *"There is a cohort of other people doing similar kinds of stuff [to me]. But my perception [...] is that the mentality of Big Data with capital B and D is still very pervasive. One of the main questions I would get if I briefly explained my research would be 'oh, are you a computer scientist?' The association was, if you're working with big data, you must come from a computer science or a mathematics background. And when I say like 'oh no, I'm an anthropologist', it just kind of goes blank."*

The scientometric analyses presented in figures 6.1 - 6.3 depict the proportion of geosocial papers with respect to three paper sets. As I discuss below, they suggest that social geosocial research is gaining popularity in certain social scientific circles at a faster rate than technical geosocial research in technical (sub)-disciplines. This provides further evidence that social scientists are actively developing their approach to geosocial research.

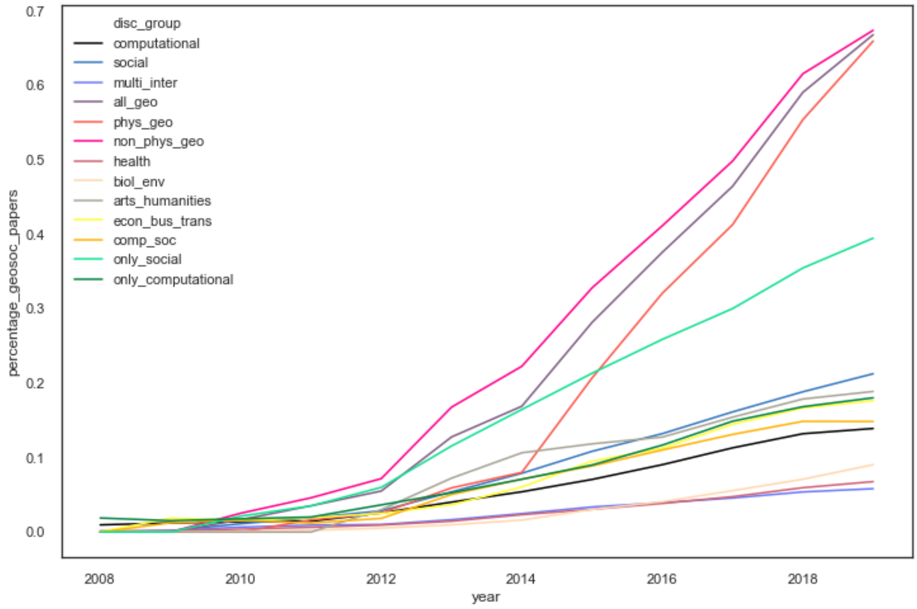


Figure 6.1: Cumulative yearly percentage of geosocial papers per disciplinary categories compared to the total number of published in the same journals between 2008 - 2019

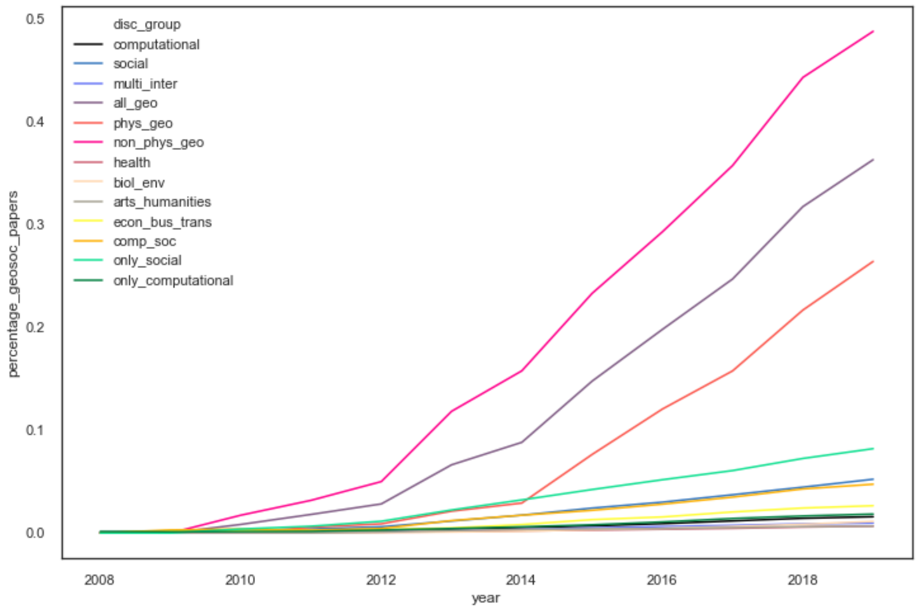


Figure 6.2: Cumulative yearly percentage of geosocial publications per disciplinary categories compared to the total number of published in all journals associated with the same Web of Science Subject Categories between 2008-2019

Figure 6.1, shows the cumulative percentage of geosocial papers with respect to the total number of papers published in the journals which publish geosocial papers over time, by Broad Disciplinary Categories. It shows that the proportion of geosocial papers is the highest in the subset of human geography ('non phys geo') journals which publish geosocial research, surpassing the percentage of geosocial papers in the subset of physical geography ('phys geo') journals which publish geosocial research. The cumulative percentage of geosocial papers is the fourth largest for the subset of 'only social' journals which publish geosocial research, and the increase of this percentage is considerably steeper than for the subset of 'only computational' journals.

Figure 6.2 shows the percentage of geosocial papers with respect to the sum of papers published in all journals listed in Web of Science whose Subject Categories match the string search outlined in Table 3.3 for each Broad Disciplinary Category, regardless of whether they publish geosocial papers or not (for more detail, see section 3.4.2). Like figure 6.1, it shows that the proportion of geosocial papers is highest, and grows fastest in human geography ('non phys geo'), followed by all geography ('all geo'), physical geography ('phys geo') and 'only social' journals.

The difference between the proportion of geosocial papers in human geography and physical geography journals is more pronounced than in figure 6.1. This suggests that while certain physical geography journals increasingly specialise in publishing geosocial research, geosocial research becomes most popular across human geography journals. In other words, my scientometric analyses suggest that human geographers are increasingly interested in geosocial research.

At the same time, the difference between the proportion of geosocial papers in 'only social' and 'only computational' journals in figure 6.2 is smaller than in figure 6.1. This suggests, that although 'only social' geosocial research does grow faster than 'only computational' geosocial research, the trend for certain 'only social' journals to specialise in publishing geosocial research is more pronounced. Elias', David's and Daniel's experiences support this, who highlight the difference between social science journals' likelihood for accepting their geosocial papers. Elias stated that the methods associated with his geosocial research are less accepted by social science journals associated with the social scientific area of his PhD research compared to other social science journals.

"The hardest thing for me these days is kind of figuring out what do

> *I put in, what is just background analysis, and what is stuff I've just got to save for another paper [... and] where do I go with with a particular paper for which audience. [...] because it's a new method, every time I would like to write a paper for a [discipline A] journal, I find myself wasting a thousand of my eight thousand words just trying to explain what I actually did methodologically. [...] It becomes a bit of a problem in terms of making the argument [...] Maybe I shouldn't send things to [discipline A], but that's what I did my PhD in. [...] in [discipline B] it's a more accepted research method."*

Like Elias, David and Daniel noted a difference in their methods' fit with different social science journals' norms. They stated that they changed their choice of publication venues to find social science journals whose methodological norms their geosocial research aligned with.

> *"David: what is now in the geography journal, we initially wanted to go to a general sociology journal, but, uhhmm [...] I think we had the realisation that the kind of urban studies angle of it was getting quite strong [...]*
>
> *Daniel: yeah, I think maybe even more generally the type of social science that we do, I don't think finds its way easily into sociology journals because they tend to be somewhat more conservative, they expect you to have sort of research questions, hypotheses, operationalisation and so on.*
>
> *David: and we would kind of have to invent those after the fact, because we do have this iterative process..."*

In sum, the scientometric and interview analyses presented above suggest that social scientific geosocial research is currently growing, but is not yet widely known.

Finally, figure 6.3 depicts the yearly cumulative percentage of geosocial papers in each Broad Disciplinary Category between 2008 and 2019, with respect to all geosocial papers in the same period. When compared to the analyses presented above, this figure highlights two aspects of geosocial research. Firstly, it suggests a discrepancy between my interview and scientometric fields. Even though my scientometric field contains a relatively high proportion of health, biology and ecology related geosocial research -

depicted by the lines corresponding to the Broad Disciplinary Categories 'health' and 'biol env' - none of my interviewees focused on these topics.

Secondly, the interview analyses and line graphs suggest that belonging to (sub)-disciplinary communities which welcome their (computational) geosocial research is key for social scientists to develop their distinct geosocial research approaches. Figures 6.2 and 6.1 - which show the pronounced increase in social geosocial research since about 2012 within the social sciences - resonated with my interviewees' narratives about the current growth, but lack of widespread recognition of social scientific geosocial research. In contrast, figure 6.3 - which suggests that the proportion of social geosocial research has been constant since 2010 and comparable to the proportion of computational geosocial research - does not resonate with interviewees quotes. I argue that my interviewees experience the recent growth but lack of widespread recognition of social geosocial research with respect to their experience of the social scientific field, not with respect to the geosocial field. This suggests that the latter is not a coordinated research community that interviewees belong to.

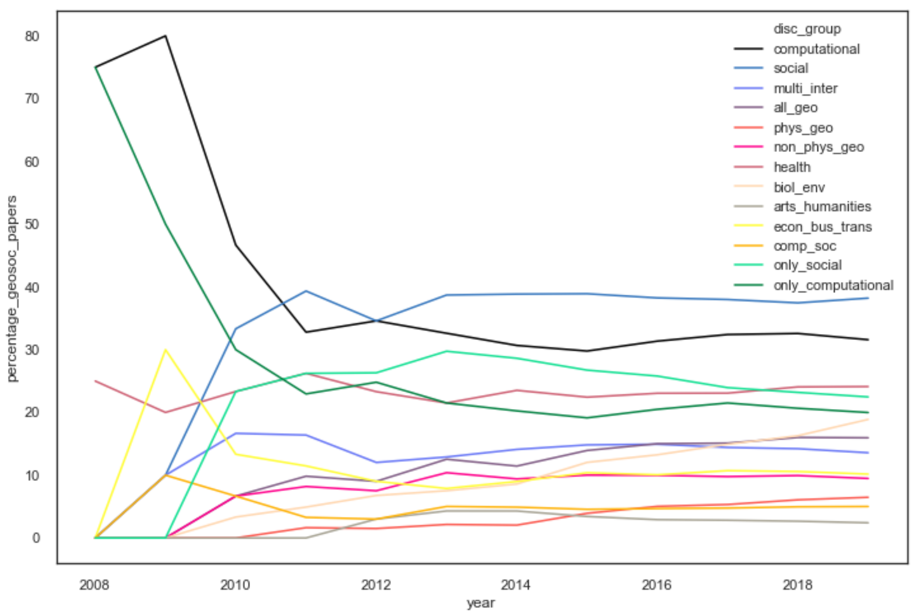


Figure 6.3: Cumulative yearly percentage of geosocial publications per subject categories with respect to the total number of geosocial papers

Altogether, through combining interviews and scientometrics, this section argued that social scientists are in the process of developing approaches to geosocial research which differ from technical geosocial research. The next section will trace the relationship between social and technical geosocial research over time and present scientometric evidence which shows that they increasingly differ.

6.3 Changing Relations between Social and Computational Geosocial Research

This section presents scientometric analyses which support the above finding that social scientists are in the process of differentiating their geosocial research approaches, and show that social and technical geosocial research increasingly diverge over time. Informed by findings in Chapters Four through Six, the analyses explore whether the (changing) connections between 'only social' and 'only computational' geosocial scholarship help the development of distinct geosocial research approaches. Chapters Four and Five argued that relations between interviewees with social and technical backgrounds are essential for the development of distinct geosocial research approaches. Chapter Four illustrated the role of interviewees' shared interest in combining computational data analysis and social research, supported by overlapping aesthetic sensibilities. Chapter Five illustrated the importance of collaboration (especially based on loose common ground) between interviewees with social and technical disciplinary backgrounds. In addition, Chapter Five and section 6.2.2 argued that social scientists are currently, actively differentiating their geosocial research from technical approaches.

Based on the above findings, I hypothesise that the links between social and technical geosocial research help differentiate them. As section 3.4.4 explained, I scientometrically study this by exploring whether edges between geosocial papers published in 'only social' and 'only technical' journals impact the structure of author-bibliographic network of these papers (over time). I study the author-bibliographic coupling network of papers published in journals in the 'only social' and 'only computational' Broad Disciplinary Categories because as section 6.2.3 argued, this subset of my scientometric field better aligns with my interview field. The scientometric analyses help study the relationship between social and technical geosocial research on larger scale than interviews and participant observation: including the research of a broader set of scholars over a longer time period.

I argue that the author-bibliographic coupling network analysis can help explore how the connections between social and technical geosocial research help differentiate approaches - first hypothesised through interview analysis - by making the following three assumptions. Firstly, I assume that author-bibliographic coupling - where papers are linked to the extent that they cite literature authored by the same scholars - provides a (partial) view on the above diverse forms of associations among scholars who conduct social and technical geosocial research. In other words, I assume that shared methodological and aesthetic interests, as well as knowledge exchanged through collaboration partially manifest in author-bibliographic coupling.

Secondly, I assume that geosocial papers published in the non-overlapping 'only social' and 'only technical' journals partially capture the scientometric footprints of the social and technical geosocial scholarship (that I argued interviewees actively differentiate) 'well enough' for analytical purposes. As sections 6.1 and 6.2 argued, interviewees (whose approaches to geosocial research differs) tend to publish in journals associated with their disciplines for research evaluation purposes and to express their belonging to associated research communities. In addition, as Chapter Five and section 6.2 argued, social scientists from diverse disciplines attempt to differentiate their geosocial research from technical approaches. Thus, I assume that the collection of geosocial papers published in 'only social' journals capture the research output of these social scientific scholars, and that geosocial papers 'only technical' journals capture the research output of technical geosocial researchers 'well enough'.

Finally, I assume that the modularity of the author-bibliographic coupling network helps study differences in geosocial research (approaches). This assumes that author-bibliographic coupling relations among papers - citing the same author - indicate similarity of their geosocial research approaches. Network clustering in Chapter Seven will illustrate the validity of this assumption. It will show that interview findings about differences in social, technical and geographic geosocial research approaches are mirrored in citation network clusters.

Figure 6.4 depicts the author-bibliographic coupling network *G1* discussed in section 3.4.4 which calculates author-bibliographic relations using cosine-similarity. The nodes of the network are geosocial papers aggregated on the journal level. The network's weighted edges correspond to the cosine similarity values which depict author-bibliographic coupling between the geosocial papers published in the journals which correspond to the nodes. As section 3.4.4 discussed, the journals were categorised into three Broad Disciplinary Categories: figure 6.4 depicts 'only computational' journals in green, 'only social' journals in orange and journals which do not fall into these categories in purple. The figure illustrates the relative separation between the 'only computational' and 'only social' journals, meaning that the geosocial papers they publish rarely cite the same authors.

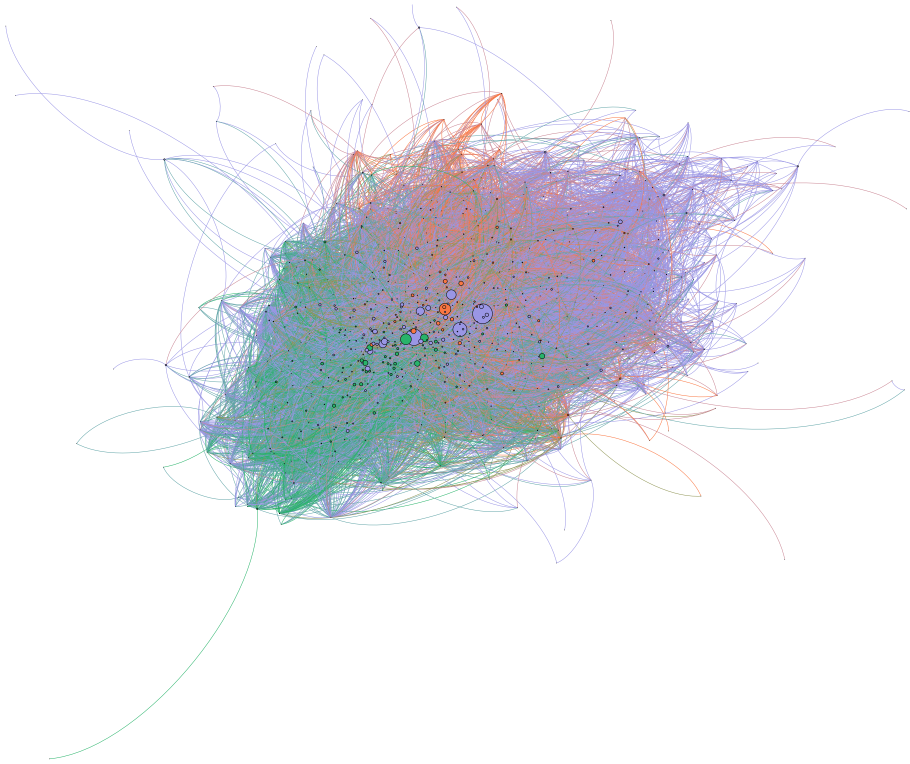


Figure 6.4: Journals author-coupling categorised by disciplines. Green: 'technical', orange: 'social science', purple: other

To explore whether edges (jointly cited authors) between geosocial papers published in 'only social' and 'only computational' journals (*inter-edges*, for short) shape the network's structure, figures 6.5 and 6.6 depict the modularity of the network *SG1* - a sub-graph of *G1* that omits the inter-edges - compared to the modularity of 1000 simulated random sub-graphs with equal number of randomly deleted edges (for details, see section 3.4.4). The red line indicates the modularity of *SG1*, and the histograms (green and blue, respectively) depict the distribution of the simulated graphs' modularities. As section 3.4.4 discussed, in addition to the number of edges, figure 6.5 also controls for the distribution of edge weights of the 1000 simulated random sub-graphs, so that the distribution of the edge weight of the randomly omitted edges follows the distribution of the edge weights of the inter-edges.

As figure 6.5 shows, it is highly unlikely for a sub-graph of *G1* with number of edges and edge weight distribution equal to that of *SG1* to have the modularity of *SG1*. The modularity of *SG1* is larger than those of the simulated graphs. The analysis shows that deleting the edges between geosocial papers published in 'only social' and 'only computational' journals - the authors that papers published in 'only social' and 'only computational' journals jointly cite - render the graph more modular, or in other words, less 'cohesive'. This suggests that the inter-edges have a central importance. Without them, the network of journals which publish geosocial papers would be significantly less connected to each other. As figure 6.6 shows, this result holds when the simulation does not control for the edge weight distribution of *SG1*.

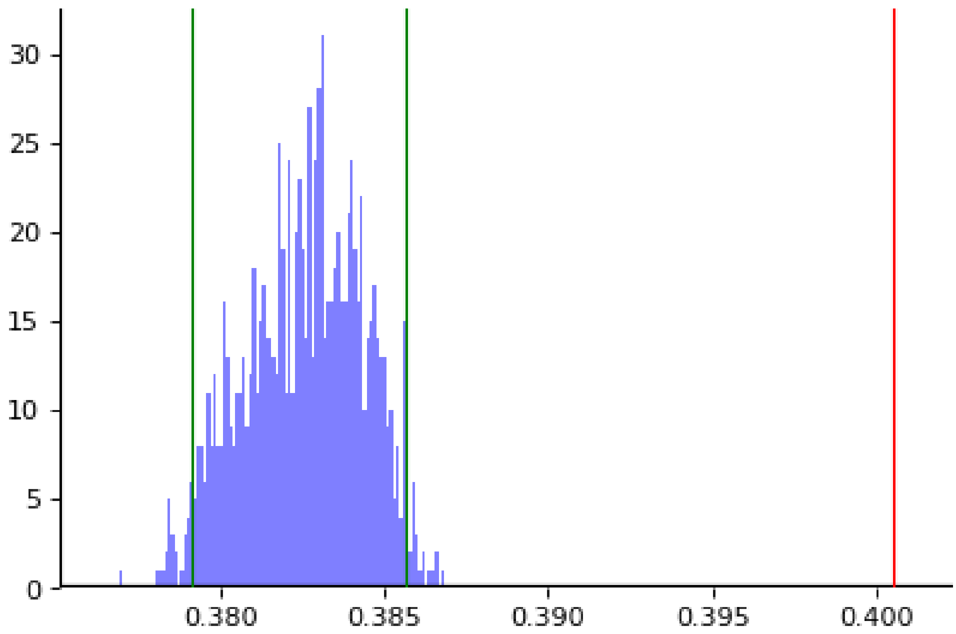


Figure 6.5: Modularity of the sub-graph $SG1$ (red line) compared to the modularity of 1000 randomly simulated sub-graphs with equal number of edges and equal distribution of edge weight, 95% of the data falling between the vertical green lines

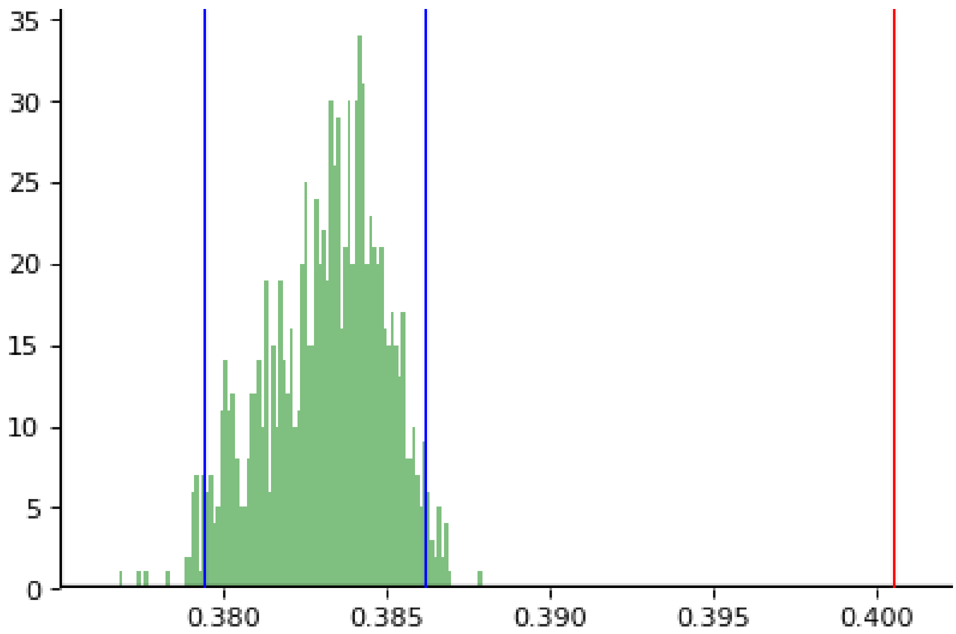


Figure 6.6: Modularity of the sub-graph $SG1$ (red line) compared to the modularity of 1000 randomly simulated sub-graphs with equal number of edges, 95% of data falling between the vertical blue lines

To study the *changing* relationship between geosocial papers published in 'only social' and 'only computational' journals, figures 6.7 and 6.8 depict the modularity of the graph *SG1* in red compared to the 95% upper and lower bounds of the modularity of 1000 simulated random sub-graphs with equal number of randomly deleted edges (depicted by the blue and green regions respectively) *over time*. Like above, figure 6.7 depicts network analysis which controls for edge weight, while figure 6.8 depicts analysis which does not control for edge weight. As the black dash line shows, the proportion of inter-edges (the number of authors cited both by papers published in 'only social' and 'only computational' journals) decreases over time. As the figures show, the modularity of the simulated graphs significantly differ from the modularity of *SG1* from 2016 onward. This shows that from 2016 onward, the author bibliographic coupling links between geosocial papers published in 'only social' and 'only computational' journals render the network less modular, or in other words, more connected. As figures 6.9 and 6.10 show, these patterns hold for the network *G2* and its sub-graph *SG2*, which, as section 3.4.4 explained, calculates author-bibliographic coupling using the normalisation method outlined by Waltman, Boyack, et al. (2020). However, in this analysis, the two approaches differentiate at a later time point, in 2018. I found that the finding also holds four Discipline Categorisation Methods, two normalisation methods and two edge weight decimal rounding methods discussed in section 3.4.4, omitted due to space limitations.

In line with the analysis presented in section 6.2, the above analyses suggests that scholars who publish in 'only social' and 'only computational' journals increasingly develop geosocial research in distinct directions. In line with Chapters Four and Five, they show that a diminishing shared literature base (jointly cited authors) between 'only social' and 'only computational' geosocial papers plays an increasingly important role in connecting disparate geosocial research approaches. This increasing separation could result from scientists' efforts to demarcate social geosocial research from technical approaches, and the rapid growth of social scientific approaches discussed in section 6.2. Thus, I argue that social and technical geosocial research are distinct approaches to geosocial research. Future work could 'qualitatively' explore which authors social and technical geosocial research jointly cite and changes over time, in addition to counting changes in the shared author citations over time. Studying the scholarship of jointly cited authors could provide hints about (changes in) the nature of common ground between social and technical geosocial research.

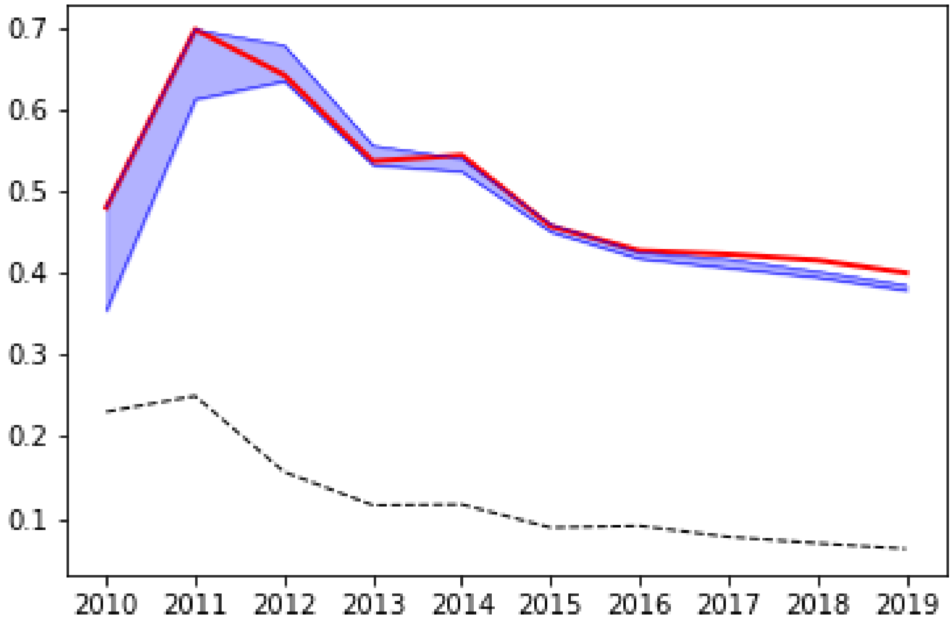


Figure 6.7: Yearly changes in the modularity of network *SG1* (normalised with cosine similarity) with respect to the 95% 'confidence interval' of the modularities of random networks with equal number of edges and comparable edge weight distribution

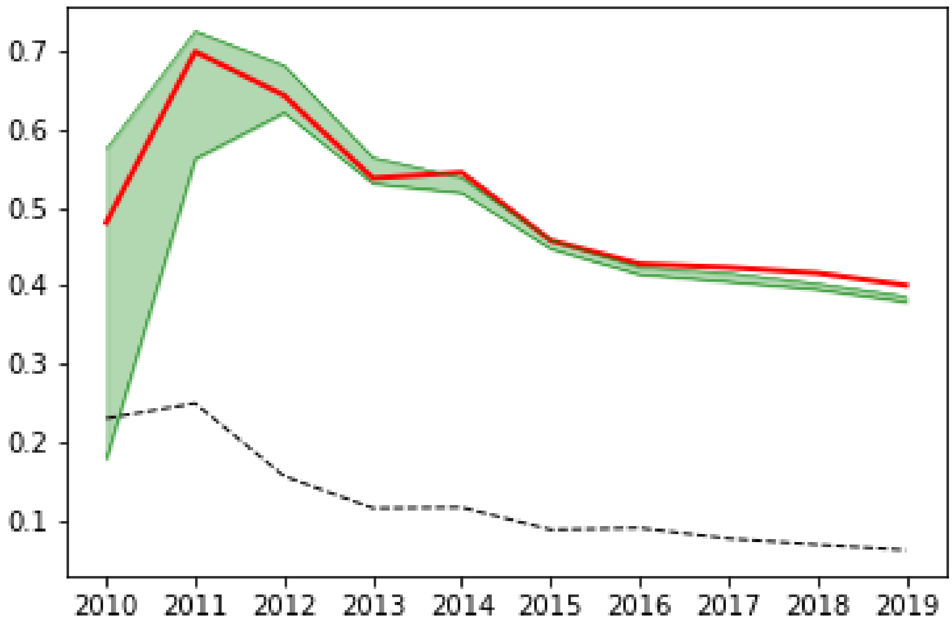


Figure 6.8: Yearly changes in the modularity of network *SG1* (normalised with cosine similarity) with respect to the 95% 'confidence interval' of the modularities of random networks with equal number of edges, not controlled for edge weight distribution

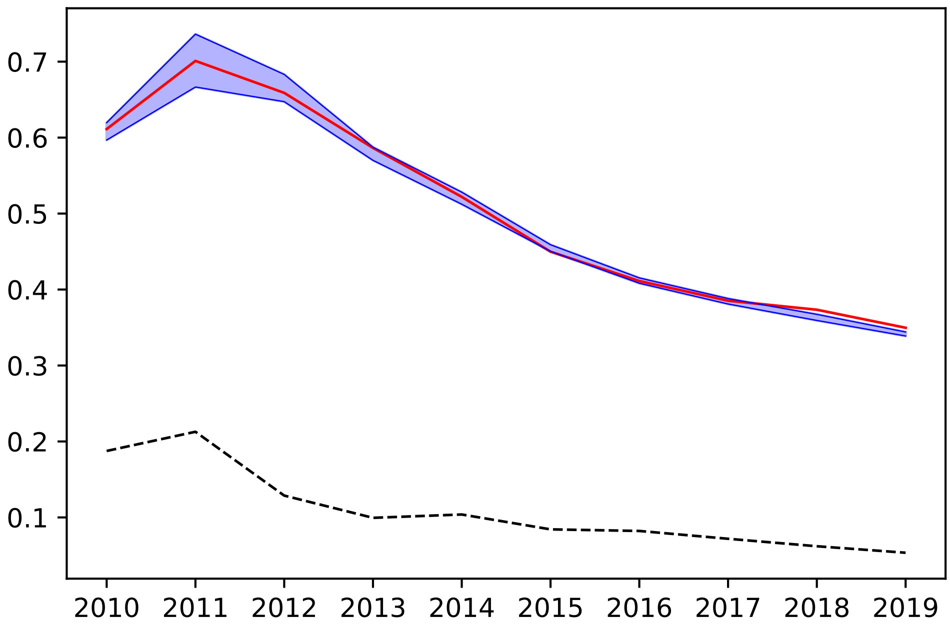


Figure 6.9: Yearly changes in the modularity of network *SG2* (normalised with the method outlined by Waltman, Boyack, et al. (2020)) with respect to the 95% 'confidence interval' of the modularities of random networks with equal number of edges and comparable edge weight distribution

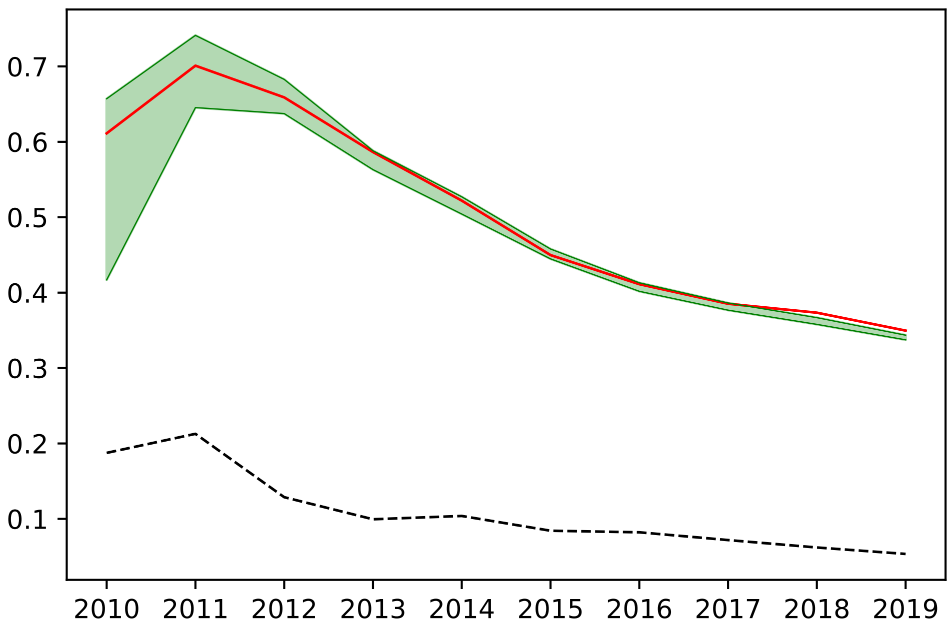


Figure 6.10: Yearly changes in the modularity of network *SG2* (normalised with the method outlined by Waltman, Boyack, et al. (2020)) with respect to the 95% 'confidence interval' of the modularities of random networks with equal number of edges, not controlled for edge weight distribution

6.4 Conclusion: Home-Making for Geosocial Research

Using mixed methods, this chapter explored interviewees' practices which help them find institutional homes and thus develop their geosocial research approaches, and argued that social and technical geosocial research increasingly differ. Similar to Johnston's (2003) findings about British geography almost two decades ago, I argued that geosocial research is a collection of research developed by scholars who belong to disciplinary or sub-disciplinary research communities, rather than a coordinated research community. Next, I discuss this chapter's findings with respect to each research question.

In response to the First Research Question, this chapter explored how interviewees find institutional homes for their geosocial research which, in return, shapes the way they develop their geosocial research approaches. It highlighted three such practices: reflecting on geosocial research in light of their institutions' perceived research foci - mainly in disciplinary terms - as well as social scientists changing affiliations and actively differentiating their research from technical geosocial scholarship. Next I discuss each of these practices in detail.

Osborne (2013) argues that research traditions meet and mingle predominantly as each advances on their own terms. For example, they often meet as scholars develop methods along the lines of existing research traditions. In line with him, section 6.1 argued that participants ground and develop their geosocial research in line with their respective disciplinary traditions both in institutions they perceive disciplinary and in those they perceive interdisciplinary. They actively re-imagine what computational methods could look like in light of their disciplinary backgrounds, thus developing geosocial research in diverse directions. I identified two main factors that facilitate participants' development of geosocial research along disciplinary lines in institutions they perceive as interdisciplinary. On the one hand, they seek to publish in journals associated with their disciplines to meet research evaluation requirements and to become members of disciplinary communities. Thus, whilst Osborne (2013) emphasizes epistemic causes for research to proceed along disciplinary lines, in line with Weszkalnys and Barry (2013), my participants' experiences also highlight the role of constraints imposed by research evaluation. In addition, they anchor their contribution to interdisciplinary teamwork in light of their existing, disciplinary knowledge. For example, similar to the interviewees of Balaban (2018) Anne became uncertain about her work when her disciplinary identity was questioned, and tried to imagine her contribution along disciplinary lines.

Whilst my analysis supports Osborne's 2013 observations that the mingling of research traditions comprise the development of research along disciplinary lines, my analysis highlights the labour required to make this possible for an emerging or new research field. As Chapter Four argued, many interviewees with social science background learn computational data analysis skills as they need to balance academic and non-academic jobs, due to financial pressures. As Chapter Five argued, several interviewees collaborate with scholars with complementary skills amid epistemological and ethical differences that they need to negotiate, to learn the skills required for geosocial research. In addition, social scientists may seek to alter computational data analysis methods to help capture units of analyses of interest to them. Finally, as section 6.2 of this chapter argued, scholars may need to change departmental affiliations or cope with experiences of forming part of a minority within their disciplinary research communities.

Section 6.2 argued that most interviewees with social science background expressed a misfit between their institutions' research focus and their geosocial research and changed their affiliations for three main reasons. Firstly, some felt that their geosocial research and associated skill set did not fit well with disciplinary divisions dominant in their home institutions. Secondly, others willfully used geosocial research as a means to change their research profiles and affiliations in the hope for better funding opportunities or to better align their research portfolio with their interests. Thirdly, three participants were prompted to change their affiliations because their home institutions did not wish to accommodate their geosocial research.

In addition, many of them actively differentiate their scholarship from technical geosocial research. For example, they highlight epistemic virtues associated with their geosocial research, such as their reflexivity and use of knowledge beyond computational data analysis, including local knowledge about the spaces they study and social theory. As section 2.4 discussed, STS scholars have demonstrated that the introduction of new technologies often brings about the re-negotiation of professional identities and epistemic authority. Similar to the radiologists in Burri's (2008) study, my interviewees re-negotiate their professional identities by changing their affiliations, and their epistemic authority, that is, their authority to make scholarly claims using geotagged social media data.

Social scientists might feel the need to differentiate their research from technical geosocial scholarship because 'technical' scholarship with 'big data' is often positioned as a new and more reliable form of social science (cf. Bartlett et al., 2018). For example, Sloan and Morgan (2015), Lazer et al. (2009), Pentland (2015) and Galam (2012) argue that the availability of digital traces opens opportunities for a new science, which they call computational social science, social physics and sociophysics, respectively. A number of social scientists highlight the relevance of their approach and push back against the claim that computational social science enables more robust or better social science. For example, Marres (2017, p. 21-22) argues that scholars who stake a claim for 'computational social science' such as Lazer et al. (2009) appropriate the label 'sociology' to describe data science projects and in doing so falsely claim that the latter can "solve the problems of social research". Marres and several interviewees with a social science background quoted in this section argue that mainstream computational social science provides a narrow perspective on how to study societies that use digital technologies. My participants seek to develop data analysis approaches which differ from technical scholarship.

In response to the First and Second Research Questions - which explore how geosocial research approaches differentiate and differ - in line with Chapters Four and Five, section 6.3 argued that a diminishing shared literature base (jointly cited authors) between 'only social' and 'only computational' geosocial papers plays an increasingly important role in connecting disparate geosocial research approaches. Thus, in response to the Second Research Question, I argued that social and technical geosocial research can be understood as distinct approaches. I analysed the strength of the author-bibliographic coupling relations between two sets of geosocial papers: those published in journals associated with the 'only social' or 'only technical' Broad Disciplinary Categories between 2010 and 2019. My analyses showed that the proportion of authors cited by both sets of papers decreases over time. In addition, it showed that since 2016 or 2018 (depending on the normalisation method used to calculate the values of the weighted author-bibliographic coupling edges) this decreasing proportion of shared citations among geosocial papers published in 'only social' and 'only technical' journals render the network less modular, or in other words, more cohesive. These shared author-citations increasingly link diverging paper sets.

As the Chapter Nine will discuss in more detail, combining the line graphs and interviews in section 6.2 helps answer Research Question 3.1 - which asks how methods mixing can help study the differentiation of and differences among geosocial research approaches. Firstly, combining line graphs and interviews enriched my argument that geosocial research is a collection of research but not a coordinated community, that finding disciplinary and sub-disciplinary research communities they belong to is essential for social scientists to differentiate their geosocial research, and that social geosocial research is currently on the rise. Secondly, plotting the disciplinary distribution of geosocial papers with respect to all geosocial papers helped explore differences in the composition of the interview and scientometric fields.

In response to Research Question 3.1, I also combined interviews and structural scientometric analysis to trace the differentiation among the social and technical geosocial research approaches. This helped me explore the differentiation of these approaches over a longer time period and accounting for a larger set of geosocial scholarship. It enriched my argument that the two can be understood as separate approaches, and their differentiation happens gradually, over time.

Finally, in response to Research Question 3.2 - which explores how to evaluate the affordances of computational methods 'for STS' - I discuss the affordances of structural, homogeneous network analysis in the interpretative context, comprising findings in Chapters Four through Six. In this chapter I calculated the network modularity of counterfactual networks to trace social scientists' efforts to differentiate their approach from technical geosocial approaches. Three main prior findings informed this method. Firstly, using mixed methods in Chapters Four and Five I argued that connections between social and technical approaches - such as social scientists' critical reading of technical scholarship, using or modifying computational methods, collaboration through loose common ground and shared aesthetics - were essential for geosocial research approaches to differentiate, and specifically, for social scientists to differentiate their approaches from technical ones. I assumed that author-citation links between geosocial research published in 'only social' and 'only technical' journals capture some of these links. Secondly, through combining interviews and descriptive statistics, in this chapter I argued that increasing belonging to social scientific communities who value geosocial research help social scientists develop their geosocial research approaches. Thus, I assumed that social and technical geosocial research increasingly differ. Thirdly, based on interview quotes presented in this chapter I argued that journals' disciplinary categorisation signify differences in the geosocial research they publish. Interviewees with different approaches tend to publish in journals associated with their disciplines for research evaluation reasons and to find venues which welcome their research. Based on these, I hypothesised that a decreasing proportion of links between geosocial papers published in 'only social' and 'only technical' journals (implicitly signaling a growing proportion of links among papers published within 'only social' or 'only technical' journals) play an important role in connecting geosocial research approaches which increasingly differ over time. In addition, I found that social scientist and geography interviewees choose publication venues associated with their disciplines whose methodological norms fit their computational research approach for research evaluation purposes. Thus, I assumed that geosocial papers aggregated on the level of Broad Disciplinary Categories 'only social' and 'human geography' - which use journal level classification - roughly captured social geosocial research output.

Importantly, this analysis does not make a strong assumption about the coherence of social and technical approaches, or about how the number of clusters map onto the social and technical approaches I hypothesise through interviews. Rather, it traces how relational practices among social scientists and between social and technical scholars - operationalised as network relations through their assumed scientometric traces - render geosocial research more or less coherent, cohesive or diverse over time.

Chapter Seven will further explore differences between approaches to geosocial research using mixed methods. It will compare the social and technical approaches this chapter differentiated. In addition, it will use scientometrics inductively to identify the scientometric traces of geosocial research approaches, and identify a third approach - geographic geosocial research.

Chapter 7

Exploring the Difference Between Social and Computational Geosocial Research Approaches

This chapter uses mixed methods to explore differences between social and computational approaches to geosocial research identified in Chapters Five and Six, contributing to all three research questions.

In response to the First Research Question - which asks how geosocial research approaches develop - section 7.1 argues that reflecting on how social media platforms and analytical decisions shape the characteristics of geosocial data, and that the knowledge interviewees create with them about spaces is central to the development of geosocial research approaches.

In response to the Second Research Question, which asks how approaches to geosocial research differ, I use mixed methods to explore differences between social and technical geosocial research approaches. Using participants' quotes, section 7.1 argues that technical and social geosocial researchers' reflexivity differs. Whilst social geosocial interviewees' reflexivity foregrounds experiential and historic perspectives, technical geosocial interviewees focus on the impact of computational calculation methods in relation to demographic categories. Using computational STS methods, section 7.2

compares the topics and spaces computational and social geosocial approaches study, and the methods they use. Contrasting the two approaches highlights the diversity of collective practices and forms of participation (e.g. in civic, intercultural, commercial, health and disaster-related practices) social geosocial research studies at specific locations (e.g. at specific cities). Comparatively, its focus on studying these situated practices and events, on diverse spatio-temporal scales and on computational methods development differentiates technical geosocial research from social approaches. Secondly, through combining scientometrics and interviews, I identify a third approach to geosocial research - geographic geosocial research - differentiated by its use of geographic methods and study of diverse types of spaces.

Finally, this chapter's mixed methods approach contributes to the Third Research Question - which explores combinations of scientometrics and STS as part of one case study - in three main ways. Firstly, it uses scientometrics inductively, and thus illustrates the potential of computational methods to identify qualitatively meaningful patterns. Secondly, it uses scientometrics to compare geosocial research approaches, rather than tracing relational practices that help scholars differentiate them as Chapters Five and Six did. I find that combining methods foregrounds the way my choice of comparative units shapes the knowledge I create. Thirdly, it visually highlights my scientometric findings' contingency on the data analysis infrastructure.

7.1 Reflexivity

This section argues that reflecting on how characteristics of geosocial data and analytical decisions shape knowledge production about spaces is core to interviewees' geosocial research. I refer to this practice as 'reflexivity'. Several social scientist interviewees differentiate their geosocial research from technical geosocial research in terms of their own reflexivity. I argue that although reflexivity is central to both social and technical geosocial research, the reflexivities these approaches entail differ. Through interviewees' narratives, I differentiate between hermeneutic and algorithmic reflexivity - discussed in sections 7.1.1 and 7.1.2 - which I associate with social and technical geosocial research respectively. I consider my descriptions to be conceptual models which simplify social and technical geosocial research practices to capture differences between them.

7.1.1 Hermeneutic Reflexivity

Six interviewees with social science background from groups A-E differentiated their approach from technical geosocial research in terms of two reflexive acts: their reflexivity about how their lived experience shapes their research, and how their geosocial research accounts for the practices or perspectives of diverse groups. These reflexive acts are considered central epistemic virtues in their home disciplines, including anthropology, 'interpretive' sociology and human geography. I refer to social scientist participants' reflexivity as 'hermeneutic reflexivity', drawing on Fortun et al.'s (2016) notion of hermeneutic expertise, which amounts to "taking into account what things mean, to whom, why, and to what end" (p. 3). Figure 7.1 depicts three key features of hermeneutic reflexivity discussed below: reflecting on how analytical decisions and social media platforms mediate (geosocial) knowledge about spaces, predominantly in experiential and historic terms.

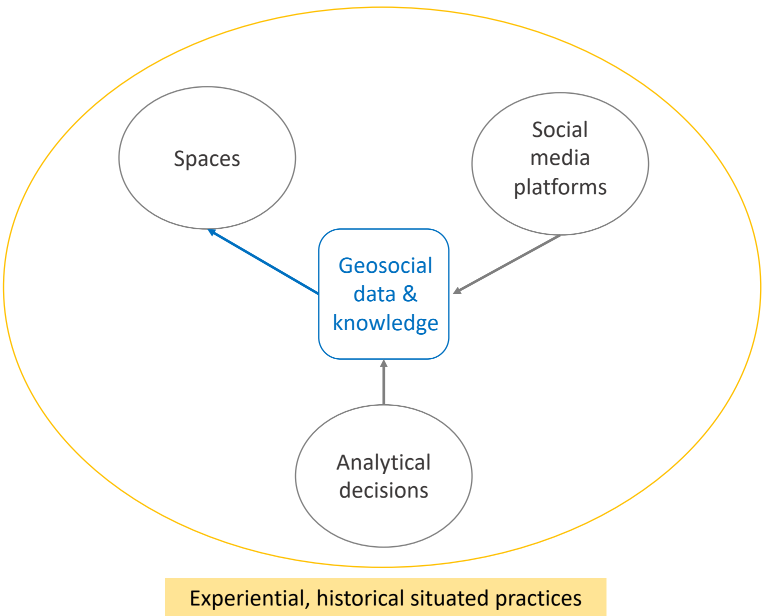


Figure 7.1: Hermeneutic reflexivity

Firstly, interviewees from groups A, B and D discussed reflecting on their analytical decisions in experiential and historic terms. Bruno, Anne and David distinguished their approach to geosocial research from technical approaches in terms of such reflection, and Brian noted the importance of reflecting on the history of algorithms.

As section 5.3 discussed, Bruno stated that his team's local and experiential knowledge about the spaces they study shape their research questions and methods. Thus, it is constitutive of their findings. Being part of a multidisciplinary PhD program, Anne was faced with a dilemma of how to use computer programming for her geosocial research. After experimenting with coding and considering outsourcing computational analysis to others, she decided to use a method - thematic analysis - with which she was more familiar. She noted her satisfaction with this choice because she thinks thematic analysis affords more opportunities to reflect on how her experiences impact the knowledge she produced.

> *"I came in [to the PhD program] with a clear idea, you know, that we need to push against [computational social science] [...] I'm glad I stuck with this kind of manual method, because I think it allows greater space for reflexivity as well. To say, 'what are the things that I find difficult or boring, or exciting about this research?' and 'How am I framing my own work?'"*

Like Anne, David stressed the importance of reflecting on how his and Daniel's lived experiences shape their results. He stressed that it is dangerous to produce knowledge about a community encountered solely through online traces, given the discrepancy between their own and their subjects' lived experiences of the neighbourhood they study.

> *"Here we are in [this city] as two white guys, and we kind of know a lot about the lives of this scene in [another city], and we personally never met [...] I think we're very acutely aware of the danger of, yeah, doing some kind of violent abstraction, and just kind of having these figures stand in for, you know, as they often are, in public discourse, some [grand claims about society]..."*

Finally, Brian reflected on the knowledge he creates in terms of the history of the algorithms he uses:

> *I use algorithms all the time [and they] often date back many decades. Sometimes to truly understand, we need to excavate, and I think these*

> *histories give you a better understanding of how an algorithm works but also why it works in that particular way. [...] They often have a set of parameters. [...] maybe somebody decides to hard code a particular value [... and] if you don't excavate that you might actually never find out even though [it] can be very important to [...] change."*

Altogether, Anne, Bruno, Brian and David stressed the importance of reflecting on how their lived experience or historical processes shape their analytical decisions and findings. Next, I discuss a second aspect of hermeneutic reflexivity: reflecting on how social media platforms mediate geosocial data and knowledge in experiential terms.

Reflecting on how social media platforms mediate users' experience and thus impact geosocial data and knowledge is central to Anne's, David's and Daniel's geosocial research. Anne stated that her method treats social media data in a way that accounts for, or mirrors, everyday social media users' experience, and thus helps her study social media platforms' impact on (users') narratives of space.

> *"I was kind of trying to integrate the kind of everyday experience of [the social media platform] into my analyses. [...For instance,] you know, what Joe gets from understanding [a city], if you do a quick search on [the social media platform]. [I used analytical solutions] to mirror the way that people see tweets in their feed."*

Like Anne, David and Daniel reflected on differences in the social media platforms' mediation of user groups' practices. Their dataset changed after the social media platform altered its geotagging policy. They discussed the effect of changes in terms of users' cultural and experiential practices: their members, spatial footprint, and the types of spaces they tagged. They stated that platform changes affected user groups differently. For example, a user group with a substantial proportion of non-white users who were tagging locations outside of gentrifying areas disappeared from their dataset after the policy change.

> *Daniel: "it was the only cluster that [...] had a substantial proportion of non-white people, and also their spatial footprint was very, very different. [...] but they just disappeared."*

> *David: [...] generally we retained a fair proportion of users [after the policy change], but in that case, like 4 out of 5 users were no longer*

> *in the data. I mean it's really dramatic, like they literally fell off the map. [...]"*

They ironically invoked the notion 'natural experiment', a method popular in some branches of social science where participants are assigned to treatment and control groups by processes outside of the researchers' control. A seminal example is David Snow's study on the 1854 cholera outbreak in London, where participants were divided into control group (healthy) and treatment group (infected with cholera) by the epidemic process itself, rather than by the researchers (Freedman, 2009). David and Daniel refrained from using the terms 'natural experiment' and 'system drift' - the latter used by Salganik (2017) - to refer to changes in their data. They felt these concepts falsely suggest that changes in the data are independent of the situated practices that create them and their own methods. Thus, they seek alternative concepts to express the combination of human and non-human agencies that shape data.

> *"We have played a little bit with regarding this [changes in the data before and after an API change] as a natural experiment [laughing] ... we're completely comfortable analysing [this data] because of our background in [an interpretivist sociological tradition] and our familiarity with the neighbourhood. And then to write about a natural experiment doesn't really fit too well... We still have to think of the good way to write this up [...]. Data disappeared from the map - but only from our map! ... Matt Salganik calls this 'system drift' [...] I don't know what the right metaphor is, cause even drift sounds like 'oh it's just a naturally occurring process'..."*

Altogether, Anne, David and Daniel reflected on how their research (methods) account(s) for how social media platforms mediate users' experiences and practices. Next, I discuss the third aspect of my participants' hermeneutic reflexivity: their reflection on the extent to which their research can take diverse groups' situated practices into account.

Anne, Brian, Colin, David, Daniel and Elias assume the spaces they study are co-constituted by diverse actors' practices, and reflected on whether their research can account for these. Anne considers spaces inherently diverse, and assumes that social media data shapes users' experiences of them. Her research explored how filtering based on geolocation - a feature of the social media platform from which she collected data - shapes the diversity of narratives users - and herself - have about spaces.

> *"when you're engaging with these subjectivities about spaces, then the idea of something being valid becomes very difficult [...] [my research was in part methodological,] exploring the subjective narratives about space and the potential of social media data to assist in those understandings. And realising that geotagged filtering is a huge, commonly used method, [...] I was keen to [...] see how it is affecting our understanding of spaces..."*

Brian differentiated his research from computational social science by contrasting the latter's quest for optimisation with interpretive social sciences' reflexive tradition that seeks to study diversity, where optimized solutions often do not exist.

> *"the word optimization is used a lot in engineering and computer science, [...where scholars are] interested in developing new algorithms because they're more computationally efficient... but [in] social scien[ce] [...] many questions [...] cannot be optimized. There is no optimal solution. And that is partly what makes social science so interesting. So yeah, there are these foundational differences..."*

Colin claimed that the main difference between physicists' and his methods is his field's predisposition to question findings, and in particular, the way the analyst's decisions foreground or obscure diversity. For example, he worries about whether his team's analyses accounts for the perspectives of a minority group living at the space they study.

> *"...What is the value of [...] being always in doubt and not truly believing the results of any data analysis? [...] We had a joint workshop with physicists, and they said that physicists don't clean data. In contrast, worrying about who we should leave out, who our analyses are leaving out, and what the sample consists of kept me up all night... [...] For example, what should we do about [a minority group]? Are they under or over represented in this data?"*

Daniel and David stated they chose to study a cluster of social media users because their social media footprint differed from the rest of the users. Thus, their research specifically aims to account for diverse situated practices. Users in the chosen cluster posted from different areas in the city, and (as outlined above) were ethnically more diverse.

> *"So all the clusters are like very heavily concentrated in the city centre, especially sort of gentrifying areas, but they, they didn't really*

> *have a centre, they were scattered all over [the city]... we wanted to look at that because they were interesting in that they were exceptional..."*

Finally, Elias claimed that his approach differs from research that privileges technically complex methods and uses quantitative criteria to assess the accuracy or validity of results. According to him, such criteria fail to account for society's diverse practices and perspectives. He also critiqued framing data quality in demographic terms, which he believes does not capture diversity.

> *"political science has really tended towards this sort of veneer of scientific-ness based on its use of complicated methods. [... For example, it has a] strong focus [...] on like – is your data representative? Is it kind of descriptive or is it rigorous and analytic? [... But] these are kind of binaries that don't really account for the messiness of society."*

In sum, Anne, Brian, Colin, David, Daniel and Elias's quotes highlight the third aspect of hermeneutic reflexivity: they conceptualise spaces and situated practices as diverse, and reflect on how their geosocial research can capture, or account for this diversity.

Altogether, this section illustrated hermeneutic reflexvity's focus on assessing how analytical decisions and social media platforms shape geosocial data and knowledge about spaces in experiential and historic terms. Next, I present algorithmic reflexivity's focus on assessing the impact of analytical decisions and social media platforms' affordances on geosocial data and knowledge about spaces in terms of calculations and demographic categories.

7.1.2 Algorithmic Reflexivity

Like social scientists, technical interviewees reflect on how analytical decisions and social media platforms shape geosocial data as well as the knowledge they enable about spaces. However, as figure 7.2 depicts, in contrast to hermeneutic reflexivity, algorithmic reflexivity focuses on the impact of calculations on knowledge about spaces assessed in demographic terms. In addition, instead of conceptualising it as a unique contribution, technical interviewees narrate their reflexivity as a necessary aspect of their research. Only Luke differentiates his geosocial research in terms of his reflexivity. This section illustrates the above three aspects of 'algorithmic reflexivity' depicted by figure 7.2 through interviewees' quotes.

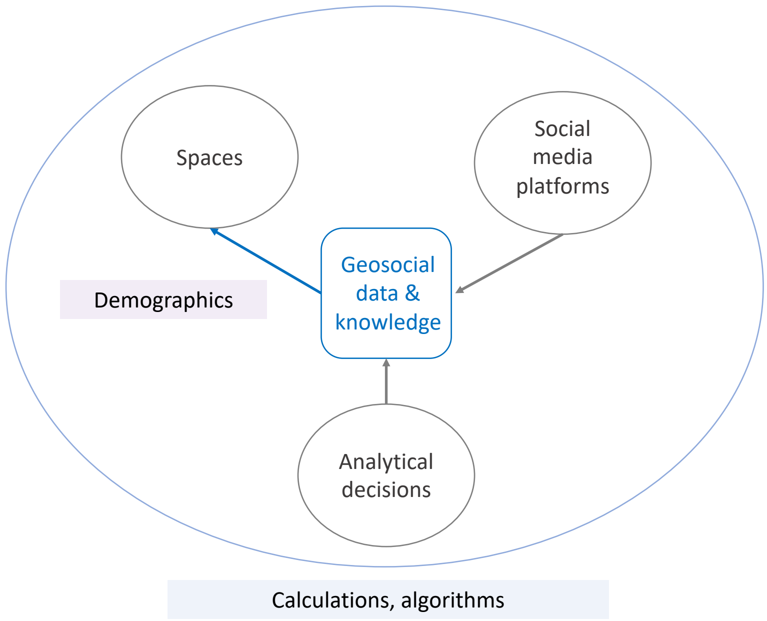


Figure 7.2: Algorithmic reflexivity

Firstly, Jane's, Kevin's, Miles', Mike's, and Isaac's quotes illustrate that reflecting on analytical choices in terms of calculations is central to technical scholars' geosocial research. As section 5.2.1 discussed, Jane is interested in reflecting on how the modeling approach popular in her discipline shapes the knowledge she creates, and Kevin reflected on the suitability of specific community detection methods in mathematical terms. Section 5.3 showed Miles' reflection on the calculations they perform when optimising a data model. Mike, quoted below, acknowledged that there are several valid ways to normalise data which yield different and equally valid findings. He believes that this heterogeneity hinders the development of standard methods accepted by the research community. Instead, he claimed that scholars must argue for the adequacy of their chosen normalisation method:

> *"There are several ways of normalising [...] In several papers, we had this exchange with reviewers [...] There are different techniques, all of them are, you know, valid, but of course, they map different things. So getting a standardised methodology which can be, you know, accepted, by everybody, I think it's a challenge."*

Isaac explained that even though his team members reflect on both the way their decisions and computational infrastructures impact the knowledge they create, they tend to omit such reflections from publications in part to avoid possible questions during peer review. This suggests that published literature may not reflect the scope of the reflexivity which forms part of technical geosocial scholars' research practice.

> *"To be honest, there's also kind of this game that you have to play, that if you write too much about your data collection and [analytical decisions], people will start asking questions. If you don't go into detail, they won't think about it as much. It might be a little bit ugly, but that's the truth."*

In sum, technical interviewees predominantly reflect on their analytical decisions in terms of the impact of their calculations, acknowledging that often, multiple acceptable calculations exist. Next, I discuss algorithmic reflexivity about how the calculations performed by social media platforms shape geosocial data and knowledge.

Jane and Luke reflected on how calculations performed by social media platforms' algorithms shape geosocial data and the knowledge scholars create with them. Jane noted that the sampling method used by the free API of the social media platform her team uses to access geosocial data can shape the data (for example, by omitting data points at regular time intervals). She explained that her team modifies data sets - for example, by removing automated posts - based on their understanding of these algorithms.

> *"Since we use the [free API], the sampling method could introduce further biases. [...] We remove bots based on the time stamp of posts because bots post at regular time intervals. [...] the [API] algorithm samples data based on the time stamp of posts. As a result, your sample can either completely over represent, or under represent specific bots, depending on [...] the sampling time frame."*

Luke demarcated his geosocial research in terms of how it accounts for the ways social media platforms' algorithms shape geosocial data. His method filters geosocial data based on patterns identifiable in diverse modalities, including posts' text and geotag.

> *"A lot was needed to clean the data spatially. A lot of the existing literature [...] identifies bots by the number of following and time of [posting] [...] and that's very limited [...] my approach took it from a more of a location perspective."*

In sum, Jane and Luke's narratives highlight algorithmic reflexivity's tendency to assess how calculations performed by social media platforms' algorithms shape geosocial data and knowledge. Finally, I discuss quotes which illustrate that algorithmic reflexivity assesses geosocial data and the knowledge they enable about spaces in demographic terms.

Isaac, Jane, Josh and Mike stated that geosocial data are limited because they do not 'represent the population' of the spaces they study. Isaac explained that geosocial data are not representative of age groups:

> *"It's certainly biased, so it's not representative of the society as a whole [...] if I only look at [social media], perhaps I have most of the people under 30, significant number of people under 50 and then relatively few older people. And perhaps for under 30 people segregation is no longer an actual thing. I don't believe that, but you know it's... I can never really test that validity..."*

Like Isaac, Jane stated that geosocial data are not representative of age, socio-economic status or ethnicity:

> *"It's not representative of age, socio-economic groups. In certain areas, people from lower socio-economic status are over-represented. In the [country we studied] it's also not representative for ethic groups..."*

Josh also stressed the non-representativity of geosocial data, and stated that their analytical methods account for this:

> *"Obviously, we have to admit that [the social media data] has its limits, for example, it's not representative. We always accounted for this in our work."*

Finally, Mike highlighted the non-representativeness of social media data by noting that highly active users in certain areas can yield biases:

> *"...you also find out the problems, like you know, of course biased representations, you can have entire areas in city that contains [posts] by the same user. You have to figure out all these biases and problems with online data."*

In sum, interviewees from groups I, J and M reflected on geosocial data and the knowledge they afford about spaces in demographic terms: they stress that geosocial data are not representative of age, socio-economic status and ethnicity of the spaces they study.

This section argued that reflexivity is a core aspect of geosocial research, and that social and technical geosocial research differ according to the type of reflexivity they foreground. While the former foregrounds experiential and historic reflection, the latter assesses the impact of calculations in demographic terms. The next section further explores differences among geosocial research approaches using scientometrics.

7.2 Computational STS Comparison: Social & Technical Geosocial Research

This section provides further scientometric evidence that support the distinctions between social and technical geosocial approaches (discussed in Chapter Six), and compares them using computational STS methods. As section 3.4.5 discussed, I clustered the author-bibliographic coupling network of geosocial papers using the Leiden algorithm.

Figure 7.3 depicts the network of the clusters of the author-bibliographic coupling network *G3*, which includes all papers in my scientometric field. Clusters 0 - 3 contain several papers, but clusters 4 and 5 each only have 2 papers that share no cited authors with the papers in the other clusters. The length of the edges indicates the proportion of shared author-citations between clusters. The closer clusters are, the higher the proportion of jointly cited authors. Figure 7.4 depicts the distribution of papers across five Broad Disciplinary Categories in each cluster using the method discussed in section 3.4.5: 'multi- or interdisciplinary', 'physical geography', 'human geography', 'only social' and 'only computational'. While figure 7.3 illustrated clusters' connections, figure 7.4 highlights their differences. 'Only social' papers dominate Cluster 0, and there is a relatively high number of geography papers in Cluster 1 compared to the other clusters. Comparatively, the proportion of 'multi-interdisciplinary' papers is the highest in Cluster 2, alongside a relatively high proportion of computational and human geography papers. Finally, 'only computational' papers dominate Cluster 3. Thus, although technical papers are distributed across clusters, I argue that this analysis differentiates the scientometric footprints of the 'social geosocial research' (Cluster 0) and 'technical geosocial research' (Cluster 3) approaches. This suggests that the distinction between social and technical approaches identified in Chapter Six through interviews and by analysing a network which comprised 'only social' and 'only computational' papers holds in the context of all geosocial papers. In addition, Cluster 1 suggests that geographic geosocial research is a distinct approach. Figure 7.3 highlights the relatively high proportion of jointly cited authors between the 'social' (cluster 0) and 'technical' (cluster 3) clusters, illustrating their connectedness. The technical cluster shares relatively more author citations with cluster 2 compared to geography dominated cluster 1, while the social cluster cluster shares more author citations with the geography cluster (cluster 1) compared to cluster 2. Exploring these edges qualitatively in detail is beyond the scope of this project, but Chapter Eight will compare the clusters of this network in more detail.

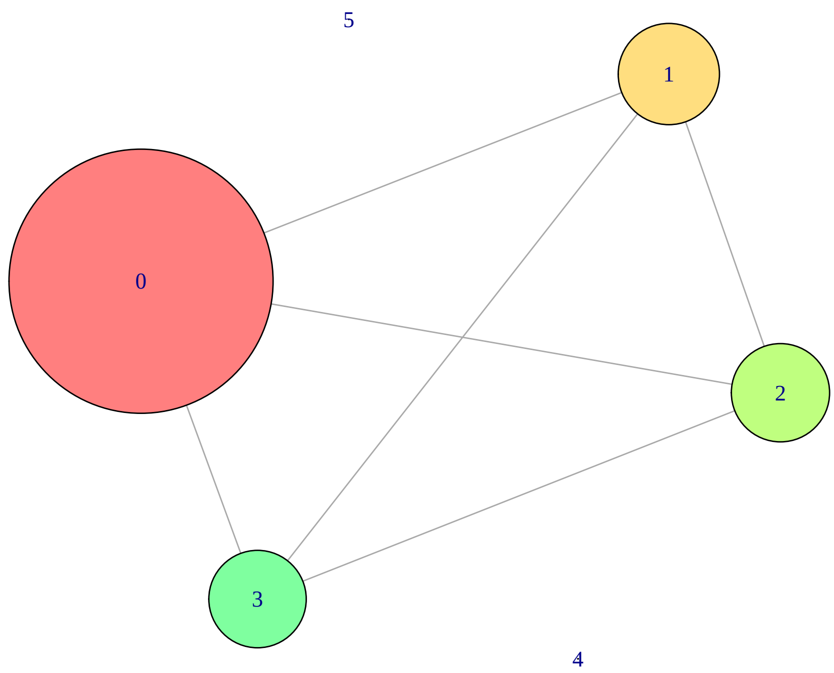


Figure 7.3: Network of the clusters of *G3*

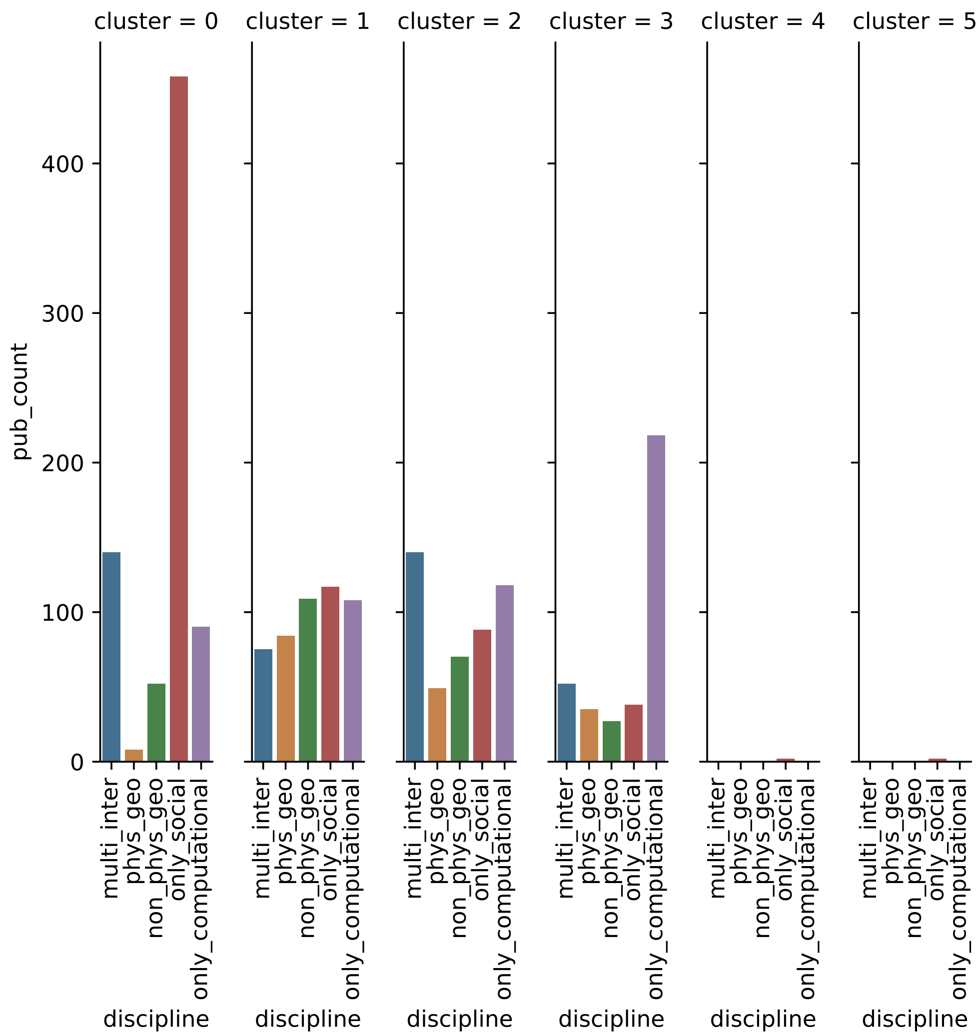


Figure 7.4: Distribution of subject categories across the clusters

To scientometrically compare social and technical approaches to geosocial research, below I cluster the author-bibliographic coupling network comprising 'only social' and 'only computational' geosocial papers *(G4)*. As section 6.3 argued, the set of 'only social' and 'only computational' geosocial papers better aligns with the disciplinary background of my interviewees than my entire scientometric field (depicted by *G3*). Thus, I assume that the comparative findings about social and technical geosocial research approaches using their scientometric footprints through clustering *(G4)* better resonate with my interview findings than comparing the above social and technical clusters in the network *G3*. As figure 7.5 shows, the community detection on network *G4* identified five clusters (Clusters 0 - 4). Figure 7.5 depicts the number of papers in clusters across the five Broad Disciplinary Categories.

As figure 7.5 shows, Cluster 0 contains the majority of 'only social' papers along with smaller number of 'only computational' papers. Cluster 1 contains the majority of 'only computational' papers along with smaller number of 'only social' and 'physical geography' papers, and an even smaller number of 'human geography' papers. The partial separation of 'only social' and 'only computational' papers in Clusters 0 and 1 support Chapter Six's findings about the growing separation between these approaches enabled by information exchange among scholars who practice them through collaboration, as well as shared methodological and aesthetic interests. Thus, I argue that Cluster 0 predominantly comprises papers associated with social geosocial research, and Cluster 1 predominantly comprises papers associated with technical geosocial research - capturing these approaches' scientometric footprints. Cluster 2 contains mostly 'only social' and 'human geography' papers, along with smaller number of 'only computational' and 'physical geography' papers. Given the relative dominance of geography papers in Cluster 2 compared to Clusters 0 and 1, I interpret this cluster as a geography oriented cluster. This suggests that the geographic approach to geosocial research can be seen distinct from social and technical approaches identified in Chapters Five and Six. Clusters 3 and 4 contain 2-2 social scientific papers that shared no cited author with other papers in the dataset. Thus, I excluded these papers from my analyses.

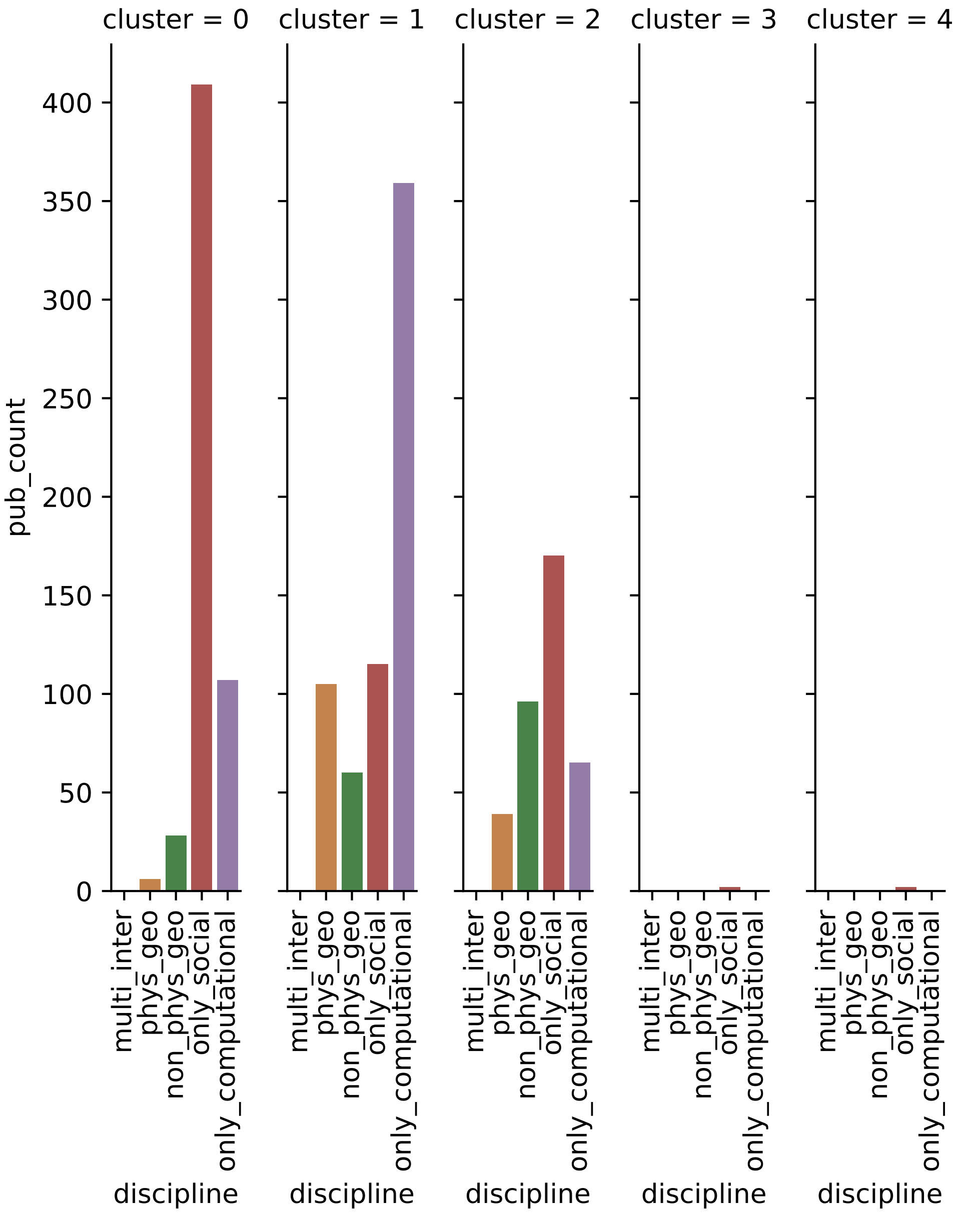


Figure 7.5: Technical and Social Geosocial Research Author Bibliographic coupling network clusters' disciplinary categories

A few interviewees' comments support the scientometric separation of geographic geosocial research. Isaac, David, Daniel and Mike noted that initially they did not consider their research explicitly geographical, and did not engage with the geographic relevance of their research in depth.

David and Daniel, who have backgrounds in sociology, noted that the interview with me made them reflect on the geographic aspect of their research. They noted that they did not relate their research to geography scholarship as much as they could:

> *"After this interview, I'm also thinking that maybe we should position ourselves to other [scientific] groups as well.[...] We also make a point out of this long term perspective. The geography, I'm not sure if we make as strong a case as we could."*

Isaac, an applied mathematician noted that he does not consider his geosocial research geographic:

> *"So, geography is not usually on my mind when I think of a problem or a [research] question..."*

Finally, Mike noted that his computer scientist team is learning about geographic quantitative methods as they develop their geosocial research. He would welcome a collaborator with background in geography, but also noted that his team could learn the relevant methods by themselves.

> *"...it'd be super great to have people from geography [on our team], because we are slowly learning about tools that geographers use [...] and they are super interesting to make the analysis much better [...] even just very simple spatial models [...] and spatial correlations [...] But [...] we could also learn these ourselves..."*

In sum, the above interviewees' narratives support my scientometric finding about the separation between geographic, and social or technical geosocial research. Next, I compare these three geosocial research approaches.

To explore differences between social and technical geosocial research approaches, also in relation to the geographic approach depicted by Cluster 2 (the Geography Cluster), figures 7.6 and 7.7 depict the top 30% noun phrases based on their TF-IDF values, extracted from the abstracts and titles of papers in Clusters 0, 1 and 2 of network

G4, categorised into themes as outlined in section 3.4.6. I further divided the top 30% into three ranges depicted by the font color. Grey terms correspond to the top third, blue noun phrases are in the second third, and yellow terms are in the bottom third within the top 30% of noun phrases ordered by their TF-IDF scores within each cluster. As section 3.4.6 discussed, due to the nature of TF-IDF scores, these figures emphasize relative, comparative differences among these clusters, rather than absolute differences or characterising clusters in detail. I highlight five main differences between the scientometric footprints of the social and technical approaches geosocial research when contrasted with one another and the Geography Cluster (Cluster 2).

Firstly, terms' categories that differentiate clusters differ. A large proportion of the noun phrases that differentiate Cluster 0 (the Social Cluster) from the other two clusters are related to collective practices and actors. In contrast, a large proportion of noun phrases that differentiate Cluster 1 (the Technical Cluster) relate to methods. In addition, compared to the Social Cluster, a large proportion of noun phrases that differentiates the Technical Cluster (Cluster 1) express spatial units. Finally, the relatively large proportion of noun phrases that express spaces differentiates Cluster 2 (the Geography Cluster), which further illustrates its distinctly geographic focus. Next, I compare the meaning of terms within each category across clusters.

Secondly, whereas frequently mentioning actors with diverse roles and subjectivities, such as 'reporter', 'respondent', 'institution', 'adolescent', 'patient' sets the Social Cluster apart; noun phrases related to technology users who create geosocial data (e.g. 'similar user', 'human') and the 'police' set apart the Technical Cluster's actors. Similar to the Social Cluster, the three actors that are relatively more frequent in the Geography Cluster compared to the other two clusters - 'activist', business' and 'resident' - relate to diverse roles and subjectivities.

Thirdly, the relatively frequent mention of specific locations, such as 'Seoul', 'New Zealand' and 'Hong Kong' differentiates the Social Cluster. In return, the Technical Cluster's relatively frequent use of noun phrases which express diverse spatial scales, such as 'district', 'city level', 'spatial scale', and spatial processes that can be expressed in spatio-temporal terms, such as 'urban vibrancy' and 'urban dynamics' sets it apart from the other two clusters. This suggests that comparatively, the Social Cluster focuses more on studying specific locations, whereas technical geosocial research places more emphasis on studying diverse spatial scales. Space related terms that differentiate the Geography Cluster include terms that refer to specific locations (e.g. 'London' or 'Australia'), types of spaces (e.g. 'public space') and spatial patterns, (e.g. 'spatial distribution', 'spatial pattern'), illustrating its rich spatial vocabulary.

Fourthly, while terms associated with media which express their use differentiate the Social Cluster (e.g. 'Facebook page', 'online activity', 'social media use'), my analysis highlights the Technical Cluster's relative focus on media in relation to methods (e.g. 'user generated content', 'social sensor') and technology development (e.g. 'prototype', 'software'). The Geography Cluster's media related concepts also illustrate its relative focus on geography, for example, through the noun phrases 'geoweb', 'geographic information system'.

Finally, intervention and power-related temporal concepts such as 'intervention', 'mobilisation', 'shift', 'resistance' set the Social Cluster apart, compared to the Technical Cluster's methods-related temporal concepts that express methodological innovation and spatio-temporal units (e.g. 'new method', 'spatiotemporal pattern'), and Cluster 2's distinct temporal terms, which include both time related to collective practices (e.g. 'activist', 'migration') and methodological time (e.g. 'trend', 'flow'.)

Altogether, I argue that a relative focus on diverse actors' situated practices, forms of participation in civic or commercial practices, and changes in power relations differentiate social scientific geosocial research from technical geosocial research. On the other hand, its relative focus on using computational data analysis methods to explore spaces at diverse spatial and temporal scales differentiates technical geosocial research from social approaches. Finally, the use of geographic methods to explore diverse types of spaces differentiates geographic geosocial research from the social and technical approaches.

Collective practice	public discourse, feeling, harassment, sociality, intervention, reflection, collaboration, publishing, public opinion, framing, professional, promotion, struggle, interact, apology, training, race, advertising, demand, trust, mobilization, reaction, assessment, political communication, public engagement, reason, commitment, matter, making, collective action, play, course, writing, credibility, exposure, inequality, shift, brand post, suggestion, creation, reputation, incident, competition, sharing, examination, exchange, resistance, racism, street art, encounter, health, distinction, intention, public relation, social capital, sentiment, awareness, history, solidarity, governance, status, norm, gender, style, civic engagement, class
media	Facebook page, new medium, online medium, affordance, app, online activity, media, television, social media use, journal, hashtag
space	Seoul, New Zeland, Hong Kong, United Kingdom, Europe, Canada, physical space, Singapore, smart city
actor	race, reporter, subject, respondent, institution, organisation, patient, adolescent, family, product, migrant, firm, employee, local government, scholar, manager, public library
method	questionnaire
time	intervention, mobilization, reaction, new medium, shift, resistance, history, class

Collective practice	public health, emergency response, crowdsourcing, travel, natural disaster, human mobility pattern, air pollution, disaster management, health, human dynamic, human activity pattern, information diffusion, behaviour, privacy, rumor, law, social interaction
media	ugc (user generated content), social sensor, prototype, software, bottari, yelp, weibo, geotagged tweet, recommendation system, smartphone, communication technology, lbsn
space	district, city level, urban dynamic, urban boundary, urban environment, urban vibrancy, physical location, human settlement, spatial scale, urbanisation, hotspot, study area, earthquake, air quality, home location, spatial interaction, spatiotemporal pattern, road segment, street, land use, footprint
actor	similar user, human, agent, police, social media user
method	error, new method, new perspective, better understanding, baseline, layer, magnitude, spatial information, test, satellite, probability, resolution proposed method, real world dataset, deep learning, retrieval, estimation, contextual information, usefulness, improvement, noise, large amount, temporal information, large number, topic model, dbscan, aerial image, location data, feasibility, metric, complexity, emergence, event detection, million, modeling, location information, social sensing, identification, implementation, matrix, signal
time	new method, new perspective, emergency response, urban dynamic, human activity pattern, spatiotemporal pattern, event detection

Figure 7.6: TF - IDF clusters 0 and 1

CLUSTER 2 – 'GEOGRAPHY CLUSTER'	
Collective practice	policy, crowd, migration, crime, crisis, trend, action, violence, participation, surveillance
media	geoweb, geographic information system, instagram, tag, lbsns, device, social media platform
space	cyberspace, proximity, London, Australia, station, public space, landscape, neighbourhood, flow, boundary, spatial distribution, spatial pattern
actor	activist, business, resident
method	spatial analysis, discipline, mapping, trend, giscience, database, interview, visualization, detection, limitation, effect
time	moment, trend, flow, year

Figure 7.7: TF - IDF cluster 2

As section 3.4.6 discussed, TF-IDF scores emphasize the comparative difference among groups. However, as Chapters Four through Six argued, social and technical geosocial research can differentiate from one another thanks to exchange between them, such as through scholars' shared aesthetics sensibilities, collaboration and efforts to link computational and social scientific disciplines. To better explore similarities between social and technical geosocial research, I will next explore noun phrases present in both the Social and Technical Clusters using the method section 3.4.7 discussed. Figure 7.8 depicts 112 noun phrases categorised using the above scheme, which occur both in the Social and Technical Clusters at least 10 times (*intersection noun phrases*), ranked according to their frequency. These 112 terms comprise 40% of all noun phrases that occur in both clusters 10 times or more. I found a total 278 noun phrases mentioned in both clusters at least 10 times. However, many of these noun phrases were generic words that were difficult to interpret in themselves, such as 'interest', 'topics', 'challenge', 'content'. As section 3.4.7 discussed, I could not categorise these and thus omitted them from further analysis.

The colors in figure 7.8 indicate the number of times noun phrases are mentioned. Yellow noun phrases are among the top 50% most frequent intersection noun phrases in both the Social and Technical Clusters. Green noun phrases are in the top 50% most frequent intersection noun phrases only in the Social Cluster. Pink noun phrases are in the top 50% frequent intersection noun phrases only in the Technical Cluster. Black noun phrases are present in both the Technical and Social Clusters, but are not in the top 50% most frequent intersection noun phrases in either cluster. Terms in all capitals will be further analysed. Due to the data loss during the analysis discussed above, I do not interpret the number of intersection noun phrases per topic category. Thus, figure 7.8 keeps the height of each row, regardless of the number of noun phrases they list.

This analysis highlights several points of connections between the social and technical geosocial approaches, including their shared interest in urbanism (e.g. 'city', 'New York city', 'smart city', 'urban space'), disasters, tourism and public governance and discourse ('citizen', 'news'). In addition, network analysis is relatively popular in both approaches. In light of earlier findings obtained through the TF-IDF analysis, I hypothesise that while social geosocial research studies these spatial events and practices in relation to diverse actors, subjectivities and forms of participation at specific locations, technical geosocial research studies them in light of the affordances of computational methods to study diverse spatio-temporal scales. Next, I use noun phrase co-occurrence maps of the

Social and Technical Clusters to explore this hypothesis.

INTERSECTION: CLUSTERS 0 AND 1	
collective practice	event, activity, use, behavior, movement, life, news, disaster, popularity, mobility, behaviour
media	social medium, tweet, twitter, social network, platform, technology, post, web, video, internet, facebook, site, social media, online, photograph, mobile device, medium, profile, website, communication technology, smartphone
space	location, CITY, place, pattern, area, china, space, context, environment, world, country, structure, region, scale, map, geography, new york city, situation, smart city, building, landscape, boundary, urban space, street
actor	user, person, group, CITIZEN, society, tourist, population, friend, resident, business
method	data, approach, model, analysis, framework, pattern, technique, case study, NETWORK, survey, factor, sample, evidence, set, experiment, dataset, social media data, distribution, methodology, valuation, prediction, comparison, limitation, mechanism, observation, uncertainty, increase, hypothesis
time, process	time, process, year, dynamic, period, real time, today, recent year, month, phase

Figure 7.8: Intersection between clusters 0 and 1

I explored the above hypothesis using the modified ego term co-occurrence networks of the three noun phrases in all capitals in figure 7.8 using the method outlined in section 3.4.8. As section 3.4.8 discussed, whilst TF-IDF and the term occurrence analysis rank noun phrases, the modified ego network maps allow me to explore noun phrases' connections within each cluster. This helps explore the meanings of the selected noun phrases in the context of social and technical geosocial research. To study the above hypothesis, I chose three noun phrases, each associated with a different topic category: 'citizen' is an actor, 'city' refers to a space, and 'network' relates to methods. I chose noun phrases that are relatively popular in both clusters to ensure that the modified ego-networks - which, as section 3.4.8 explained, only contain noun phrases mentioned at least 10 times in the data - include a large number of the chosen interface noun phrases' connections, to help compare their meanings in the two clusters.

As section 3.4.8 discussed, I created modified ego-networks by subsetting the noun-phrase co-occurrence networks of each cluster. Figures 7.9 and 7.10 depict the noun phrase co-occurrence networks for each cluster. As outlined in section 3.3.2, these are screenshots of the interactive data visualisations available through the VOSviewer user interface. The screenshots do not show all of the nodes and edges of the network, and I analysed the maps using VOSviewer's interactive, zoom-able data visualisations. The colors of the nodes depict the association of each noun phrase with clusters within each network, identified by VOSviewer's clustering algorithm. These clusters - and the colors - indicate the extent to which noun phrases co-occur in the abstracts and titles of papers used for the analysis. As section 3.4.8 explained, I created modified ego networks by subsetting these noun phrase co-occurrence networks and keeping only the ego nodes ('citizen', 'city' and 'network') and their alters.

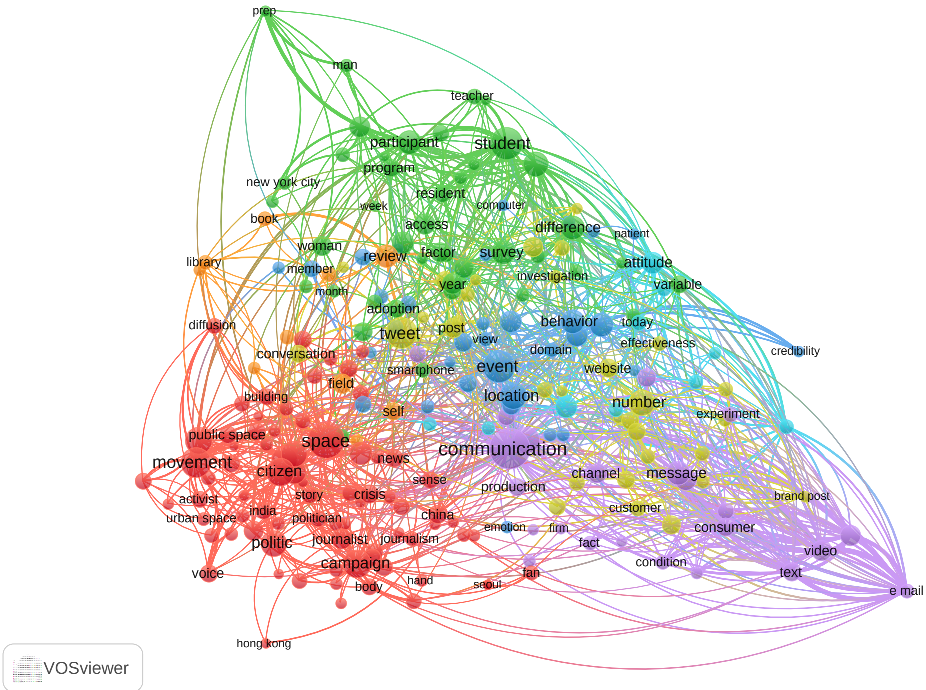


Figure 7.9: Social Cluster (Cluster 0) Term Map

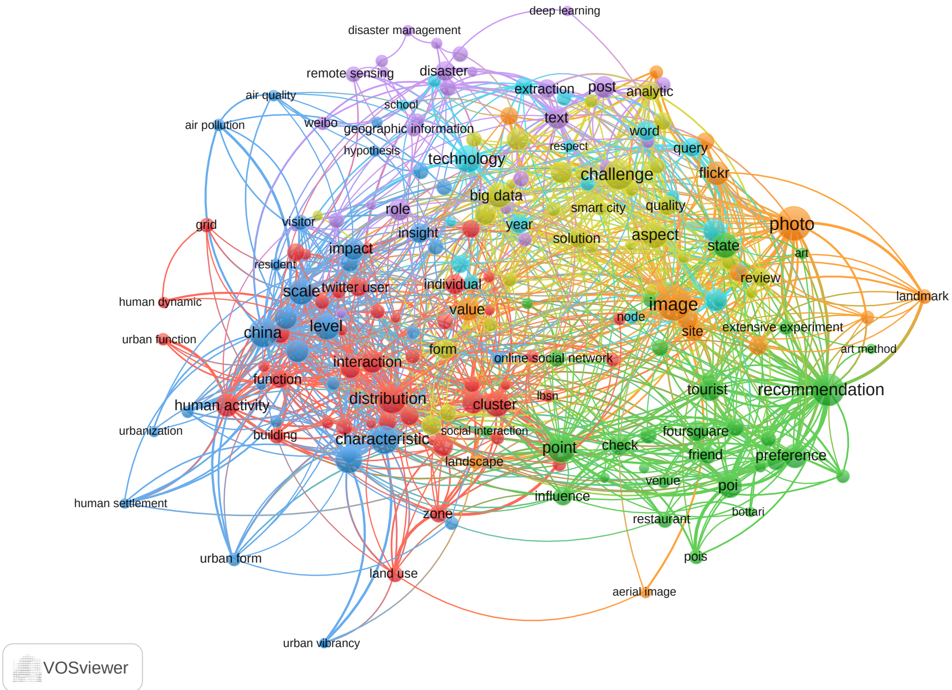


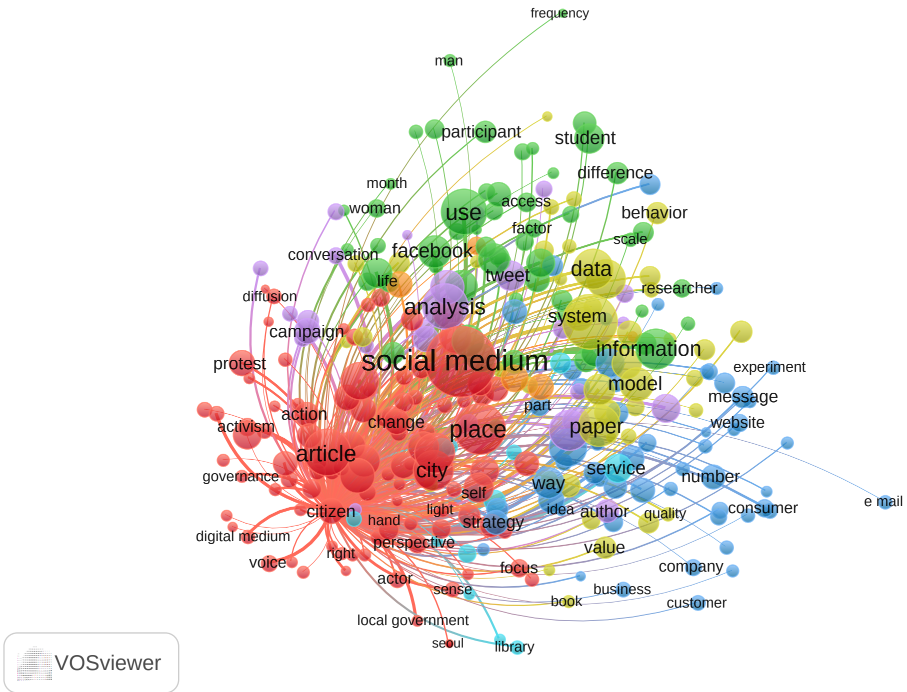
Figure 7.10: Technical Cluster (Cluster 1) Term Map

I first discuss the modified ego network of the term 'citizen' in Clusters 0 (top figure, the Social Cluster) and 1 (bottom figure, the Technical Cluster) depicted by figure 7.11. As section 3.4.8 discussed, these screenshots of modified ego networks only include the edges that originate from the ego-node (in this case, 'citizen'), but the position and color of each noun phrase is the same as in the term co-occurrence maps presented in figures 7.9 and 7.10. Thus, terms' position and color contain information about their relationship within the term maps of Clusters 0 and 1, without visually depicting their edges. In addition, the VOSviewer interface provides some descriptive information about the networks, visible at the bottom panel of each image, including the number of edges in the network ('Links'), the 'Total link strength' (calculated by adding up the edge weight value for each edge or link in the network), and the number of times the selected node (in this case, 'citizen') occurs in the data (in this case, the abstracts and titles of papers in the cluster, depicted by the 'Occurrences' variable).

The ego network of 'citizen' in the Social Cluster (303 links, total link strength 3081, top image) is denser than the ego network of 'citizen' in the Technical Cluster (172 links, total link strength 822, bottom image). The 'citizen' of the Social Cluster highlights diverse forms of participation in cultural practices. The red and purple noun phrase clusters on the top imagine in figure 7.11 highlight citizen's connection to diverse forms of participation in civil society as well as power relations, such as 'activism', 'protest', 'civic participation', 'social movement', 'culture', 'government', 'journalist', 'race' (red cluster) and 'campaign', 'audience', 'election', and the method 'content analysis' (purple cluster). The blue cluster depicts the consumer citizen though noun phrases such as 'consumer', 'customer', 'identity'. The green cluster on the top figure highlights the technology user and learner citizen, with noun phrases like 'learning', 'education', 'student', 'technology', 'YouTube', 'Facebook'. The yellow cluster combines terms related to research methods ('paper', 'study', 'data', 'approach', 'model', 'technique') and two main topic areas: the 'geography' of 'disaster' and 'ideology'.

In contrast, the Technical Cluster - the bottom image in figure 7.11 - mostly depicts the 'citizen' embedded in *senseable* environment - in other words, traceable with computational data and methods - studied with a wide array of computational methods and data. The red cluster on the bottom image highlights the 'data' available about citizens' 'activity' in the 'city', their senseable environment through 'air pollution' and the 'visitor' citizen's experience of 'landscape[s]'. The blue cluster depicts the 'remote sensing' 'social sensing' of 'event[s]', threats to citizens' senseable environment such as 'natural disaster[s]' for 'disaster management'. The green cluster depicts the 'classification' of citizen's senseable visual perspective using 'flickr', 'panoramio', 'photo', in addition to 'twitter'. The yellow cluster in the bottom figure combines noun phrases about citizens studied with diverse methods in and the citizen of the 'smart city'. The purple, light blue and orange clusters highlight more methods to study technology 'use[r]' citizens and their situated practices, using 'prediction[s]' and 'algorithm[s]'.

In sum, the modified ego networks of 'citizen' support my hypothesis, that whilst social geosocial research explores actors' participation in diverse cultural practices, technical geosocial research treats 'citizen' as a social sensor and studies spaces by exploring 'sensed' practices on diverse spatio-temporal scales afforded by computational methods. Whilst these modified ego networks help study differences in the meaning of noun phrases in the two clusters, they do not allow to study the nuances of meanings associated with the ego-nodes: they depict all *alters* (the noun phrases connected to the ego-node), regardless of the strength of these relations. To study the strength of relations between the ego nodes and their alters - which can help study the emphases in meanings they carry in social and geosocial research - I next discuss a version of modified ego networks where edges are filtered based on their weight. I use the VOSviewer user interface to filter edges based on their weight.



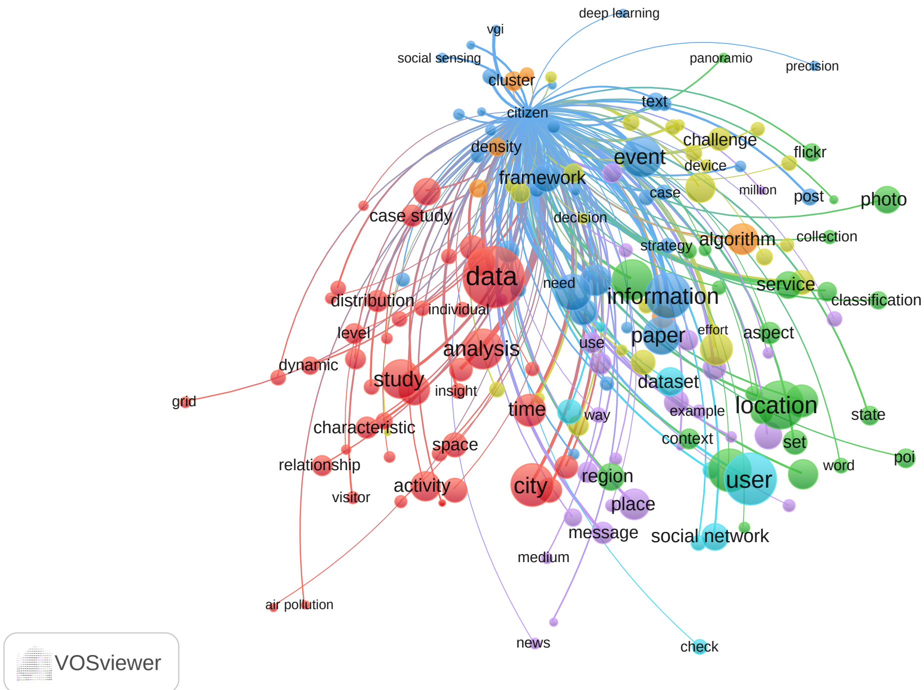
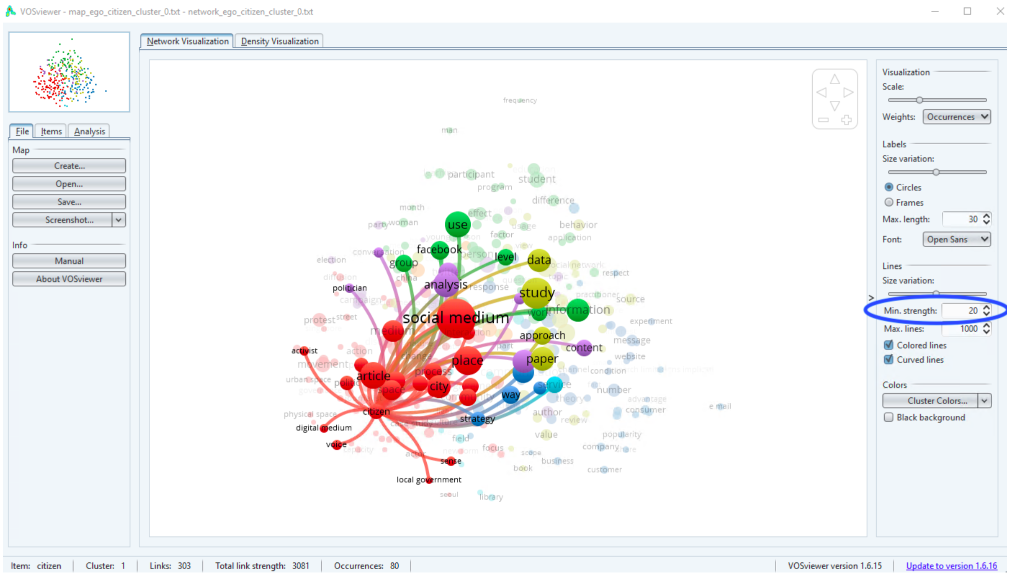


Figure 7.11: 'citizen' ego networks in the Social Cluster (top figure) and Technical Cluster (bottom figure)

Figure 7.12 depicts the modified ego network of 'citizen' showing edges only above a certain edge weight values. The top image in figure 7.12 depicts the modified ego network of 'citizen' in the Social Cluster, highlighting noun phrases which are connected to 'citizen' with edge weight 20 or above. I highlighted the edge weight filter function with a blue oval on the right side of the image. The bottom image in figure 7.12 depicts the modified ego network of 'citizen' in the Technical Cluster, highlighting noun phrases which are connected to 'citizen' with edge weight 12 or above. Filtering edges based on weight helps highlight the noun phrases most strongly connected to 'citizen', and study meaning of citizen in the two clusters in more detail.

Filtering the Social Cluster based on edge weight depicts a citizen similar to the earlier analysis. However, the filtering shifts the emphasis to the citizen who participates in civic life and in changing power relations, depicted by the red and purple clusters. The terms 'activist', 'voice', 'local government' and 'politician' remain connected to 'citizen' in the filtered map. Filtering the Technical Cluster based on edge weight further emphasizes the observed, sensed 'citizen' and their data provider role to understand them and their 'city'. Noun phrases most strongly connected to 'citizen' include 'sensor', 'vgi' which stands for volunteered geographic information, 'event', 'data', 'analysis', 'time', 'event', 'service', 'message' and 'content'.



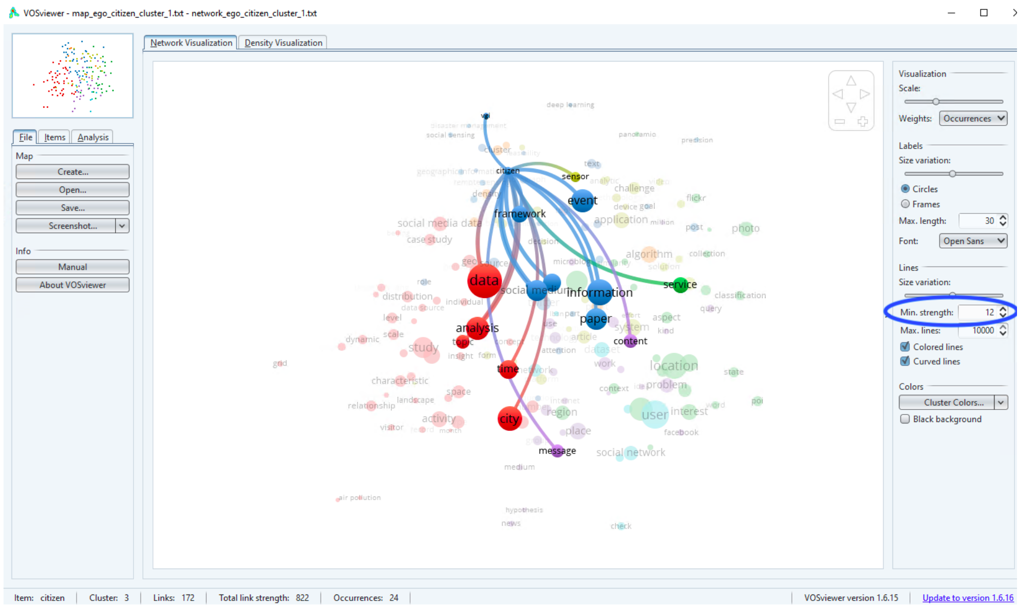
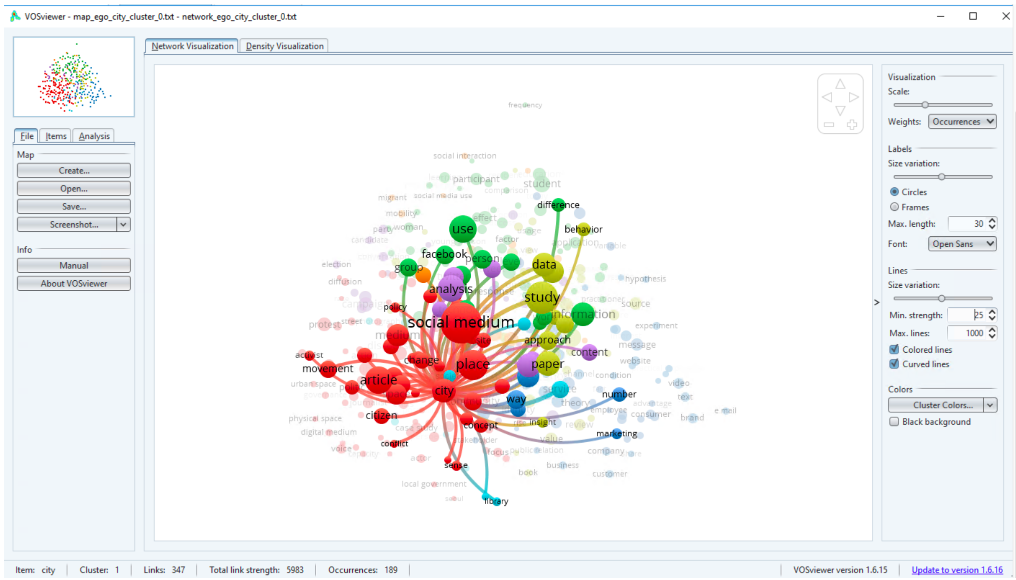


Figure 7.12: 'Citizen' ego networks in the Social Cluster (top figure) edge weight equal to or over 20, and the Technical Cluster (bottom figure), edge weight equal to or over 12

Next, I discuss the modified ego networks of the term 'city'. As the figure 7.13 shows, the *weighted* ego network of 'city' is denser in the Technical Cluster (the bottom image) than the Social Cluster (top image). The themes highlighted by the 'city' modified ego networks are similar to the themes highlighted by the 'citizen' modified ego networks. The Social Cluster depicts the 'city' as a space constituted diverse forms of participation, including civic and commercial practices. In contrast, the Technical Cluster emphasizes the sensed city. In addition, the Technical Cluster emphasizes the city in relation to practices conceptualised as spatio-temporal patterns, through notions like 'human activity', 'human dynamic', 'land use' as well as digitally mediated activities such as 'recommendation'. Both the Social and Technical Clusters of 'city' discuss the built environment, though differently. The Social Cluster mentions 'urban space' and 'physical space' in the red cluster which focuses on civic participation. The Technical Cluster mentions the notions 'urban form', 'human settlement' and 'urban planning' in the red cluster which focuses on "senseable" natural and built cityscape. I observed a similar difference observed for the terms 'space' and 'space' whose visualisations I omit due to space limitations.

Using the filtered ego-network method, figure 7.14 depicts the modified ego networks of 'network'. 'Network' is more strongly connected to noun phrases in the Technical Cluster (bottom image, total link strength 3712) than in the Social Cluster (top image, total link strength 2722), but it is connected to more terms in the Social Cluster (top image, 315 links) than in the Technical Cluster (bottom image, 293 links). Thus, the filtered modified ego network of 'network' is denser for the Technical Cluster compared to the Social Cluster. While the Social Cluster highlights networks of civic and intercultural participation as its connections to the terms 'migrant', 'debate', 'protest', 'movement' show, the Technical 'network' emphasises the 'activity' of 'human[s]' and user[s]' studied though diverse data ('foursquare', 'aerial image') and 'algorithm[s]'.

Altogether, the modified ego networks of 'citizen', 'city' and 'network' show that research in the Social Cluster predominantly frames them in relation to diverse forms of civic, commercial, intercultural participation, whereas research in the Technical Cluster frames them in relation to computational methods that allow sensing the natural and built environments socio-temporal events.



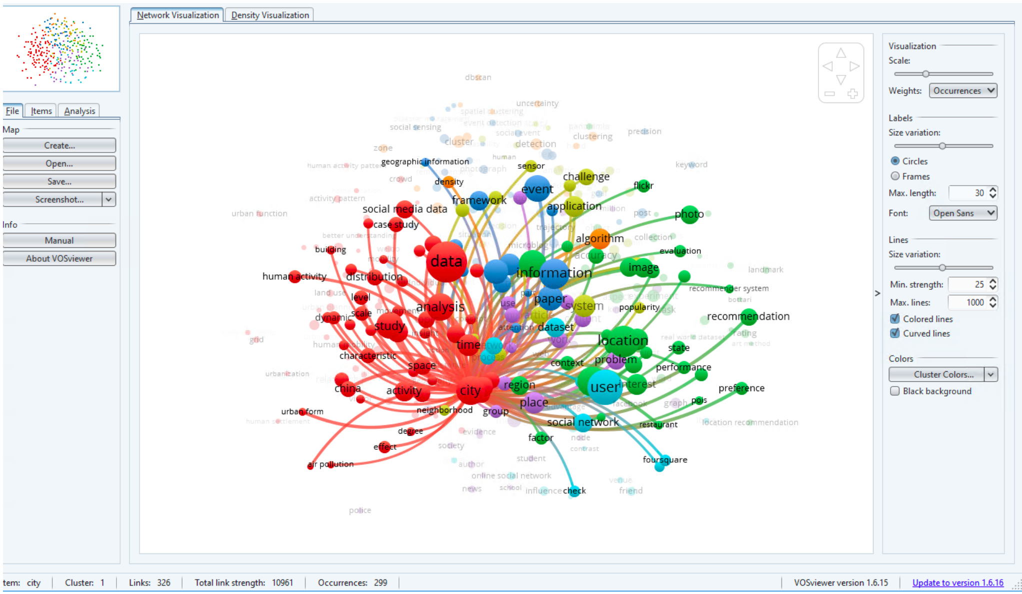
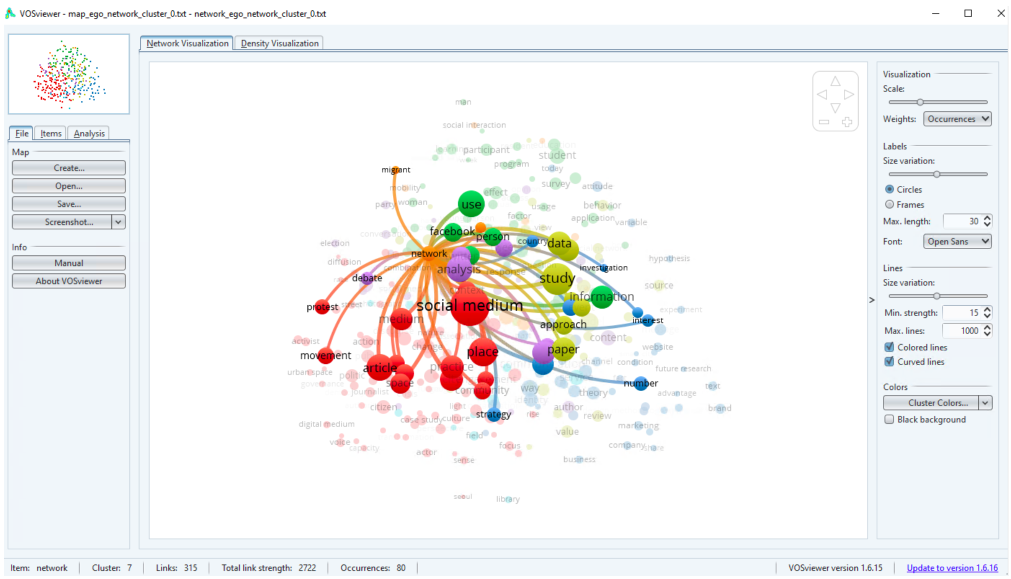


Figure 7.13: 'City' ego networks in the Social Cluster (top figure) edge weight equal to or over 25, and the Technical Cluster (bottom figure), edge weight equal to or over 25



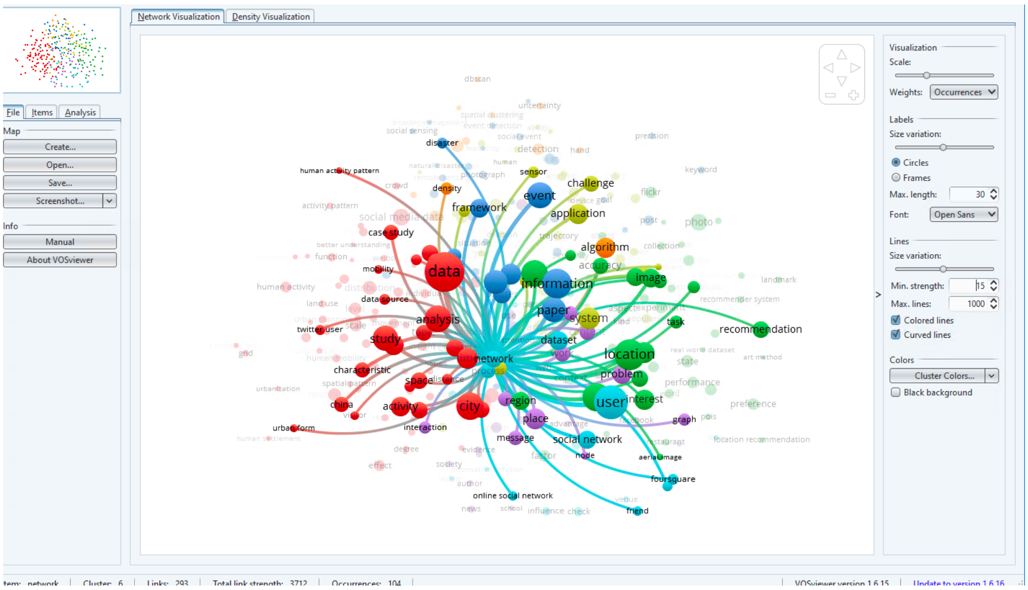


Figure 7.14: 'Network' ego networks in the Social Cluster (top figure) edge weight equal to or over 15, and the Technical Cluster (bottom figure), edge weight equal to or over 15

This section identified geographic geosocial research and compared social and technical geosocial research using computational STS methods. I argued that these approaches differ in terms of the method they use and types of spaces or situated practices they study.

7.3 Conclusion

This chapter focused on exploring differences between the social and technical approaches to geosocial research using mixed methods (explored by the Second Research Question), but addressed all three research questions. Below I discuss my findings with respect to each research question.

In response to the First Research Question, based on interviews, I argued that reflecting on how analytical choices and social media platforms shape geosocial data and the knowledge interviewees create about spaces with them is essential for participants to develop their geosocial research approaches. I argued that reflecting on geosocial data's characteristics is essential because geosocial traces were not created for scientific research and scholars cannot shape their characteristics (cf. Sloan and Morgan, 2015). Reflecting on how methods shape findings may also form part of disciplinary norms to account for how knowledge creation (cf. Strathern, 2004) and necessitated by the multiple analytical options computational data analysis affords (cf. Mackenzie and McNally, 2013).

This chapter addressed the Second Research Question, which explores how approaches to geosocial research differ, by identifying and comparing approaches using mixed methods. Firstly, in addition to the social and computational geosocial research approaches Chapter Six discussed, this chapter identified a third approach: geographic geosocial research. I identified this approach inductively using scientometrics, and reflected on interviews in light of the scientometric finding. Interviewees' quotes support the scientometric separation between geographic, social and computational geosocial research approaches. Several participants noted that the geographic aspect of their research is under-developed, and that they do not use geography methods. I compared the scientometric traces of the geographic approach to those of the social and technical approaches through combining TF-IDF and thematic analyses. Terms associated with geographic methods and its focus on diverse types of spaces including specific locations, types of environments and spatial patterns differentiated geographic geosocial research.

Secondly, this chapter explored differences between social and technical geosocial research approaches using interviews and three computational methods: TF-IDF and term frequency analyses (categorised thematically) and analysing the modified ego-networks of three terms used both by social and technical approaches: 'city', 'citizen' and 'network'. Altogether, I argued that a relative focus on diverse actors' situated practices, experiences and their changes over time - such as changing power relations - differentiates social geosocial research from technical geosocial research. On the other hand, its relative focus on using computational data analysis methods to explore spaces at diverse spatial and temporal scales and in terms of calculable units differentiates technical geosocial research from social approaches. Next, I discuss interview and computational STS findings in more detail.

Through interviews, I differentiated between hermeneutic and algorithmic reflexivity associated with social and technical geosocial research respectively. I argued that hermeneutic reflexivity - associated with social geosocial research - mainly assesses the impact of social media platforms and analytical decisions on geosocial data and the knowledge they afford about spaces in experiential and historic terms. Social scientist participants reflect on how social media users' experiences shape geosocial data, and how their own experiences and the history of data infrastructures shape their methods and findings. They treat these experiential and historic practices constitutive part of their findings. I refer to this as hermeneutic reflexivity, drawing on Fortun et al.'s (2016) notion of hermeneutic expertise. Finally, many social scientist interviewees stated that their geosocial research differs from technical geosocial research in terms of their reflexivity. As Chapter Six also argued, STS scholars have demonstrated that the introduction of new technologies is often associated with the re-negotiation of professional identities and epistemic authority. Similar to the radiologists Burri (2008) studied, my interviewees re-negotiate their authority to make scholarly claims using geotagged social media data in relation to research produced by other scientific disciplines, by claiming the epistemic values associated with hermeneutic reflexivity.

In contrast, through interviews I found that participants with technical backgrounds tend to reflect on their data, analytical decisions and findings about spaces in terms of the calculations that enable them, and in demographic terms. Furthermore, in contrast to interviewees from groups A-D, technical scholars quoted in this section do not frame reflexivity as a distinct feature of their research. Similar to Neff et al. (2017), I found that for technical geosocial scholars, reflecting about the origins, attributes and potential uses of data and their own analytical choices are necessary and often unacknowledged aspects of their research. As Isaac's quote illustrates, technical scholars can even feel incentivised to omit such reflexivity from their published work to reduce the opportunities for reviewers to question their decisions. Algorithmic reflexivity also contrasts with hermeneutic reflexivity in that scholars whose reflexivity I labeled 'algorithmic' often aim to 'control for' demographic or the impact of their analytical choices. They often argue they produce valid findings because they 'account for' potential 'biases'.

Thus, in line with Strathern (2004), I found that disciplines differ in how they account for - or reflect on - how they create knowledge. Reflexive acts that comprise hermeneutic reflexivity, such as reflecting on situated practices and knowledge creation in terms of lived experience and history, are considered central epistemic virtues in the disciplines my interviewees who practice such reflexivity are trained in, including anthropology, 'interpretive' sociology and human geography. Discussing assumptions of quantitative calculations and reflecting on the benefits and weaknesses of specific calculations - as with algorithmic reflexivity - are key to computational and mathematical disciplines (e.g. Fortunato, 2010; Freedman, 2010). Unlike Borges Rey (2017) who found that few of the digital journalists he interviewed reflected concurrently on the affordances of computational data analysis tools and their subject matter - news stories - the majority of my interviewees reflect on how both computational data analysis infrastructures and their analytical decisions shape the subject matter - the knowledge they create about spaces.

Using computational STS methods, I compared social, technical and geographic geosocial research. The computational findings resonate with the above interview findings. Similar to the interview narratives about reflexivity, computational STS analyses highlight social geosocial research's relative emphasis on studying spaces in relation to actors' practices and forms of participation. In contrast - similar to algorithmic reflexivity which foregrounds calculations methods and units - I find that the frequent use of concepts related to computational data analysis methods, technology users (e.g. 'similar user', 'human') and 'senseable' and calculable spatial scales differentiate technical from social geosocial research. The scientometric differences I found also resonate with Meeteren et al. (2016) who argue that approaches to policentricity research differ in terms of the methods they use and spatial units they study. Firstly, while its focus on specific locations differentiates social geosocial research, its focus on spatial units which can be expressed in terms of spatio-temporal scales differentiate technical geosocial research. Secondly, technical approaches' comparative focus on computational methods and the geographic methods' comparative focus on geographic data analysis methods signal methodological differences.

To explore Research Question 3.1, this chapter combined methods to identify and compare geosocial research approaches in three main ways. As outlined above, I identified approaches to geosocial research with scientometrics inductively, highlighting a third approach - geographic geosocial research - and reflected on interview quotes in light of this scientometric finding. I also compared approaches by thematically analysing noun phrases ordered according to TF-IDF and word frequency values. TF-IDF helped me emphasize the difference between social, technical and geographic geosocial research, and the word frequency analysis helped me identify terms both approaches use. However, these methods take noun-phrases out of context. Thus, finally, I visually analysed the modified ego networks of three noun phrases relatively frequently mentioned by both social and technical geosocial papers - 'city', 'citizen' and 'network'. This allowed me to explore the different ways social and technical approaches frame spaces and actors and use methods.

In addition, to visually illustrate the construction of my scientometric analyses and my findings' reliance on (digital) infrastructure (cf. D'Ignazio and Klein, 2020; Cambrosio, Bourret, et al., 2014), I used screenshots of the interactive user interface. When used 'as intended' (such as exporting visualisations), data analysis software remove the traces of most analytical decisions.

In response to Research Question 3.2 - which explores how we can assess the affordances of computational methods 'for STS' - next, I discuss how the interpretative context informed each computational method used in this Chapter.

The interpretative context suggested that I can identify the scientometric traces of geosocial research approaches through clustering author-bibliographic coupling networks for four main reasons. Firstly, in Chapters Four through Six, using mixed methods I argued that different approaches to geosocial research exist, and can be traced through author-bibliographic citation relations. The latter was suggested by Chapter

Six's citation network analysis which showed the separation between the 'only social' and 'only computational' geosocial papers. Secondly, I conceptually argued that approaches can only be differentiated with respect or in comparison to each other. Network clustering suits this purpose because it identifies sets of nodes that are relatively strongly connected to each other in comparison to their connections to rest of the network.

Thirdly, network clustering allowed me to identify sets of papers that correspond to approaches bottom up, rather than pre-defining them based on disciplinary categories as in Chapter Six. This is important for two reasons. On the one hand, 'only social' and 'only computational' journals do not perfectly correspond to 'social' and 'technical' approaches. Although interviewees noted they *mainly* published their geosocial research in journals associated with their disciplines, they also published in journals that I categorised as 'interdisciplinary or multidisciplinary' left out from the above analysis. In addition, as noted earlier, 'social' and 'technical' geosocial research differentiate through relations between the two. Thus, I assume that the boundaries of the scientometric of footprint of these approaches do not fully correspond to 'only social' and 'only computational' paper sets. In addition, network clustering determines the number of geosocial approaches bottom up, which is beneficial because I do not know *a priori* how many approaches to geosocial research I can differentiate. As section 9.3.2 will discuss, this bottom up approach opens opportunities for surprises.

Fourthly, the interpretative context helped interpret the results of the network clustering. The clustering of both networks yielded clusters that I could assume captured the social and technical approaches identified earlier. In addition, I found interview evidence that the geosocial cluster identified scientometrically can be considered a separate approach.

The interpretative context also informed my use of the disciplinary distribution of papers in the clusters of *G3* and *G4* to identify geosocial research approached in two ways. Firstly, similar to the network clustering, it assumed that approaches can be identified and characterised in comparison, with respect to one another. This analysis compared the distribution of papers between clusters across the same Broad Disciplinary Categories. Secondly, as outlined above, previous results discussed in Chapter Six suggested that the disciplinary categorisation of geosocial papers signals epistemic differences among them because scholars tend to publish in disciplinary journals for research evaluation purposes and to belong to research communities.

The interpretative context informed my use of TF-IDF in three main ways. Firstly, it required understanding these sets of papers as footprints of geosocial research approaches as outlined above. Secondly, as section 3.4.6 explained, ranking terms in the abstract and title of geosocial papers using TF-IDF - which, for each cluster, highlights terms that are relatively frequent within them and relatively infrequent in other clusters - helped explore comparative differences among approaches. Finally, I interpreted the collection of terms highlighted by the TF-IDF analysis using thematic analysis informed by my conceptual framework.

Finally, the modified ego-network analyses hinged on the interpretative context in three main ways. As above, I assumed that the set of papers in the 'social' and 'technical' clusters of the author bibliographic coupling network *G4* - which formed a basis for this analysis - correspond to the scientometric footprints of geosocial research approaches. In addition, interviews and previous scientometrics suggested similarities among geosocial research approaches. I assumed that shared concepts partially reflected these similarities. Finally, the modified ego-networks allowed me to understand differences in the meaning of these terms through depicting their co-occurrence with other terms. The networks helped explore how the terms are used and framed in context.

Chapter Eight will further compare approaches to geosocial research by altering the units identified and compared by scientometric methods. I will illustrate the diverse uses of computational methods in more detail, approaches' shared interest in studying specific locations and interviewees' use of local knowledge when studying specific locations. In addition, it will explore how research methods mediate knowledge about spaces using heterogeneous network analyses. I will argue that the spatial units brought forth using machine learning and social network analyses differ.

Chapter 8

Exploring Methods' Mediation

This chapter explores how research methods mediate knowledge about spaces and my findings about geosocial research approaches, contributing to all three research questions. Firstly, section 8.1 explores the Second and Third Research Questions by diversifying the scientometric analyses of geosocial research approaches. It compares findings about geosocial research approaches by clustering the author-bibliographic coupling network *G3* (comprising all geosocial papers) and a term map of geosocial papers with findings discussed in Chapter Seven through clustering *G4* (comprising 'only social' and 'only computational' geosocial papers). These analyses help illustrate the heterogeneity of geosocial research approaches and similarities among them.

Secondly, in relation to the First and Third Research Questions, section 8.2 explores how two methods - machine learning (ML) mostly affiliated with technical geosocial research, and social network analysis (SNA), primarily associated with social geosocial research - mediate geosocial research. I argue that both ML and SNA enable a new set of scholars to conduct geosocial research, but scholars who use ML position their geosocial research emphasizing its affordances to study diverse types of spaces. In addition, heterogeneous network analysis shifts my analytical perspective from methods' mediation and instead highlights spatial units' mediation.

Thirdly, to address the First Research Question, section 8.3 explores how interviewees who study specific locations - of interest to geosocial scholars from diverse disciplines - use local knowledge. It highlights four main roles local knowledge plays: interpret computational findings (more quickly); validate analytical decisions and assess data quality; motivate scholars or slow down the data analysis process.

8.1 Geosocial Research Approaches Further Compared

This section uses two scientometric analyses to diversify the way I differentiate among approaches to geosocial research, allowing me to explore similarities and differences among them in more detail: a citation network analysis and a term co-occurrence network analysis. As section 2.3 argued, the comparative differences I identify among geosocial research approaches depend on the analytical units. The two analyses presented below help diversify the analytical units I use to compare approaches.

Firstly, I compare the cluster analysis of the author-bibliographic coupling network of all geosocial papers *(G3)* - depicted by figure 7.3, reproduced in figure 8.1 below - with the analysis of network *G4* discussed in section 7.2. Like my analysis of *G4*, I analyse the clusters of *G3* based on the disciplinary distribution of papers therein and the TF-IDF method outlined in section 3.4.6. Figures 8.2 - 8.5 depict the TF-IDF analysis of the clusters of *G3*. As section 2.6.3.3 discussed, community detection is non-deterministic and is contingent on the network used for the analysis. Thus, analysing the clusters of *G3* allows me to alter the analytical units used for comparison. A detailed analysis of the clusters of *G3* is beyond the scope of this thesis. Instead, I highlight one similarity and four main differences compared to the analysis of *G4* discussed in Chapter Seven.

Like the clusters of *G4*, the clusters of *G3* include a Social (Cluster 0), Technical (Cluster 3) and Geography (Cluster 1) Cluster. The existence of these three clusters in both networks *G3* and *G4* supports my argument to treat them as distinct approaches. In addition, there is a fourth cluster containing more than two papers - Cluster 2 - which comprises a relatively high percentage of papers published in journals with Broad Disciplinary Categories 'interdisciplinary or multidisciplinary' and 'only computational'. I refer to this as the Mixed Cluster hereafter, to stress the preponderance of 'interdisciplinary' papers, but avoid to use the word 'interdisciplinary' which would suggest a that this cluster, unlike others, is interdisciplinary.

Comparing the clusters of *G3* highlighted three further differences with respect to the analysis of *G4* discussed in section 7.2. Firstly, the clusters of *G3* better highlight the diversity in the use of computational data analysis for geosocial research. The Technical Cluster of *G3* - containing less than 400 papers - is smaller than that of *G4*, containing around 600 papers. At the same time, The Geography Cluster of *G3* contains less 'only social' and more 'only computational' papers than that of *G4*, and 'only computational' papers are also present in the Mixed Cluster. The distribution of 'only computational' papers across the clusters, as well as the increase in computational methods-related terms that differentiate the Geography and Mixed Clusters suggest shared focus on computational method use or development among the Technical, Geography and Mixed Clusters. Secondly, the presence of 'urban research' related noun phrases in those with high TF-IDF values in all three clusters highlight their shared urban research focus. Thirdly, noun phrases which refer to specific locations appeared among the terms that differentiate all approaches identified using *G4*. This highlights the approaches' shared interest in studying these types of spaces. Altogether, comparing the clusters of *G3* illustrated how knowledge about approaches to geosocial research is contingent on the units of comparison. Next, I discuss difference among approaches to geosocial research through the lens of noun-phrase co-occurrence network analysis.

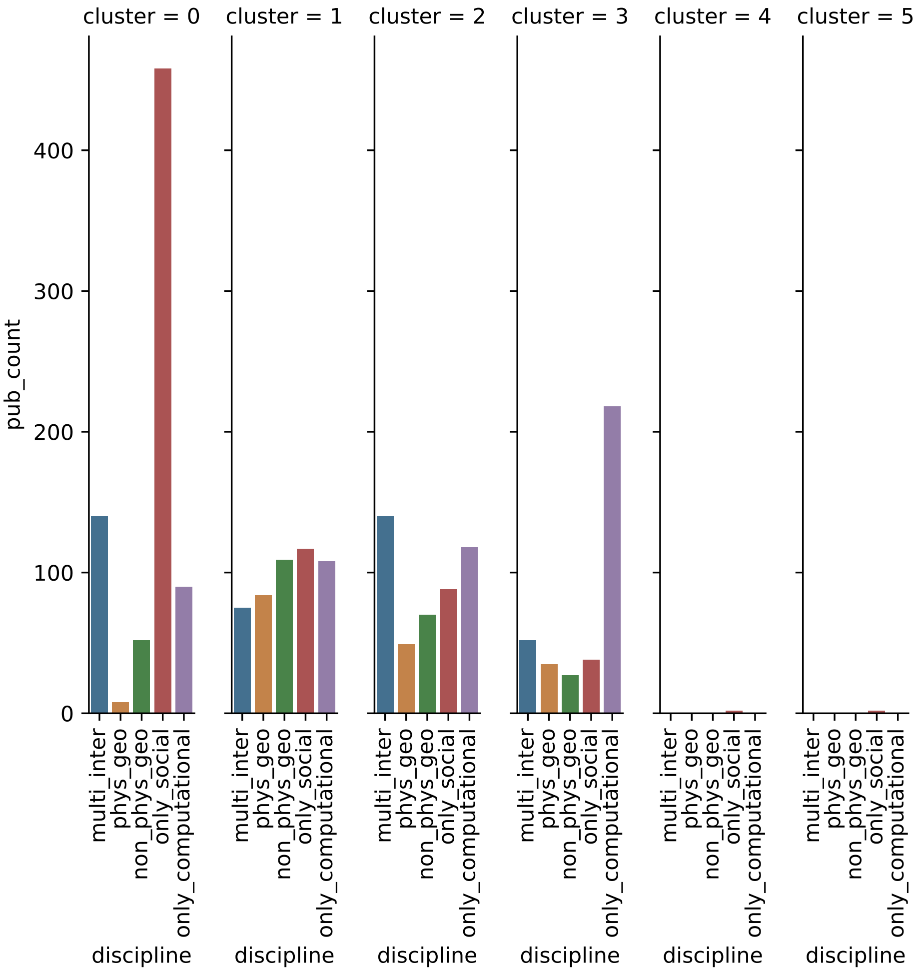


Figure 8.1: Distribution of subject categories across the clusters

'SOCIAL CLUSTER' - CLUSTER 0	
topic / practice	urology, spanish, arab spring, reproductive health, money, waterpipe, information sharing, psychological distance, public attention, plastic surgeon, blood, social norm, fidelity, leisure, brand post, condom, entertainment, religion, sexual health, coordination, sale, social distance, daily life, brand equity, cyberbullying, public engagement, social media usage, emotional support, alcohol post, ebola, sustainable development, adolescence, political participation, health behaviour, hiv risk, edible oil, hypertension, business model, disadvantage, diet, crime, anxiety, racism, hurricane sandy, self presentation, employment, proliferation, recognition, civic engagement, teaching, health condition, sleep, ethics, operation, uptake, wildfire, residence, stigma, street art, diabetes, solidarity, hiv testing, encounter, freedom, public relation, inclusion, reputation, birth, mental health, authenticity, sexual behaviour, empowerment, alcohol use
media	youtube video, online social medium, osns, social media tool, snapchat, facebook advertising, grindr, social media content, mainstream medium, sms, snss, digital platform, pinterest, online medium, radio, social media usage, Airbnb, cyberspace, google, social media technology, tumblr, texting, mobile medium, twitter account, new technology, mobile technology, online platform, tripadvisor,, traditional medium, whatsapp, online review, text messaging, digital age, chat, web site, social network site, media use, email, text message, mass medium, internet use
space	netherlands, real world, new york, seoul, rural area, digital space, brazil, European capital, Egypt, cyberspace, south korea, online space, hong kong, Singapore, Malaysia, high school, nation, south Africa, library, Chicago, Germany, Texas, usa, united kingdom, spain, tourist destination, new Zealand, physical space, neighbourhood, England, classroom
actor	intervention group, coyote, homeless youth, guest, boy, local health department, municipality, celebrity, healthcare professional, pregnant woman, public library, virtual community, infant, young woman, high school, service provider, nation, small business, age group, library, latino, doctor, Germany, general public, clinician, employer, learner, applicant, midwife, african American, health professional, human, college, nurse
method	informed consent, focus group discussion, digital data, altmetric, mean age, random sample, social media content, empirical study, validation, descriptive statistic, control group, social science, theoretical framework, likelihood, reliability, assumption, api, ethics, interpretation, depth interview, baseline, essay, discourse analysis, literature review, cross sectional study, data collection, research question, argument, qualitative analysis
time	month period, timing, June, new opportunity, august, minute, progress, February, digital age, September, transition, December, new form, origin, hour, tendency, april

Figure 8.2: *G3* network - Cluster 0 TF-IDF

'GEOGRAPHY CLUSTER' - CLUSTER 1	
topic / practice	wildlife, situational awareness, public participation, hurricane sandy, disaster management, natural disaster, politics, resilience, conservation, climate change, crisis, business, aesthetic value, sustainability, public, forest
media	social media photograph, geoweb, social media activity, geotagged photo, smartphone, geotagged photograph, geotagged tweet, lbsns, weibo,, social media
space	geoweb, Madagascar, urban form, urban park, study area, Singapore, spatial clustering, hot spot, urban environment, London, urban space, new york city, Australia, proximity, Beijing, united states, smart city, ecosystem, forest
actor	citizen scientist, institution, general public, local government, human, lemur, author, stakeholder
method	geotagged social media data, social sensing, citizen scientist, temporal resolution, new data source, geographic information system, spatial information, volunteered geographic information, spatial clustering, remote sensing, feasibility, citizen science, utility, occurrence, lbsns, usefulness, giscience, discipline, temporal pattern, proxy, spatial analysis, crowdsourcing, identification, efficiency, content analysis, keyword
time	temporal resolution, season, April, today, shift, recent year, temporal pattern

Figure 8.3: *G3* network - Cluster 1 TF-IDF

'INTERDISCIPLINARY CLUSTER' - CLUSTER 2	
topic / practice	influenza, urban dynamic, human behaviour, social activity, friendship, rumor, natural disaster, segregation, information diffusion, human movement, human activity pattern, travel, population density, gender, earthquake, tourism, human mobility pattern, trip, spatial distribution
media	geolocated tweet, recommender system, geotagged tweet, communication technology, mobile device, weibo
space	locality, Shenzhen, shanghai, city level, home location, spain, urban dynamic, cyberspace, spatial scale, natural city, urban environment, united states, spatial interaction, population density , landmark, London, situation, district, street, spatial distribution
actor	traveller, social media user
method	bottari, community detection, foursquare data, edge, magnitude, new approach, reliability, gravity model, assumption, large number, lbsns, experimental result, gender, gis, spatial analysis, keyword, lbsn, poi, hypothesis, inference, estimate, spatial distribution, node, proxy
time	date, hour, real time, recent year, week

Figure 8.4: *G3* network - Cluster 2 TF-IDF

'TECHNICAL CLUSTER' - CLUSTER 3	
topic / practice	crisis management, emergency event, food safety, police, location recommendation, art, public opinion, spatial pattern, urban emergency event, human mobility, urban planning, trip purpose, recommender system, emotion
media	panoramio, sina weibo, social media site, spatial clustering, camera, online social network, gps, lbsn, weibo, microblog, yelp, geotagged photo, mobile device
space	important location, large scale, road segment, spatial pattern, urban emergency event, urban planning, new york city, spatial distribution, Beijing, landmark
actor	human, consumer, respondent, visitor, twitter user
method	deep learning, real world dataset, aerial image, dynamic feature, ugc, recall, latent dirichlet allocation, temporal information, crowdsourcing crowdsourcing, spatial clustering, spatial analysis, pois, spatial information, probability, precision, remote sensing, sentiment analysis, spatial distribution, extensive experiment, location information, confidence, node, complexity, cloud
time	today, recent year, month

Figure 8.5: *G3* network - Cluster 3 TF-IDF

Secondly, I explore differences among approaches by altering the type of scientometric network. Figure 8.6 depicts the co-occurrence network of noun phrases from the abstracts and titles of all geosocial papers using the method discussed in section 3.4.9. VOSviewer's clustering algorithm identified 9 clusters. I do not treat all nine as separate approaches to geosocial research. Rather, I argue that this analysis helps explore similarities and differences among geosocial research endeavours in a more continuous space through depicting their distance in two-dimensional space. In this analysis, noun phrases can only be part of one cluster, but their position contains information about their specificity to geosocial research endeavours, or to the contrary, their popularity across approaches. For example, the green term 'hashtag' - being centrally located - is likely to be rather popular across geosocial papers. In contrast, the terms in red 'poi' (point of interest) and 'dbscan' (a data clustering algorithm), the light blue term 'yoga' and the yellow term 'lemur' are more specific to papers whose abstracts and titles contribute to their respective clusters. I conducted data analysis discussed below using VOSviewer's interactive user interface.

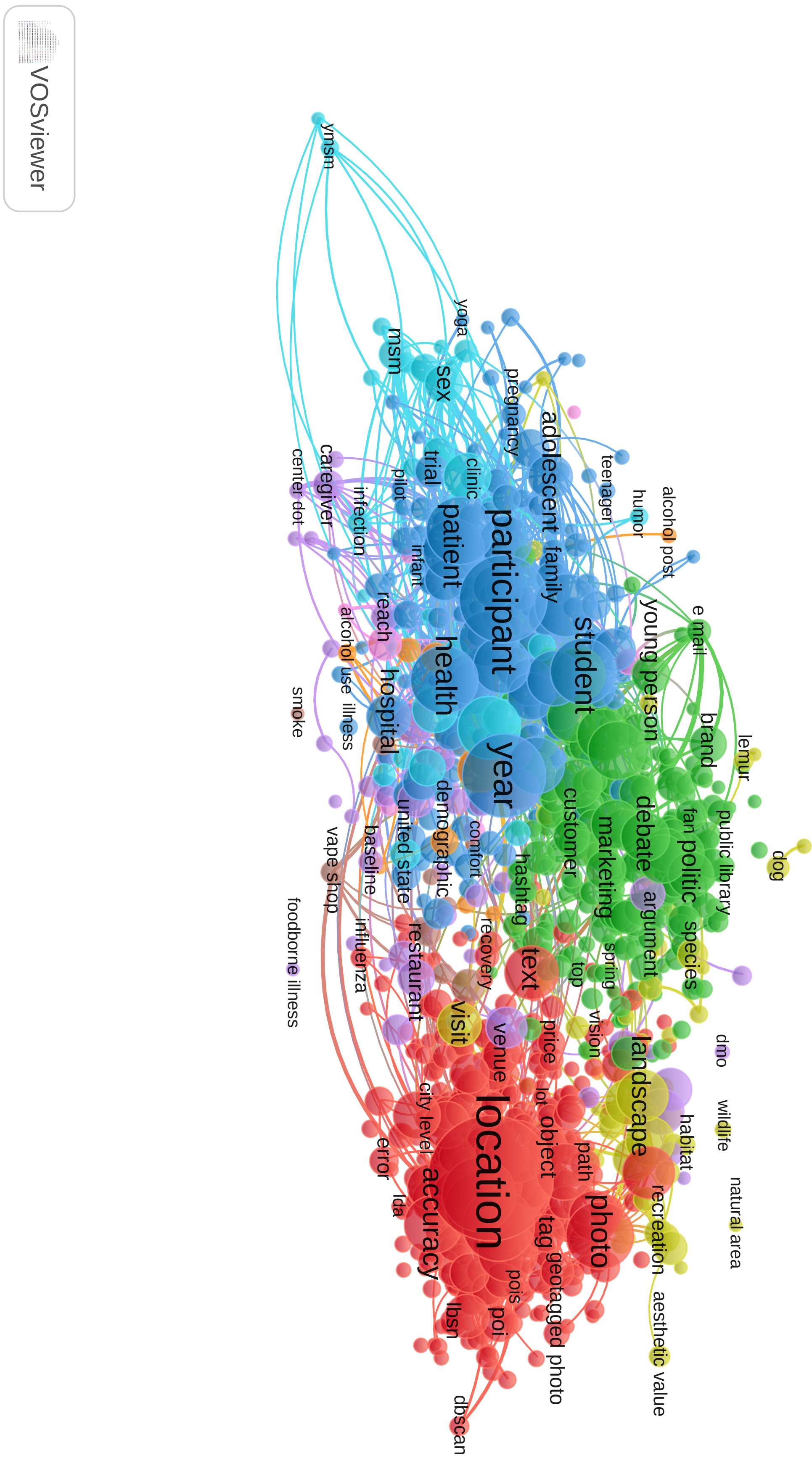


Figure 8.6: Noun phrase co-occurrence map

Figure 8.7 presents a second visualisation of the noun phrase co-occurrence network: it depicts the density of nodes (noun phrases) across the network. It shows a dense cluster on the right side of the map, depicted in more detail by figure 8.8 where the term 'location' and computational methods related terms - such as 'distribution', 'algorithm', 'prediction', 'spatial analysis', 'geographic information', 'data mining', 'new method' and 'machine learning' - are prominent. This corresponds to the red cluster on figure 8.6. This dense area is separated by smaller, denser clusters above and to the left of it. Closest is a dense cluster above it, which corresponds to the yellow cluster on map 8.6, focusing on ecosystems and landscape (aesthetics) research, as figure 8.9 depicts. Given the terms focused on computational methods and ecosystems - in addition to those related to 'crisis management' and urban research (e.g. 'San Francisco' and 'segregation', 'city scale') depicted by figures 8.8 and 8.9, I assume that the red and yellow clusters on figure 8.6 - the dense areas on the right side of figure 8.7 - mainly correspond to the Technical, Geography and Mixed Clusters discussed above.

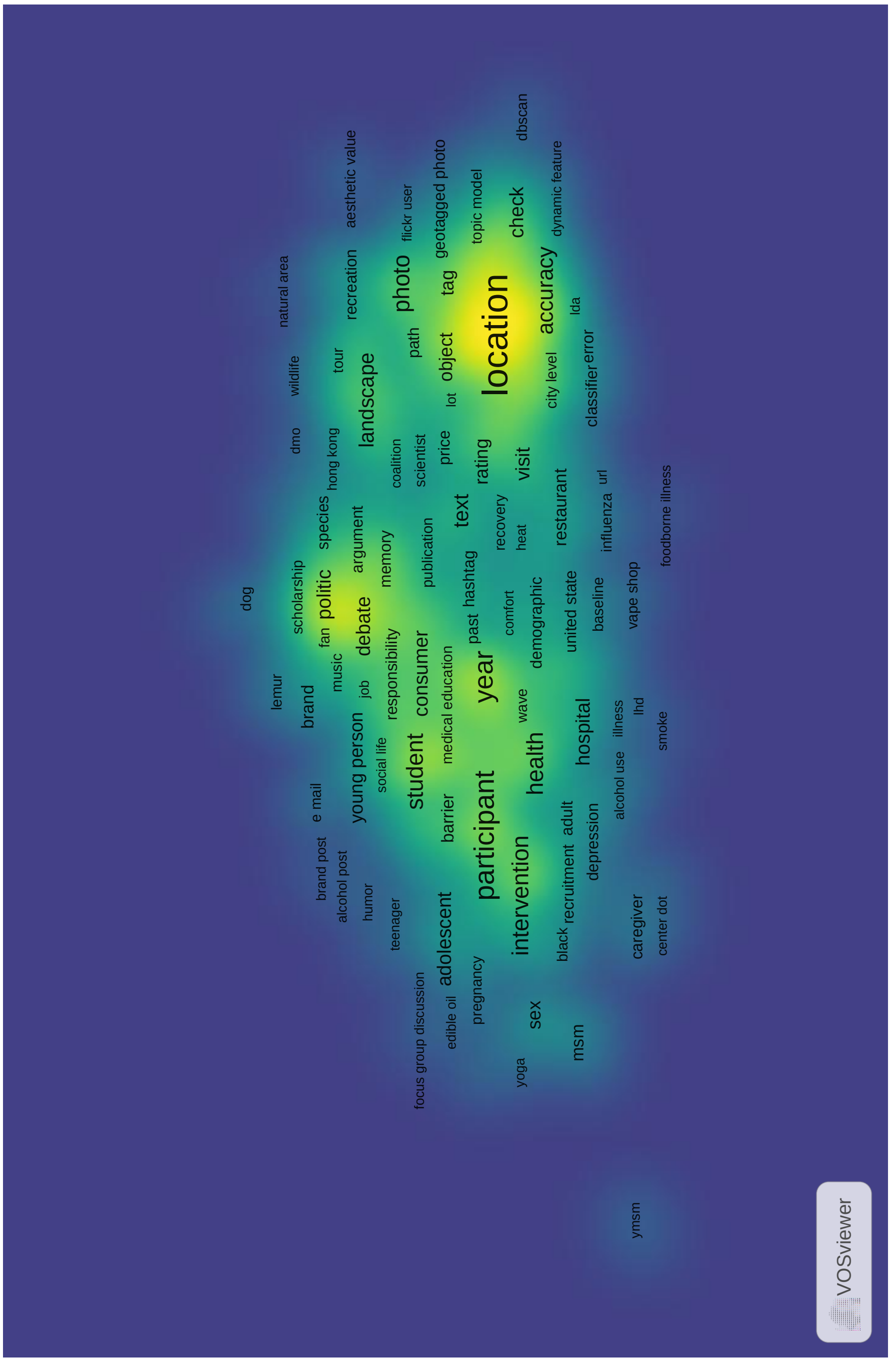


Figure 8.7: Noun phrase co-occurrence density map

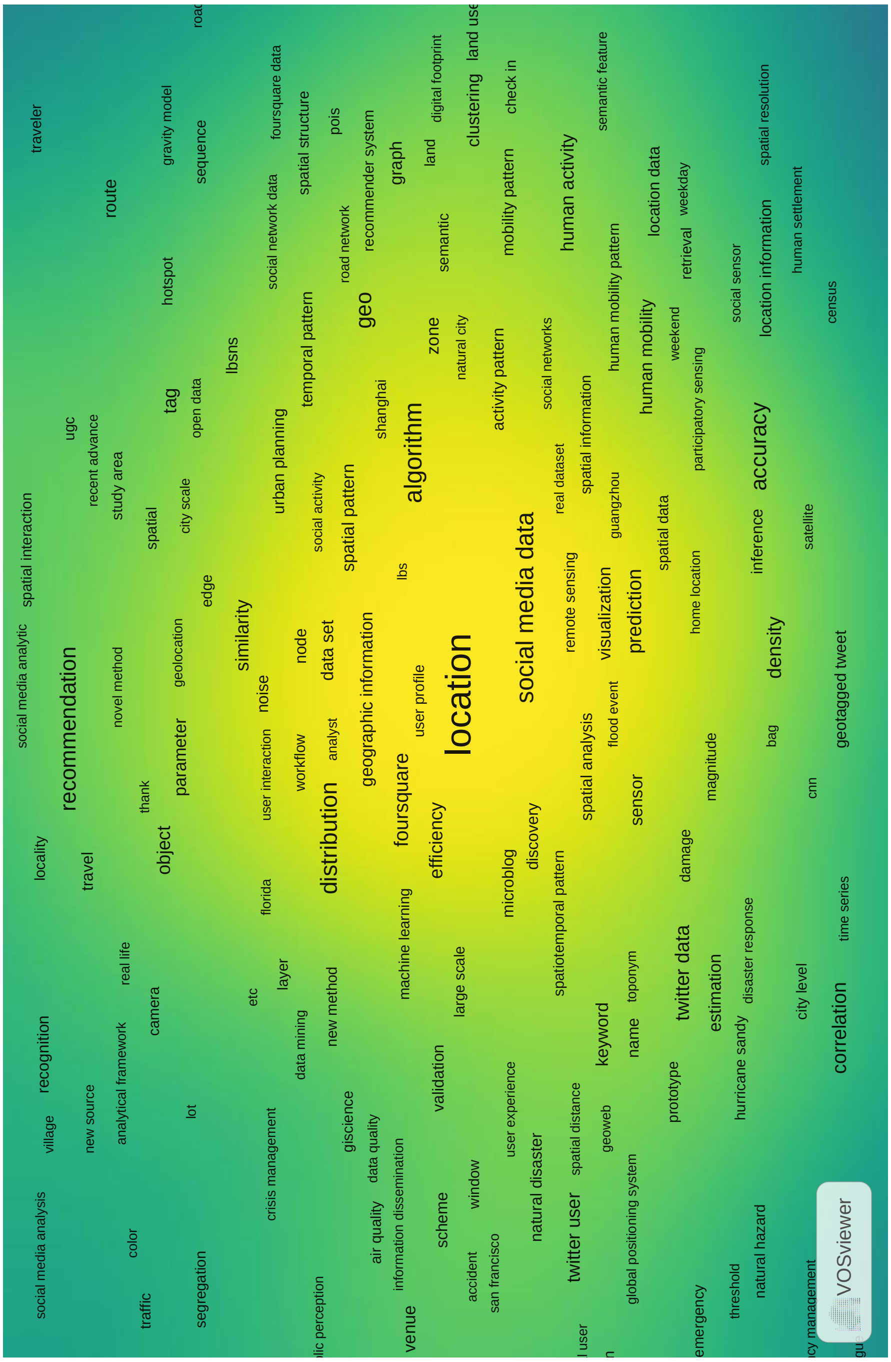


Figure 8.8: Methods focused area - noun phrase co-occurrence density map

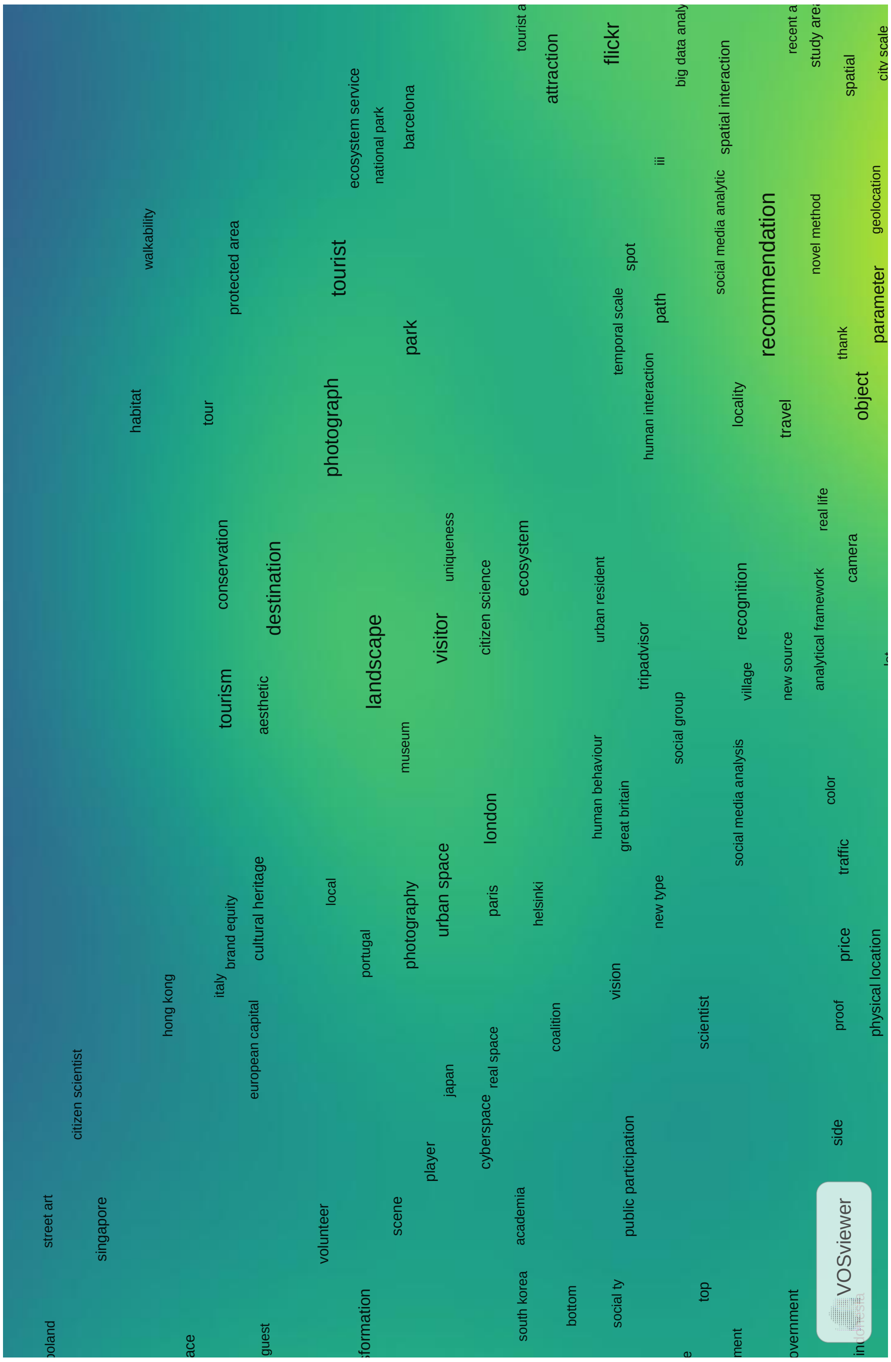


Figure 8.9: Tourism focused area - noun phrase co-occurrence density map

In addition, the dense area on the right of figure 8.7 is separated by four denser areas on the left. Figures 8.10 and 8.11 depict the latter in more detail. As these figures and figure 8.6 show, the left side (or 'top') of the map depicted by figure 8.6 is dominated by terms which refer to situated practices, related to public opinion formation, commercial activities, education and health. These resonate with the Social Clusters of *G3*. I argue that the gap between the left and right side of the map depicted by figure 8.7 highlights a methodological difference among approaches to geosocial research. I argue that the further clusters of terms are from the dense cluster on the right, the less connected research is to computational data analysis methods therein. Thus, I argue that the left side of the map in figures 8.6 and 8.7 depict social geosocial research project's varied distance from computational data analysis methods.

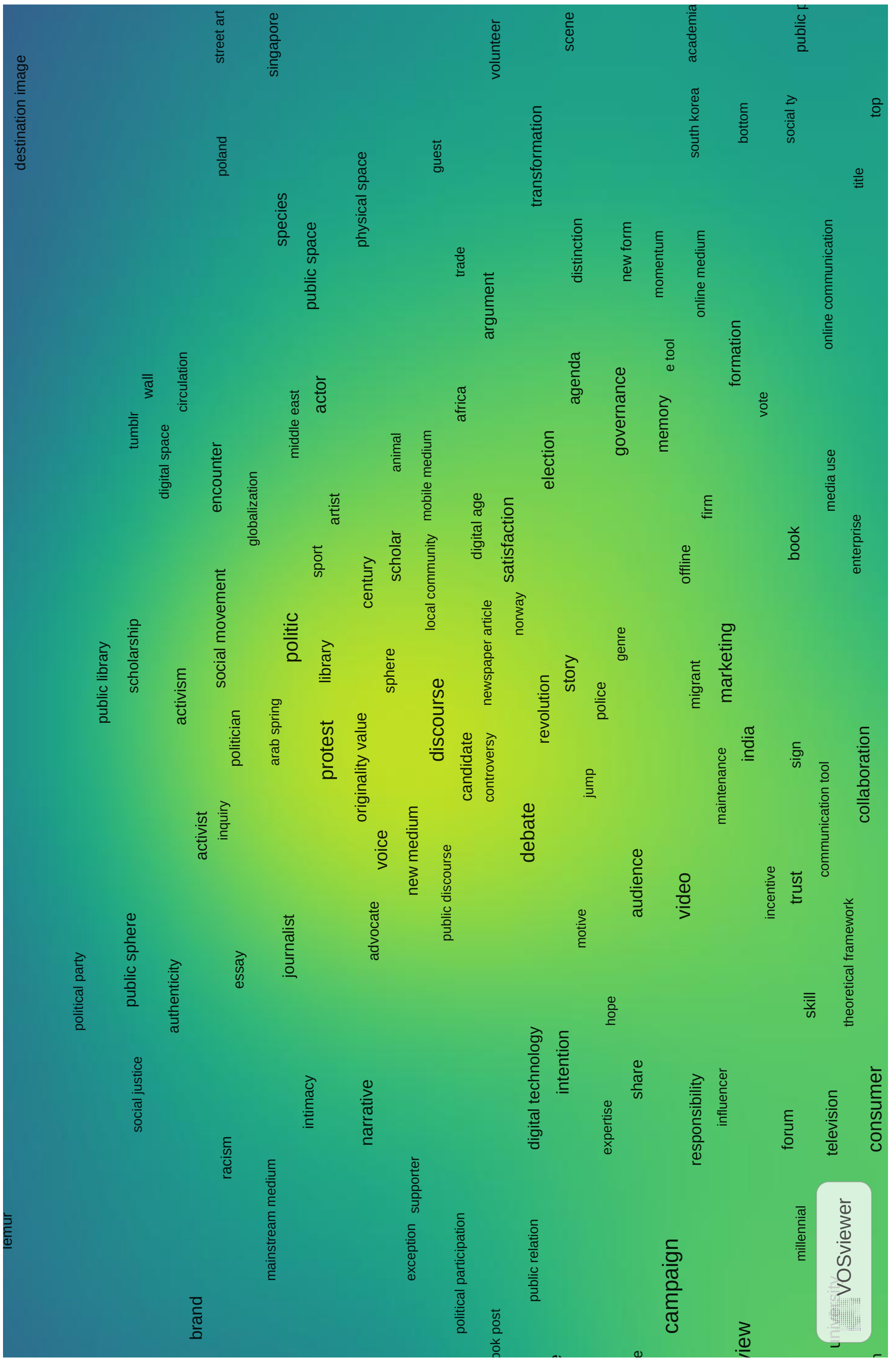


Figure 8.10: Social geosocial research, power relations focused area - noun phrase co-occurrence density map

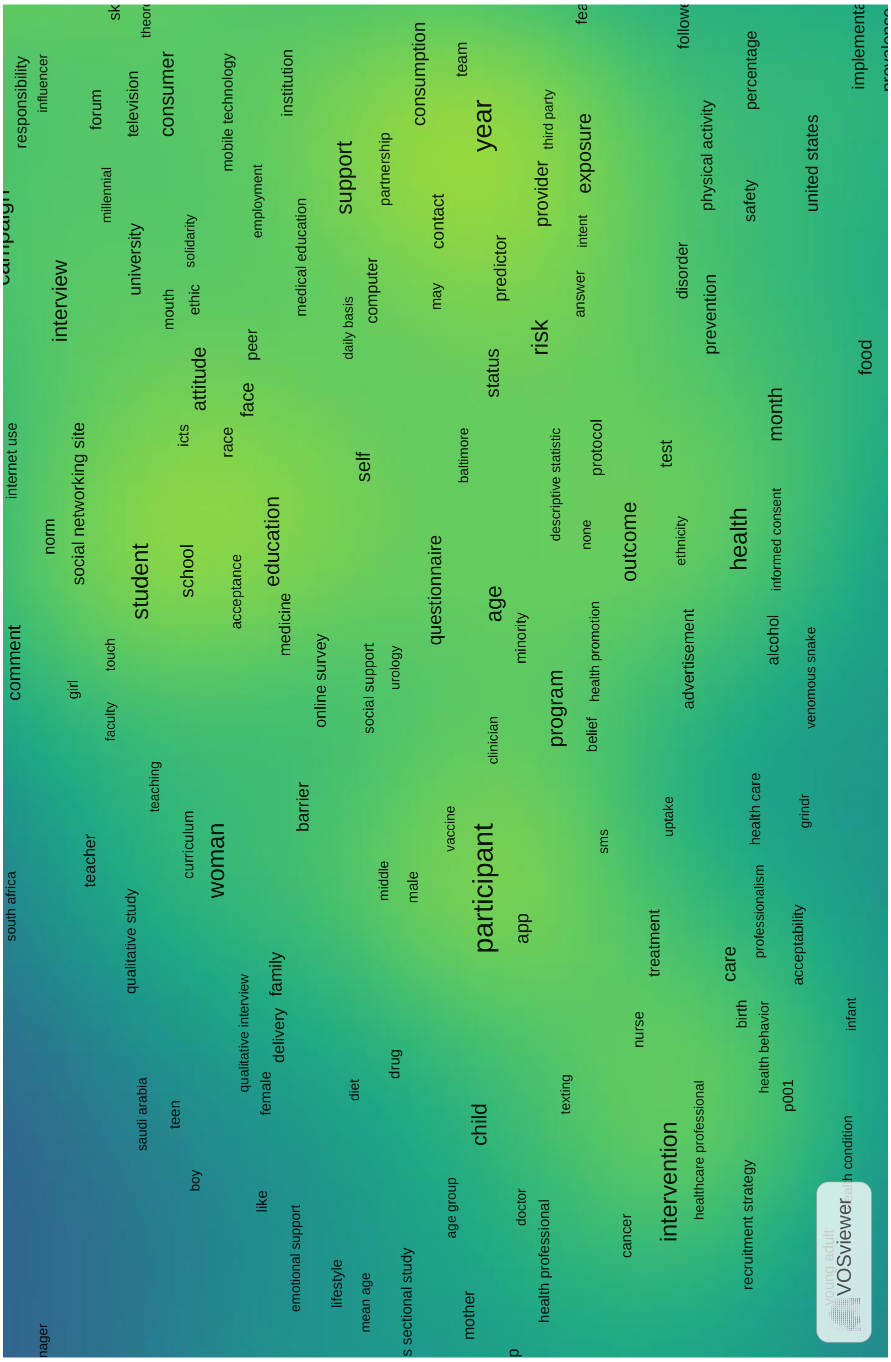


Figure 8.11: Social geosocial research, education, health and commercial focused area - noun phrase co-occurrence density map

Altogether, like the TF-IDF analysis of *G3*, the noun phrase co-occurrence network analysis - in particular, the map's right side - illustrates the shared use of computational data analysis methods to conduct urban research - including the study of specific cities, to study crisis management, and geographic geosocial research's focus on ecosystems and landscape preference. However, in contrast to the TF-IDF analysis of *G3* - which also comprised all geosocial papers - the noun phrase co-occurrence network visually obscures differences in how computational data analysis methods are used. Figure 8.6 differentiates a number of clusters on the left side of the map (among approaches further away from computational methods depicted on the right side), and fewer clusters on the map's right side. Figure 8.7 highlights the gap between geosocial research which uses computational methods and that which do not by emphasising the gap between the left and right side of the map.

The discrepancies between the author-bibliographic coupling and noun phrase co-occurrence network analyses highlight the importance of diversifying scientometric network analyses when exploring patterns scientometrically. I interpret the scientometric findings in light of interview analysis. The TF-IDF analysis which highlighted the diversity of computational geosocial research is more in line with my interviewees' research and narratives, which show the diverse ways they use computational data analysis methods. The rest of this chapter further explores two findings highlighted by the scientometric analyses in this section. Section 8.2 will explore differences in how computational data analysis methods are used for geosocial research. Section 8.3 explores how interviewees from diverse backgrounds study specific locations.

8.2 Spaces and Methods

This section uses heterogeneous network analyses to explore how two computational methods - machine learning (ML) and social network analysis (SNA) - mediate knowledge about spaces. As section 3.4.3 discussed, I perform two heterogeneous network analysis for each method. Firstly, to study how scholars position their geosocial research using these methods with respect to existing geosocial scholarship, and to explore the extent to which the methods enable a new set of scholars to conduct geosocial research, I study heterogeneous networks whose nodes are authors and author keywords. Secondly, to explore how methods mediate knowledge about spaces in more detail, I explore networks whose nodes are noun phrases from the abstracts and titles of papers categorised as relating to 'spaces', 'methods' or 'other'.

Chapters Five and Seven argued that geosocial research approaches differ with respect to the methods they use and the types of spaces they study. Section 7.2 provided scientometric evidence, and sections 5.2.2 and 5.3 provided insights based on interviews. For example, social scientist quoted in section 5.3 state that they modify computational data analysis techniques. My conceptual framework outlined in section 2.4 and interviewees' quotes also suggested that methods mediate knowledge about spaces. For example, Chase's attempt to modify a computational method to increase the number of variables it can model discussed in section 5.2.1 provides an example. This section explores in more detail how methods mediate knowledge about spaces. I find differences in the types of spatial units explored with ML and SNA.

As section 2.6.3.2 discussed, visual heterogeneous network analysis in STS has mainly been used to study how human and non-human actors coordinate collaborative, distributed scientific practices, such biomedical research. These projects often mapped the heterogeneous actors pertaining to a coordinated community - for example an institution - to trace the role of human and non-human agencies. In contrast, findings in Chapters Five and Six suggest that geosocial research is not a coordinated community in the above sense. I argued that my participants are primarily embedded in the epistemological traditions of their disciplines or research areas, and develop diverse approaches to geosocial research in line with these. I differentiate approaches to geosocial research - which differ with respect to methods and substantive foci - but I do not argue that they are coordinated communities. Thus, instead of studying heterogeneous networks of geosocial research or of different approaches, I study heterogeneous networks of sets of geosocial papers which use specific methods - machine learning (ML) and social network analysis (SNA). This shifts the analytical focus from tracing diverse actors' - such as scholars' and methods' - role in coordinating geosocial research, to exploring how methods bring forth or afford specific spatialities. Studying all geosocial papers which use a certain method also helps me capture some of the variation in how it is used and performed (cf. Marres and Gerlitz, 2016). Next, I discuss my choice of ML and SNA.

I explore how ML mediates knowledge about spaces because as section 5.1.1 argued, several interviewees with diverse disciplinary backgrounds are interested in exploring its affordances for geosocial research. At the same time, ML is a method that originates in computer science and statistics, and as figure 8.8 shows, used for computational geosocial research. Thus, interviewees' interest in it points to computational methods' capacity to travel across research areas. Interpreting the heterogeneous network maps of geosocial research which uses ML necessitates a comparative heterogeneous network, because I do not have detailed interview information about how participants use ML for geosocial research. I chose SNA as a comparative case because ML and SNA illustrate the diversity of computational methods used for geosocial research and their scientific footprints allow comparison. Both methods are popular among geosocial scholars, and allow quantitative, computational data analysis. In addition, the geosocial papers that mention them in their abstracts or titles appear in my scientometric dataset in similar years: ML in 2015 and SNA in 2014. Thus, the term maps which depict these papers are semantically rich. However, they originate in different research traditions and thus can help explore differences among approaches to geosocial research. SNA - as opposed to 'network science' - has, since its birth developed in relation to social scientific theories (cf. Erikson, 2013; Wasserman and Faust, 1994). In this sense, it stands in contrast with ML, which originates in computer science and statistics.

Firstly, as discussed in section 3.4.3, through comparing the author - author keyword networks of geosocial papers which use ML and SNA (ML and SNA papers, for short), I explore how scholars position geosocial research mediated by these methods with respect to previously published geosocial research, and the extent to which these methods enable new set of scholars to conduct geosocial research. Figure 8.12 depicts the network of authors and author keywords associated with papers which mention ML in their abstract, title or among the author keywords. The colors indicate the novelty of authors and keywords in the dataset. Authors in orange (orange dots) have published research which is part of my scientometric field prior to their paper using ML, whereas for authors in red (red dots), the first time their paper appears in my scientometric field is the one which uses ML. Orange author keywords (orange squares) were used by geosocial papers in my scientometric before they appear in papers associated with machine learning, whereas the first time red keywords (red squares) appear in my scientometric data set is in association with ML papers. To help interpret this network, I compare it with figure 8.13, which depicts the network of authors and author keywords associated with papers which mention SNA in their abstract or among the author keywords. Like for the ML network, the color of nodes indicate whether the author has published research in the dataset or not, and if the keyword was used before by geosocial scholars or not. The red colour indicates authors and keywords present in the dataset before their association with SNA papers.

Both ML and SNA papers are dominated by scholars and author keywords that were not present in my scientometric field before they appear in association with the two methods. Thus, both methods seem to enable a new set of scholars to conduct geosocial research. The position of existing and new authors and author keywords seem to be similarly distributed in the two networks. This suggests that in both cases, papers may use a few keywords previously mentioned by geosocial papers, but the majority of author keywords are novel. I note two further similarities. Both ML and SNA papers share methods related keywords with previously published geosocial papers. For SNA papers, these include 'event detection', 'mixed methods', 'social media', 'social networks' and 'networks'. For ML papers they include 'spatiotemporal data', 'gis', 'text mining', 'classification', 'social media analysis', 'data mining', 'volunteered geographic information', 'topic modelling', 'big data', 'crowdsourcing', and 'social network analysis'. In addition, ML and SNA papers share technology related keywords with existing geosocial research, which may refer to methods or users' technology mediated situated practices. For SNA, these include 'media', 'internet' and 'social networking sites'. For ML, these include 'yelp', 'digital technology', 'flickr' and 'twitter'.

However, the number of space-related keywords differentiate the two networks. Several of the ML papers' author keywords previously present in my scientometric dataset express spatial units or situated events or practices, such as 'temporal-spatial patterns', 'urban space', 'natural disasters', 'human mobility', 'Singapore', 'public space' and 'protest'. For SNA-related papers, the two keywords related to situated practices are 'transportation' and 'information diffusion', but none of the keywords already present in my geosocial dataset express spatial units. In line with diverse interviewees' interest in ML, the heterogeneous network analysis suggests that scholars position their geosocial research which uses ML by mentioning diverse spatial units of interest to geosocial scholars.

However, it is important to note two limitations when interpreting these maps. Firstly, absent data can impact my findings, which poses pronounced challenges because of the relatively small size of this dataset. Excluding even a few papers - either due to my field delineation method or WoS' selectivity - can impact the relative number of new authors or author keywords in the networks. Secondly, I interpret the novelty of author-keywords tentatively, because this categorisation is impacted by uncertainties around the meaning of publication dates. My analysis creates a binary variable based on publication year: it signals whether author keywords were present in the dataset before papers which use ML or SNA appeared. However publication time lag renders the meaning of publication year uncertain. My next analysis further explores how methods mediate knowledge about spaces, and offsets some of these limitations by using noun phrases in the abstract and titles of geosocial papers, omitting the temporal perspective.

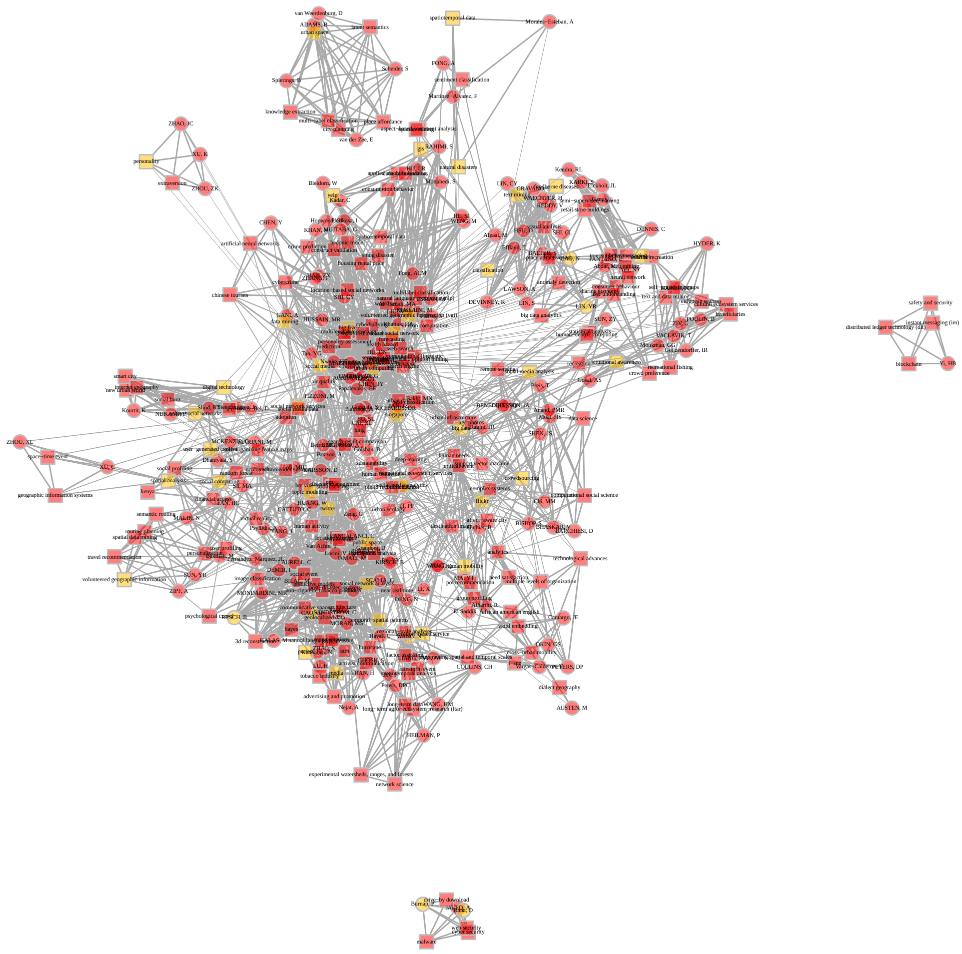


Figure 8.12: Ego network of 'machine learning' - authors and author keywords

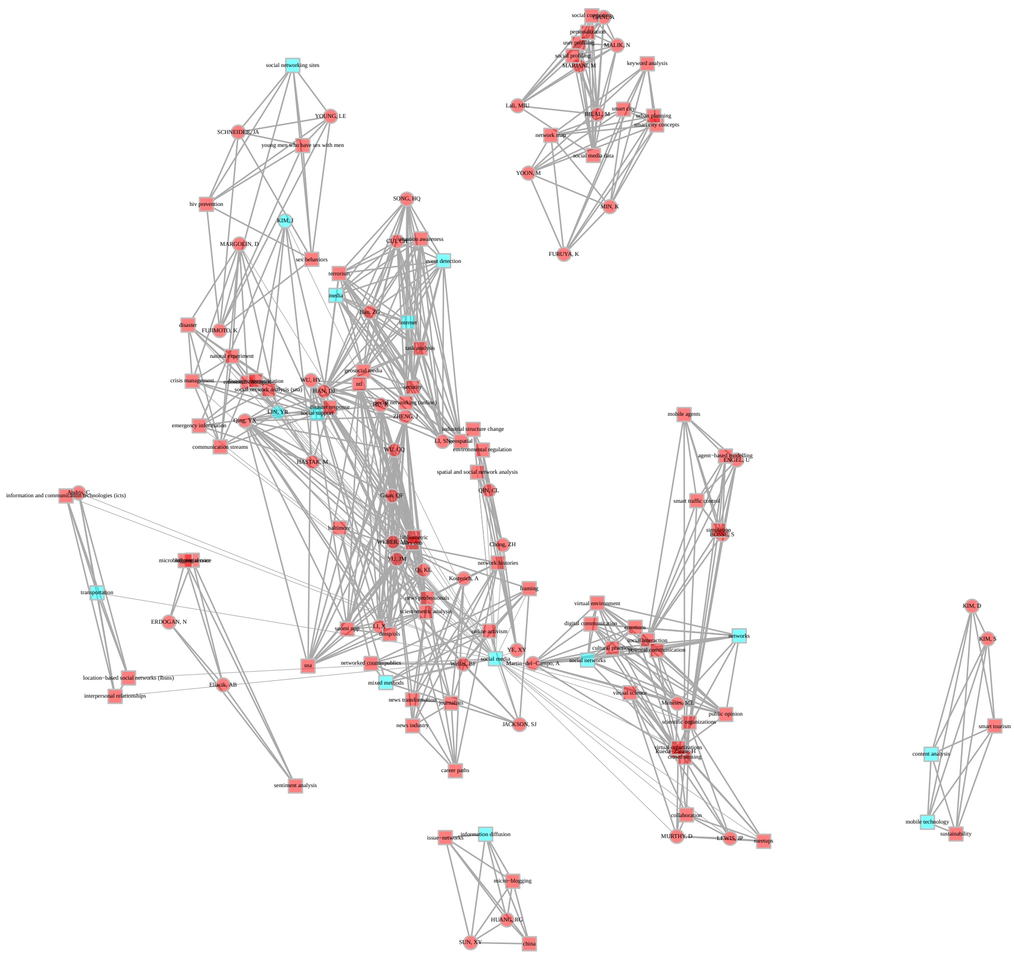


Figure 8.13: Ego network of 'social network analysis' - authors and author keywords

To explore in more detail the types of spaces ML and SNA afford, I study heterogeneous networks whose nodes are noun phrases identified in the abstracts and titles of papers categorised as related to 'methods', 'spaces' or other. As section 3.4.3 discussed, this helps explore how methods mediate the types of spaces scholars study in more detail because noun phrases in the abstracts and titles of geosocial papers - more numerous than author keywords - contain more information about the methods and spaces scholars study.

Figure 8.14 depicts the heterogeneous network of ML papers, and figure 8.15 depicts the heterogeneous network of SNA papers. I used VOSviewer's user interface to zoom in and out of maps for the visual analyses discussed below. Methods-related noun phrases are relatively evenly distributed in both networks. However, the position of space-related noun phrases differ in the two networks. In the SNA network they tend to be in the periphery of the network - with the exception of the noun phrase 'location' in the middle of the map (hidden in the static screenshot depicted by figure 8.15, but shown by figure 8.16). In addition, it is likely that 'united states' would also be more centrally located, if the noun phrases 'usa' and 'united states' were merged into one noun phrase. In contrast, in the ML map, space related noun phrases are present both in the network's centre and periphery. Relatively centrally located space related noun phrases include 'location', 'city', 'homogeneous region', 'area', 'spatial scale', 'spatial pattern', 'disaster' and 'place'. This suggests that these space related noun phrases, many expressing spaces in terms of spatio-temporal units, help coordinate geosocial research which uses machine learning, and 'hold this network together'. Thus, this heterogeneous network analysis shifts my focus from methods' mediation to asking how spatial units coordinate geosocial research.

Finally, the meaning of space-related noun phrases also differs in the two networks. The pattern I find resonates with my findings discussed in section 7.2 about social geosocial research's comparative focus on situated practices at specific locations and technical geosocial research's comparative focus on spatial scales. In the SNA network map they mostly refer to types of environments, such as 'smart city', 'border wall', 'real world', 'virtual world', 'geographic space', 'important place' (6 out of 19, 31%) or specific locations such as 'china', 'USA', 'Baltimore', 'Lousiana', 'Boston' (5 out of 19, 26%). Space-related noun phrases that refer to spatial scale units are 'city', 'world', 'country' and 'location' (4 out of 19, 21%). In contrast, in the ML network map, a large proportion of space-related noun phrases refer to units measurable on spatial scale, such as 'area', 'city', 'country', 'homogeneous region', 'location', 'neighbourhood', 'place', 'region', 'space', 'spatial pattern', 'spatial scale', 'urban area' (12 out of 29, 40%), and relatively small proportion of noun phrases refer to types of environments ('landmark', 'environment' and 'south', 'store building' and 'geographic location' (5 out of 29, 17%)). The proportion of noun phrases that refer to specific locations - including 'Amsterdam', 'Beijing', 'China', 'London', 'Shenzen', 'United States' (6 out of 29, 21%) - is also smaller than in the SNA map. Finally, the proportion of noun phrases which refer to situated events is similar in the two networks. In the SNA network they include 'disaster', 'protest', 'bombing' and 'environmental regulation' (4 out of 19, 21%), and in the ML network they include 'disaster', 'air quality', 'ecosystem ecologist', 'sustainability', 'urban leisure', 'urban planning' (6 out of 29, 21%). While there are semantic overlaps among them - several of them relate to disasters and sustainability - research which uses SNA studies more cultural conflicts, and ML papers study urban planning and leisure activities in cities.

This section and section 8.1 showed that diverse approaches to geosocial research, and both papers that use ML and SNA study specific locations. The next section explores the role of local knowledge when studying specific locations.

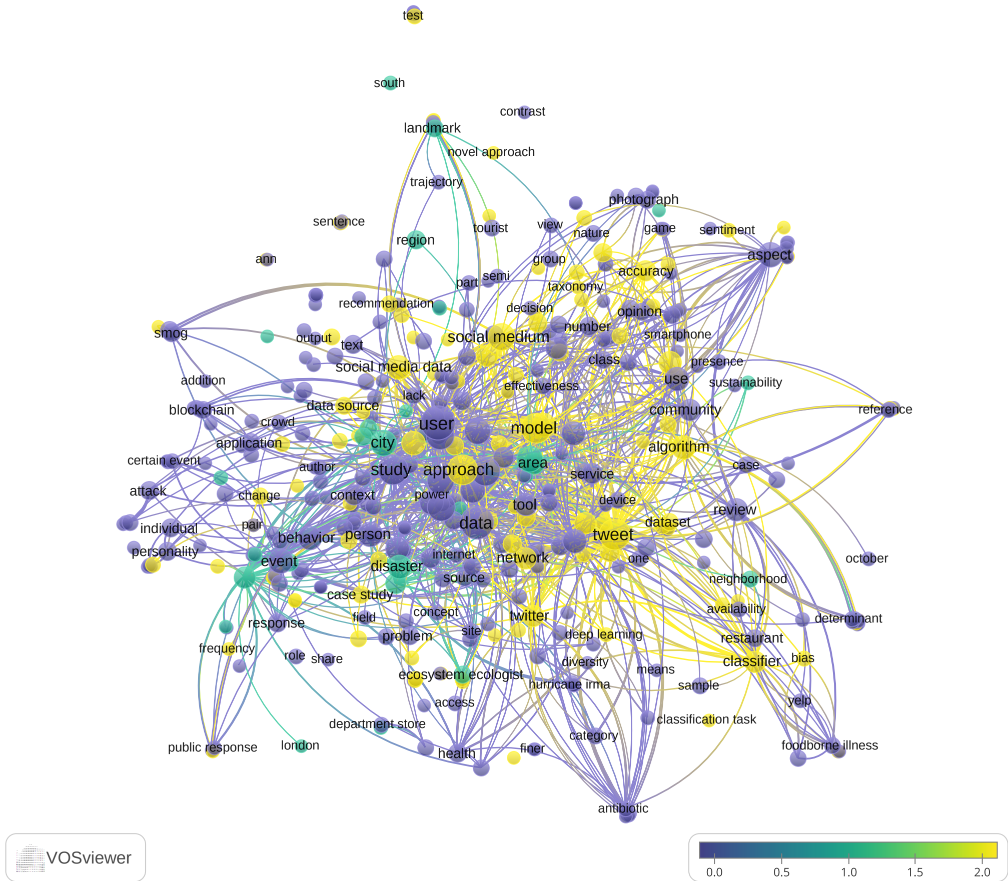


Figure 8.14: Machine learning papers: methods and spaces heterogeneous network. Green = space, Yellow = method, Purple = other

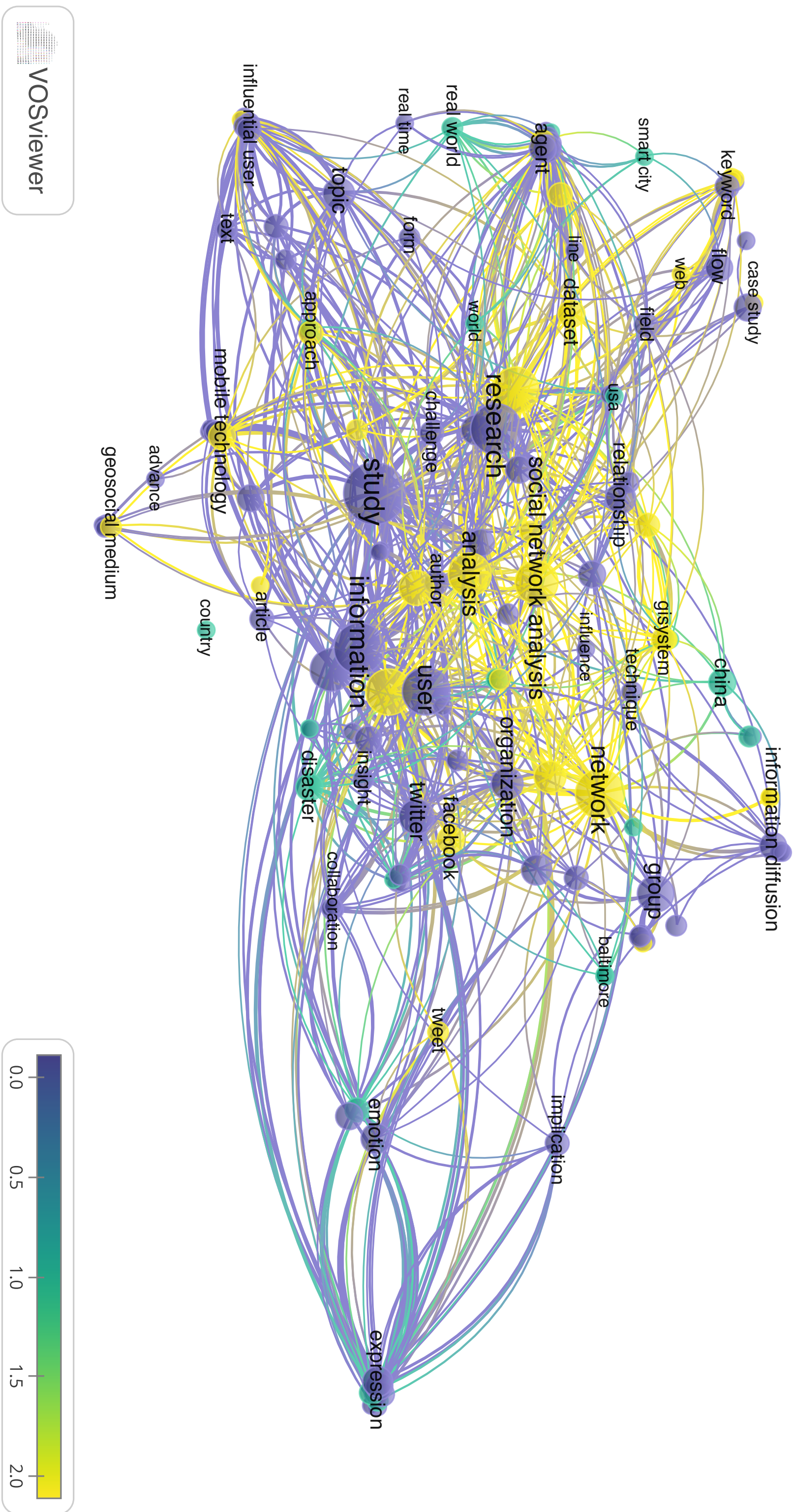


Figure 8.15: Social network analysis papers: methods and spaces heterogeneous network. Green = space, Yellow = method, Purple = other

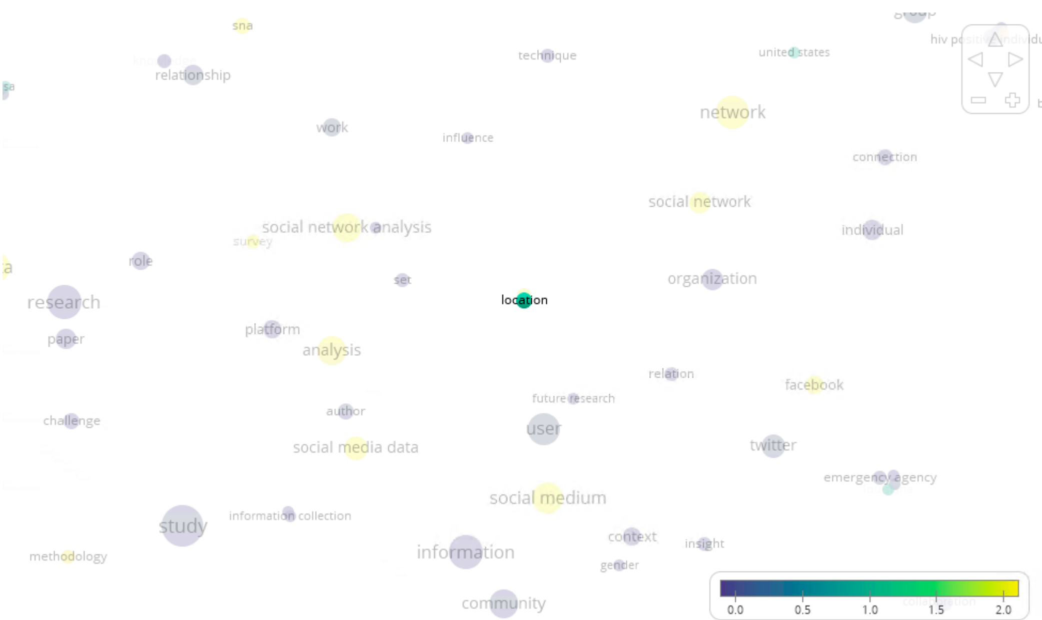


Figure 8.16: The position of the noun phrase 'location' in the heterogeneous map associated with social network analysis

8.3 Using Local Knowledge to Study Specific Spaces

This section argues that interviewees with diverse disciplinary backgrounds use local knowledge to study specific locations. Sections 7.2 and 8.2 differentiated between three types of spaces geosocial scholars study using scientometrics - spatio-temporal units (e.g. 'spatial scale', 'area', homogeneous region', 'city', 'country'), spatial patterns (e.g. 'spatial distribution'), types of environments without immediate scalar connotation (e.g. 'environment', 'landscape') and specific locations (e.g. London, Singapore). As sections 8.1 and 8.2 argued - and the quotes below will show - diverse geosocial research approaches study specific locations. This section explores how local knowledge helps interviewees study these spaces.

The majority of my interviewees with diverse disciplinary backgrounds - those in groups A, B, C, D, F, G, I, J, K, L and M - study specific locations, or the way social media platforms mediate situated practices at specific locations. A minority - Henry and Elias - state that geosocial data is best suited to study 'online' spaces - the social media platforms themselves. This section focuses on the geosocial research practices of the former group - participants who study 'offline' specific locations. As the quotes below illustrate, interviewees agree that local knowledge is important for such geosocial research, and some even find it essential. I highlight four main ways local knowledge helps interviewees study specific locations. Firstly, a number of interviewees use it to interpret computational findings, or develop them more quickly. Secondly, it can help validate analytical decisions and assess data quality. Thirdly, researching locations one knows can help motivate scholars. Finally, it can help participants to engage with the local practices they study and slow down the data analysis.

Firstly, David's, Anne's and Gary's, Jane's narratives and the practices of researchers at the summer school where I conducted participant observation show that having local knowledge about locations helps researchers interpret the findings of geosocial research (more quickly). As section 5.3 outlined, Jane stated local knowledge helps her team interpret their findings. Below I quote interviewees who emphasize that local knowledge 'speed up' the data analysis process because it helps them interpret findings more quickly. David created a data visualisation for a collaborator based at another research group, about a city David didn't know. He noted that his collaborator could *immediately* make sense of his data analysis:

> *"I would send [my collaborator] a map, or visualization, or table [...] there was a definite division of labour, in the sense that I could process the data [...] which he immediately was able to make sense of. It was really interesting."*

Anne studied two cities - City A with which she was familiar, and City B she had not visited. She chose City A due to her familiarity with it, and City B because she thought the similarities between two cities made them comparable. She claimed that her familiarity with City A helped her interpret the data about it faster.

> *"[City A], I chose because [...] I know it well enough to be able to kind of understand at a relatively quick glance what people are talking about. [City B] I picked because [...] I felt that the [...] urban flow of the cities vaguely mirrored each other [...] [My familiarity with city A] probably did make a difference in the sense that it was easier to categorise the data faster because I didn't have to Google as much."*

Finally, Gary stated that when he encountered an area with high data density, thanks to his local knowledge, he immediately knew the reason for the higher number of social media posts there. He believes that without local knowledge, one can see the change in density, but cannot immediately interpret it.

> *"Like I know exactly where there's high density there, because there's this kind of [a landscape element] there. [Someone with no knowledge about the place] also would have seen the density [of data points], of course. And would have asked the question, 'so what's going on there?'"*

In sum, local knowledge about specific locations can help interpret findings, or interpret them more quickly.

Secondly, Isaac, Miles, Gary and Luke stated that local knowledge helps them validate analytical decisions, findings, and assess data quality. Isaac's local knowledge informed his analytical decisions: his geosocial research required his team to categorise social media posts. He noted that local knowledge was essential for this task. Although he is familiar with the country they studied, he is not a native speaker of the local language. He chose a collaborator skilled in relevant computational data analysis methods and who was a native speaker. He noted that using their local knowledge, they categorised the social media data together, with little disagreement.

> *"I think I could never have written this paper about another country that I don't know or where I don't speak the language. Plus we sought a co-author who is a native speaker in addition to having strong methods skills... [...] we coded [categorised] the [social media posts] together [...] we didn't disagree very much, it was very obvious in almost all cases. [...] So the local expertise is very good."*

Mike claimed that local knowledge helped them assess and validate their analytical decisions and findings on an ongoing basis by providing a baseline. This helped them ascertain that their research was proceeding in the right direction and produced valid findings.

> *"The reason why most of our work prioritises studies [city A] and [city B], is because in the team people know them. And it's much easier, because most of this stuff, needs some qualitative assessment... [...] you need to assess validity. [...] one way [you can do that] is quantitatively by looking at [other data], but another way qualitatively, so people who have a bit of knowledge about the city, check if these things make more or less sense. [...] knowing the landscape of the cities helped a lot to understand that we are going in the right direction."*

Gary's and Luke's narratives show that local knowledge can help assess data quality and platform effects. Luke stated that local knowledge helped him identify an incorrect geospatial coordinate (a geotag), frequent in the data, resulting from platform effects.

> *"I think [knowing the city] made it significantly easier to interpret the data. [...] I needed that extra spatial knowledge to realise that something had gone wrong [in the spatial accuracy of geotags] [...] [I found a cluster of posts in the city] centrum. At first I actually thought [...] 'this must just be a popular spot'. But then I realized, when I zoomed in really really closely that it was actually in a middle of a cross road [...] I think if I'd studied [other cities], I wouldn't have been able to tell the difference."*

Finally, Gary stated that thanks to his local knowledge, he added data which improved his analysis. As outlined above, he encountered an area with high data density where, thanks to his local knowledge, he knows that there is a specific landscape element. Given his knowledge, he sought alternative data which included this landscape feature, knowing that including it would improve the fit of his model.

> *"I know exactly why there's high density there, because there's [a landscape element] there. [...] That was a little tricky, I didn't have data on it, [...] so I had to like look for data source that had this, or just add this point myself, because I knew that if I included it, it would increase my model fit."*

In sum, these interviewees state that local knowledge helps them validate the results of their data analysis or assess and correct their data.

Thirdly, Chase and Kevin noted that researching locations one knows well can help motivate scholars by creating locally-relevant knowledge. Chase stated that it is interesting and enjoyable to study a country he knows well, because it helps create findings to which he and others who read his research can relate. In addition, his local colleagues propose new locally-relevant topics to study with geosocial data.

> *"It's great that we can link our research to other data and concrete events [...] and say [statements with societal and policy relevance.] That's a really good feeling. [...] And it's great to understand things that happened here. [...] And that people are actually interested in it, because it's about us. [...] When we present our work [in this country], colleagues keep proposing new topics we could explore with the data..."*

Kevin noted that local knowledge helped his geosocial research, and he appreciated the opportunity the project gave him to extend his knowledge about the country where he lives, even though he does not consider local knowledge essential for geosocial research.

> *"I know [this country] the best, and this project allowed me to get to know it even better. You never know it as much as you think you do... Naturally it helps a lot to study a country that you know more or less. But in the end we want data. If I get data about [a country I have never been to], I will study that country."*

The summer school Frank organised, which aimed to explore a city 'remotely', through geosocial data without 'offline' research further illustrates the above three roles of local knowledge. Participants of the workshop were divided into groups of five. All groups had at least one participant local to the studied city. Throughout the workshop, the locals helped their groups interpret geosocial data maps using their experiential knowledge about the city. In addition, they researched news stories and other information about the city, using their knowledge about relevant news sites and data sources. Locals helped their groups use geosocial data in three main ways. Firstly, as above, they helped interpret the data and data analysis results by telling stories about locations with particularly high or low density of geosocial data, interpreting the meaning of social media data density. Secondly, they provided confidence and facilitated cohesion and motivation in their groups by answering non locals' questions and ensuring fellow group members that their research was 'locally relevant'. Finally, they proposed research questions relevant to both the city and urban research. Some suggested focusing analyses on the city's waterfronts, some of which have been undergoing redevelopment, gaining new attention or were in decay. Others suggested to study local societal challenges such as gentrification, or local practices, such as local holidays.

Finally, Bruno, Colin, Daniel and David stated that local knowledge made them question the results of the data analysis and slow down the data analysis process. As section 5.3 explained, Bruno stated that group B seek to study spaces they know, and modified computational data analysis methods using their local knowledge. As section 7.1.1 argued, Colin claimed that he worries about whether his team's analyses accounts for the perspectives of a minority group living at the space they study in an ongoing manner. Finally, as section 7.1.1 outlined, David and Daniel highlighted the importance of reflecting on how their lived experience and local knowledge shape their results. They stressed the danger of interpreting situated practices without visiting the locations they study in person. In sum, Bruno, Colin, Daniel and David claimed that local knowledge can help slow down data analysis.

Altogether, this section highlighted the importance of local knowledge for studying specific locations.

8.4 Conclusion

This chapter contributed to all three Research Questions by diversifying the scientometric units of analyses used to compare geosocial research approaches and studying how research methods mediate knowledge about spaces using mixed methods. Below I discuss my findings with respect to each research question.

The chapter contributed to answering the First Research Question - which asks how geosocial research approaches develop and differentiate - in two ways. Firstly, it explored how computational research methods mediate knowledge about spaces. Given the diverse ways geosocial scholars use computational methods - illustrated in Chapters Five and section 8.1 - I used heterogeneous network analysis to compare how two computational methods - machine learning (ML) and social network analysis (SNA) - mediate knowledge about spaces. This also helped answer Research Question 3.1 (which asks how methods mixing can help study the differentiation of and differences among geosocial research).

I performed two types of heterogeneous network analyses. Firstly, author-keyword and author networks suggested that both SNA and ML enable a new set of scholars to conduct geosocial research, but ML-related geosocial research shares more keywords related to spatial practices with prior research than SNA-related geosocial research. However, missing data and data's small size caused interpretative challenges.

Secondly, heterogeneous noun phrase co-occurrence network analysis (where I classified terms as related to 'methods', 'spaces' or 'other') helped explore how ML and SNA mediate knowledge about spaces in more detail. I found a difference in the types of spaces the two methods help study. Comparatively, noun phrases which refer to 'types of environments' were more frequent among SNA-related geosocial research, while terms which express 'spatial scales' were more frequent ML-related geosocial research. The heterogeneous network analysis also highlighted the position of these terms in the network. I found that 'spatial scale' terms occupy relatively central position in the term map of geosocial scholarship which uses ML, in contrast with concepts which refer to 'specific locations'. The latter are more peripheral in both heterogeneous networks. This raised the question of whether, and how 'spatial scale' units help coordinate geosocial research which uses diverse methods, which further research can explore.

Secondly, with respect to the First Research Question, section 8.3 argued that interviewees with diverse disciplinary backgrounds use local knowledge to study specific locations. As section 2.13 discussed, using local knowledge during mapping is often seen as a way to diversify the cartographic knowledge - to ensure that maps capture diverse local perspectives and local knowledge rather than those of the mappers (Pain, 2004). However, my interviewees' practices suggest that including local knowledge in mapping does not always result in more heterogeneous maps. While in some cases local knowledge motivated interviewees to question the data patterns they found and prompted them to think about their research in light of diverse local experiences, in other cases local knowledge helped speed up participants' geosocial research - to interpret computational findings more quickly. In addition, similar to the New York based urban data analysts Taylor and Richter (2015) studied who validated a fire risk model using the local knowledge of the city's fire inspectors, my interviewees used their local knowledge to validate analytical decisions and to assess data quality. However, my interviewees work in academic institutions and mostly draw on their own experiential knowledge rather than that of other people or expert groups.

I explored the Second Research Question (which explores how geosocial research approaches differ) and Research Question 3.1 (which asks how methods mixing can help study differences among geosocial research approaches) by comparing the scientometric footprints of geosocial research approaches through a series of inductive network analyses. I compared the analysis of the term co-occurrence network of geosocial papers and the analysis of the clusters of the author-bibliographic coupling networks *G3* (comprising all geosocial papers) and *G4* (comprising 'only social' and 'only computational' geosocial papers, outlined in Chapter Seven). This helped illustrate geosocial research approaches' diversity, similarities among them, as well as my findings' contingency on data analysis infrastructure.

Analysing the clusters of *G3* yielded two main insights. Compared to *G4*, it better highlighted the diverse uses of computational data analysis methods for geosocial research. Noun phrases which refer to diverse computational methods differentiated three out of the four clusters I analysed. For example, my analysis highlighted the Geography Cluster's comparative focus on citizen science, volunteered geographic information, geographic data analysis methods and ecological focus. In addition, my analysis highlighted that the Mixed Cluster places more emphasis on (urban) and spatial practices and events which can be modeled as spatio-temporal diffusion and mobility patterns, and the Technical Cluster is differentiated in terms of its study of large spatial scales and use of deep learning. In addition, the analysis of *G3* highlighted diverse geosocial research approaches' focus on studying specific locations. Secondly, I analysed the noun phrase co-occurrence network of terms from abstracts and titles of geosocial papers. This obscured the diverse uses of computational methods and instead highlighted the difference between geosocial research which uses computational methods and that which does not. It helped illustrate the diversity of social geosocial research by depicting the distance of social geosocial research projects from geosocial research which use computational data analysis methods. The contrast between the author-bibliographic coupling and noun phrase co-occurrence analyses based on the same set of papers (all geosocial papers) highlights the importance of diversifying scientometric network analyses when exploring patterns.

Finally, this chapter's mixed methods approach helps answer Research Question 3.2, which asks how we can evaluate the affordances of computational methods 'for STS'. Below I discuss how the interpretative context informed my use of each computational method in this chapter.

The interpretative context informed the cluster analysis of the author-bibliographic coupling network *G3* similar to the clustering of *G4* in Chapter Seven. In addition, the sociological relevance of the clusters of *G4* found in Chapter Seven - overlapping with clusters of *G3* - gave me confidence that the additional cluster (the Mixed Cluster) I identified also was relevant to geosocial research practice. In addition, I compared these findings with the analysis of *G4* and the term-map, and interpreted them in light of interview findings about the diversity of computational data anlaysis and approaches' shared interest in studying specific locations.

I visually analysed the clusters of the noun phrase co-occurrence network. This analysis hinged on interviews, the conceptual framework and prior scientometrics, because I compared the differences it highlighted about how scholars use geosocial data with respect to earlier findings about approaches to geosocial research.

Interviews, prior scientometrics and my conceptual framework informed the heterogeneous network analysis in three main ways. Firstly, both interviews and scientometrics highlighted the diversity of computational methods used for geosocial research, and my conceptual framework suggested that these methods shape knowledge about spaces. Thus, I assumed that comparing geosocial research which uses specific computational methods is a relevant unit of analysis. Secondly, interviews suggested that participants from diverse disciplines were interested in exploring the affordances of ML for geosocial research, highlighting it as a method to further explore. Thirdly, based on my conceptual framework and literature review, I assumed that ML and SNA can help illustrate the diversity of computational methods used for geosocial research. I assumed that SNA had closer links to the social science research tradition (cf. section 2.6.3.1), whilst ML had originated in computational disciplines.

As section 9.3.1.3 will discuss in more detail, compared to previous STS studies which used heterogeneous network analysis to study how large scale biomedical research collectives are 'held together' by diverse human and non-human agencies, my use of heterogeneous network analysis highlighted their affordances to study the mediated nature of knowledge.

Finally, as section 9.3.2 will discuss in more detail, comparative scientometric analyses and visual, heterogeneous network analysis of ML and SNA-related geosocial research illustrated the possibility to create 'surprising' findings - not hypothesised based on interviews - with scientometric network analyses.

Chapter 9

Conclusion

This chapter concludes the thesis by summarising its main findings and discussing how future work may build on it. The thesis explored three main research questions:

1. How do different approaches to geosocial research develop?
2. How do approaches to geosocial research differ?
3. How can we combine scientometrics and STS to study geosocial research?
   3.1. How does mixing methods help study the development of and difference among approaches to geosocial research?
   3.2. How can we assess the suitability of computational methods for STS?

I explored these questions for five main reasons listed in table 9.1: to explore the diversity and diversification of computational social research; to reflect on the ESRC's (2013) call to 'close the gap' between quantitative and qualitative human geography by emphasizing methods' complementarity; to contribute to methodological discussions in academic literature which call for combining STS and scientometrics; to create dialogues between distinct ways of knowing; and finally in the hope that this project's findings can inform research methods curriculum development in the social sciences.

Table 9.1 summarises the thesis' key findings in terms of the three research questions and the thesis' five goals. Section 9.1 briefly summarises the thesis' findings about geosocial research (discussed in Chapters Four through Eight in detail), which inform my reflection about the ESRC's call outlined in section 9.2 and recommendations for curriculum development are discussed in section 9.4. Finally, section 9.5 outlines future research avenues.

The thesis' main contribution to academic literature, discussed in section 9.3 is the methodological development and application of its mixed methods approach. It contributes to literature which explores connections between STS and scientometrics conceptually by highlighting the need to develop mixed methods designs and evaluate the affordances of scientometric or computational methods 'for STS' in light of the interpretative context, including the research questions, conceptual framework, characteristics of the studied research practice and previous findings. Empirically it contributes by combining methods in diverse ways as part of a single case study, and illustrating the affordances of homogeneous statistical network analysis and descriptive statistics 'for STS' often critiqued by STS scholars. Given the importance of the interpretative context, using diverse methods as part of a single case study helps assess their affordances to study diverse relational practices which comprise scientific practice. Reflecting on how STS and scientometrics can 'co-compose' knowledge, I argue that aligning units of analyses can help study research practices in depth and on larger scales; using them inductively helps highlight partial perspectives and generate 'surprising' findings. In addition, I argue that calculation practices can inform qualitative analysis even if the scientometric analytical units do not capture research practice, but this may require the analyst(s) to actively remember the discrepancy between units of analyses and relations.

<table>
<tr><th>Section</th><th>RQ</th><th colspan="2">Relevance</th><th>Main findings</th></tr>
<tr><td>9.1

Chapters 4-8 in more detail</td><td>RQ 1 & RQ 2</td><td colspan="2">Study the diversity and diversification of computational (social) research, a dominant way of knowing.</td><td>Practices that help geosocial scholars develop diverse approaches; differences among approaches.</td></tr>
<tr><td>9.2</td><td>RQ 1

&

RQ 2</td><td colspan="2">Reflect on ESCR's (2013) call for 'closing the gap' between quantitative and qualitative human geography.</td><td>• Report focuses on 'complementary'. However, I argue that maintaining the diversity of (digital) geographic scholarship also requires highlighting differences,
• Support the development of sub-disciplinary communities for scholars with broadly shared epistemic commitments interested in digital / computational research,
• Curate opportunities for scholars to capture reflexivity about data and analytical decisions.</td></tr>
<tr><td rowspan="3">9.3

Summary of and reflection on findings discussed in Chapters 4-8</td><td rowspan="3">RQ 3</td><td rowspan="2">Evaluate affordances of digital methods 'for STS'.</td><td>Digital STS: 'import' methods.</td><td rowspan="2">• Conceptual contribution: evaluate affordances computational methods 'for STS' in light of interpretative context (RQs, conceptual framework, specificities of research practice, previous findings),
• Empirical contribution: compare diverse mixed methods solutions as part of 1 case study.</td></tr>
<tr><td>STS & scientometrics interface</td></tr>
<tr><td colspan="2">Social sciences' 'compositionist' agenda (cf. Latour, 2010): mixing methods.</td><td>• Convergence and divergence of methods (section 9.3.2):
  o Align units of analyses: study depth and scale,
  o Use methods inductively: highlight partial perspective & 'surprising' findings
  o Reflect on mixed methods findings and practice separately. Calculation acts can inform qualitative analysis even if the scientometric analytical units do not capture research practice; may require the analyst(s) to actively remember the discrepancy between units of analyses and relations,
• Skill development (see also below).</td></tr>
<tr><td>9.4</td><td>All RQs</td><td colspan="2">Research methods curriculum development in social sciences:
• personal interest,
• ESRC's (2013) call to increase mixed methods training in human geography (and across the social sciences).</td><td>• Skills: familiarity with the basics of diverse computational methods,
• Use aesthetically pleasing data & methods,
• Foster reflexivity about how distinct computational methods enable diverse types of spatial units,
• Importance of exploratory data analysis (EDA),
• When students with diverse backgrounds work together:
  o EDA,
  o Reflexivity,
  o Local knowledge (for geography).</td></tr>
<tr><td>9.5</td><td>All RQs</td><td colspan="2">Future research.</td><td>See section 9.5</td></tr>
</table>

Table 9.1: Chapter Structure, Research Questions, Relevance and Main Findings

9.1 Diversity and Diversification of Geosocial Research

This thesis examined the development and differences of geosocial research approaches to explore the diversity and diversification of computational (social) research, an increasingly popular way of knowing. I argued that geosocial research is a collection of approaches rather than a cohesive research community. I identified fourteen main practices through which geosocial scholars develop their own geosocial research approaches. A core challenge for interviewees was combining two research traditions which they perceived distinct: computational data analysis and socio-spatial interpretation. Participants were motivated to combine these due to their concurrent aesthetic appreciation of social research and geosocial data (analysis) and the need to balance academic and non-academic jobs. Thus, developing distinct geosocial research approaches required interviewees to actively align academic research traditions with each other and non-academic research they felt were diverging, all amid financial uncertainties. At the same time, developing their own geosocial research approaches as these grew in popularity across social and computational sciences provided interviewees with opportunities to craft coherent narratives about their research foci across their diverse research engagements. With the rise of undergraduate social science training programs which include computational training elements, scholars' non-academic work may play a less important role in fostering their engagement with geosocial or computational research.

Collaboration with scholars with complementary skills with whom they share methodological or theoretical common ground, and setting up their own geosocial laboratories where they could experimentally search for data patterns, helped interviewees combine computational data analysis and socio-spatial interpretation. Interviewees succeeded in using geosocial data for academic research by using local knowledge to assess findings and by reflecting on how social media platforms and analytical decisions shaped their data and the knowledge they created about spaces. Such reflexivity is an important aspect of developing geosocial research in line with disciplinary traditions, given disciplines' distinct capacity to account for their knowledge creation. In addition, reflexivity helps interviewees conduct scientific research with social media data which were not designed for scientific use and whose characteristics they cannot control. I also argued that methods mediate scholars' geosocial research and afford different spatial units.

In addition, scholars imagined their geosocial research in line with their institutions' perceived research foci in search for institutional homes.

Finally, social scientists actively differentiate their approaches from technical approaches through several practices. They highlight distinct epistemic virtues associated with their research. If necessary, they change departments and modify computational methods in light of the social scientific frameworks they work with. They also critically reflect on technical scholarship and seek (sub)-disciplinary communities and publication venues who value their geosocial research.

I identified five main differences among geosocial research approaches and illustrated that these findings are contingent on the comparison units. Through interviews I identified three main differences between social and technical geosocial research. I argued that their rhythm and knowledge validation methods differ. Social geosocial research combines computational data analysis and socio-spatial interpretation iteratively and seeks to identify data patterns using both computational and social scientific criteria. In contrast, technical geosocial research combines the above research traditions sequentially and identifies data patters based on statistical criteria. I also argued that social and technical geosocial research differ with respect to scholars' reflexivity. 'Hermeneutic reflexivity' associated with social geosocial research comprises reflection on how social media platforms and analytical decisions shape knowledge about spaces in historical and experiential terms - treating these a constitutive part of findings. In contrast, 'algorithmic reflexivity' associated with technical geosocial research comprises reflection on how social media platforms and analytical decisions shape knowledge about spaces in terms of calculations and demographics. In addition, using scientometrics, I highlighted differences among geosocial research approaches - including social, technical and geographic approaches - with respect to the methods they use and their spatial units of analysis.

In addition, through comparing citation and term co-occurrence network analyses (also with interviews), I illustrated the diversity of approaches and similarities among them. For example, I showed that social geosocial studies differ to the extent that they use computational data analysis methods, and technical geosocial studies use diverse types of computational methods, ranging from geographic methods, approaches to study mobilities or flows, to the development of advanced image recognition (remote sensing) and deep learning methods. Similarities among approaches include their interest in urban studies.

Altogether, my project highlights differences in the rhythm, methodological practice and (spatial) units of analyses associated with geosocial research approaches. This suggests that geosocial research can account for situated practices from diverse, complementary perspectives. Next, I discuss how these findings inform my reflection on the ESRC's (2013) call to close the gap between qualitative and quantitative methodological traditions in human geography and the social sciences more broadly, and section 9.4 will reflect on these findings with respect to curriculum development.

9.2 How to close the gap between geography's methods traditions?

Exploring the diversity and diversification of geosocial research helped me reflect on the ESRC's (2013) call to close the gap between quantitative and qualitative methodological traditions in human geography (and more broadly the social sciences) by emphasizing their complementarity. As the Introduction explained, the ESRC (2013) called to increase computational research in British geography in part through undergraduate and postgradute training programs that emphasize the complementarity of the quantitative and qualitative methodological traditions - for example, their ability to study phenomena at diverse scales. They call for changing dominant views that highlight their divisions and differences.

In contrast to the report's emphasis on highlighting the complementarity of quantitative and qualitative geography methodological traditions, my findings show the importance of developing diverse combinations between them, beyond emphasizing their complementarity. Social scientist interviewees develop their geosocial research approaches by critiquing and if necessary modifying computational methods. In order to maintain the diversity of digital and computational geography scholarship, I argue that institutional structures should facilitate small scale collaborations - including scholars with computational skills - that help researchers develop computational methods in line with distinct geographic research heritages, akin to the collaborations and geosocial laboratories established by my interviewees.

My interviewees' experiences discussed in Chapters Five through Seven show that quantitative and qualitative methods are often associated with contradictory epistemologies and ethical considerations, as well as the effort required to develop computational methods in light of (interpretative) social science theories. For example, sociologists David and Daniel developed computational methods aligned with social theories by combining computational and ethnographic work, and carefully examining whether terms proposed by computational social scientists describe their research practice. Chase (an economic geographer) altered a modeling method proposed by computational collaborators which cannot account for variables key to his study, but struggled to make this innovation recognised by computational collaborators. Human geographer interviewees in group B developed computational methods in light of local knowledge. Brian also noted the challenge of working with computational collaborators amid ethical differences. Altogether, highlighting the differences between existing computational methods or research traditions and social scientist interviewees' research goals was key to their success in developing computational research approaches.

Thus, in contrast to the ESRC's recommendation to emphasize methods' complementarity, I argue that curating diverse computational research in geography (or the social sciences more broadly) requires exploring differences, frictions as well as complementarities among methodological traditions, and collaborations among computationally skilled scholars and experts in distinct geographic research traditions. Overly focusing on methodological complementarities may hinder the development and use of computational and digital methods in line with distinct geographic research traditions, and demotivate scholars who feel that existing computational methods, or their current uses oppose their theoretical sensibilities.

In addition, based on Chapter Six's findings about the importance of belonging to (sub)-disciplinary communities for the differentiation of social geosocial research, I argue that the development of such communities for scholars interested in digital research with shared epistemic commitments in the social sciences is essential. This highlights the importance of fostering community building efforts, for example by funding workshops or conferences, publication venues or online discussion forums that help develop research communities within geography or related social sciences for scholars with broadly shared epistemic commitments who wish to pursue computational research (similar to the community building efforts of the digital STS community outlined by Vertesi et al., 2019b).

Finally, computational methods training could help scholars articulate their reflections about the way data and analytical decisions shape their findings. Such reflexivity can help foster the long-term development (cf. Strathern, 2004) of diverse computational approaches in geography (or across the social sciences). In addition, it can help explore complementarities, similarities and differences among computational or digital research approaches, further aiding their development and differentiation.

Such practices can build on geography's rich reflexive tradition outlined in Chapter Two. In addition, my project illustrates the centrality of such methodological reflexivity of successful geosocial research for scholars from diverse disciplines, but that it is potentially under-acknowledged and under-articulated. As Chapter Seven argued, reflexivity about how analytical decisions and social media platforms shape geosocial data and knowledge about spaces is core to geosocial research. However, my technical interviewees did not consider such reflexivity a standalone practice. Isaac, an interviewee with technical background also noted that such reflexivity may be discouraged because of the questions it can provoke during the peer review process.

9.3 Main Contribution: Combining STS and Scientometrics

This section discusses my findings related to the Third Research Question, which reflects on the project's mixed methods approach the thesis' main contribution. I asked how combining STS and scientometrics can help explore the development of and differences among approaches to geosocial research (Research Question 3.1) and how we can evaluate the affordances of computational or scientometric methods 'for STS' (Research Question 3.2). As Table 9.1 depicts, I combined STS and scientometrics for three main reasons. I aimed to contribute to literatures which have recently called to combine STS and computational analysis or scientometrics in practice (e.g. Wyatt, Milojević, et al., 2017; Marres and Gerlitz, 2016; Cambrosio, Bourret, et al., 2014; cf. Neff et al., 2017; cf. Vertesi et al., 2019b) and digital STS' call to 'import' digital methods to STS (Vertesi et al., 2019b). Section 9.3.1 summarises my conceptual contribution to the above literatures by highlighting the need to assess the affordances of digital or computational methods 'for STS' in light of the interpretative context, and empirical contribution showing how diverse combinations can help study diverse aspects of knowledge practices, also contrasting my project with prior computational STS studies of biomedicine.

I also explore how STS and scientometrics - two traditions of thought that have developed in increasingly different directions since the 1980s - can help compose a common world, in line with STS' 'compositionist' agendas (e.g. Latour, 2010; Haraway, 2016) - in section 9.3.2, which reflects on how methods mixing can help compose knowledge which benefits from the partial perspectives methods afford and also acknowledges the uncertainty of knowledge.

9.3.1 Mixed Methods Study of Geosocial Research

As section 2.6.2 argued, most existing studies which reflect on the affordances of computational methods 'for STS' or related fields, such as digital sociology or digital anthropology evaluate methods in light of epistemological frameworks, such as actor-network theory (e.g. Cambrosio, Bourret, et al., 2014; Venturini, Munk, and Jacomy, 2019) or ethnographic field work (theories) (e.g. Munk, 2019). For example, they reflect on the types of nodes, edges and temporalities diverse network analysis methods afford (e.g. Cambrosio, Bourret, et al., 2014; Bourret et al., 2006); and the way analyses may benefit from the multimodality of digital data, such as the relational information online digital traces contain (e.g. hyperlinks, reactions, mentions) (Munk, 2019) in light of ANT's insights about heterogeneous agencies and the dynamic unfolding of knowledge practices. They also emphasize the need to carefully choose data which forms part of computational analyses either by manual curation (Munk, 2019) or the diversification of data sources (Cambrosio, Bourret, et al., 2014).

I developed my methodological approach in light of the above insights. However, answering Research Question 3.2 - which asks how we can evaluate the affordances of computational methods 'for STS' - I argue for better emphasizing computational methods' contingency on the interpretative context when assessing their strengths and weaknesses 'for STS': including the conceptual framework, the characteristics of the research practice under investigation, the research questions and prior findings (cf. Lury and Wakeford, 2012). Shaped by publication norms, papers often discuss how methods help study research questions (similar to this thesis' Research Question 3.1). However, the way research questions - and the broader interpretative context - enable the development and use of certain methods has been discussed less extensively.

I argue that such a contextual development and evaluation of methods helps develop mixed methods research without restricting methods development to a type of method - such as visual or statistical, network analysis or descriptive statistics - *a priori*. This flexibility, in turn, helps explore how computational data analysis and STS or other interpretative social scientific research traditions *do* or *could* meet. As section 2.6.2 discussed, in a recent study Elgaard Jensen (2020) reached a similar conclusion by arguing that collaborative digital STS projects can succeed by introducing project specific criteria for *evaluating* combinations of computational or digital methods and STS. In my project, the interpretative context - including characteristics of the scientific practice under investigation, similar to Elgaard Jensen's (2020) project specific criteria, as well as previous findings and the conceptual framework - informed both my *evaluation* and *development* of computational methods.

As I will discuss below in more detail, supported by the interpretative context, I benefited from the affordances of structural, statistical heterogeneous network analyses - often critiqued by STS scholars for obscuring the dynamism of research practices and heterogeneous agencies therein highlighted by ANT; descriptive statistics, often disregarded by STS scholars who prefer network analysis methods; and illustrated that affordances of heterogeneous network analyses - popular in computational STS for its resonances with ANT - also depend on the interpretative context.

As section 2.6 discussed, relatively few STS studies focus on the social sciences and humanities, and even less combine scientometrics and STS. Most studies which combined STS and scientometrics to date studied biomedical research and primarily used computational methods to explore how heterogeneous actors coordinate such large scale research collaborations. These studies primarily explored the affordances of visual (e.g. Bourret et al., 2006; Cambrosio, Bourret, et al., 2014) and statistical (e.g. Shi, Foster, and Evans, 2015) heterogeneous, temporal network analyses to help combine insights from STS with scientometrics. Studying heterogeneous actors' roles in achieving collective, coordinated research over time with these methods can help study epistemic shifts - such as changes in specialisations (e.g. Bourret et al., 2006; Navon and Shwed, 2012) and the impact of computational methods on research practice (e.g. Cambrosio, Bourret, et al., 2014), or even predict the development of biomedicine (Shi, Foster, and Evans, 2015).

Like these authors, I combined methods to study relational practices - such as collaboration - which help 'hold together' approaches to geosocial research. However, as figure 9.1 (answering Research Question 3.1 which asked how mixed method can help study the difference and differentiation of geosocial research approaches) depicts, to study the diversification and diversity of geosocial research, I also sought methods to study approaches' differentiation and to identify comparative differences among them. The same practices that 'hold together' geosocial research approaches - such as collaboration, reflexivity or experimental data analysis - also help differentiate approaches from one another. Figure summarises 9.2 how the interpretative context informed my use of each computational method.

Next, I summarise how assessing the affordances of computational methods in light of the interpretative context helped me develop diverse mixed methods approaches, and how each method helped study the development and differentiation of geosocial research. My discussion will follow the categorisation outlined by the font color in figures 9.1 and 9.2: I will discuss my use of structural, statistical network analyses; descriptive statistics (including the combination of TF-IDF and thematic analysis); and visual (heterogeneous) network analyses.

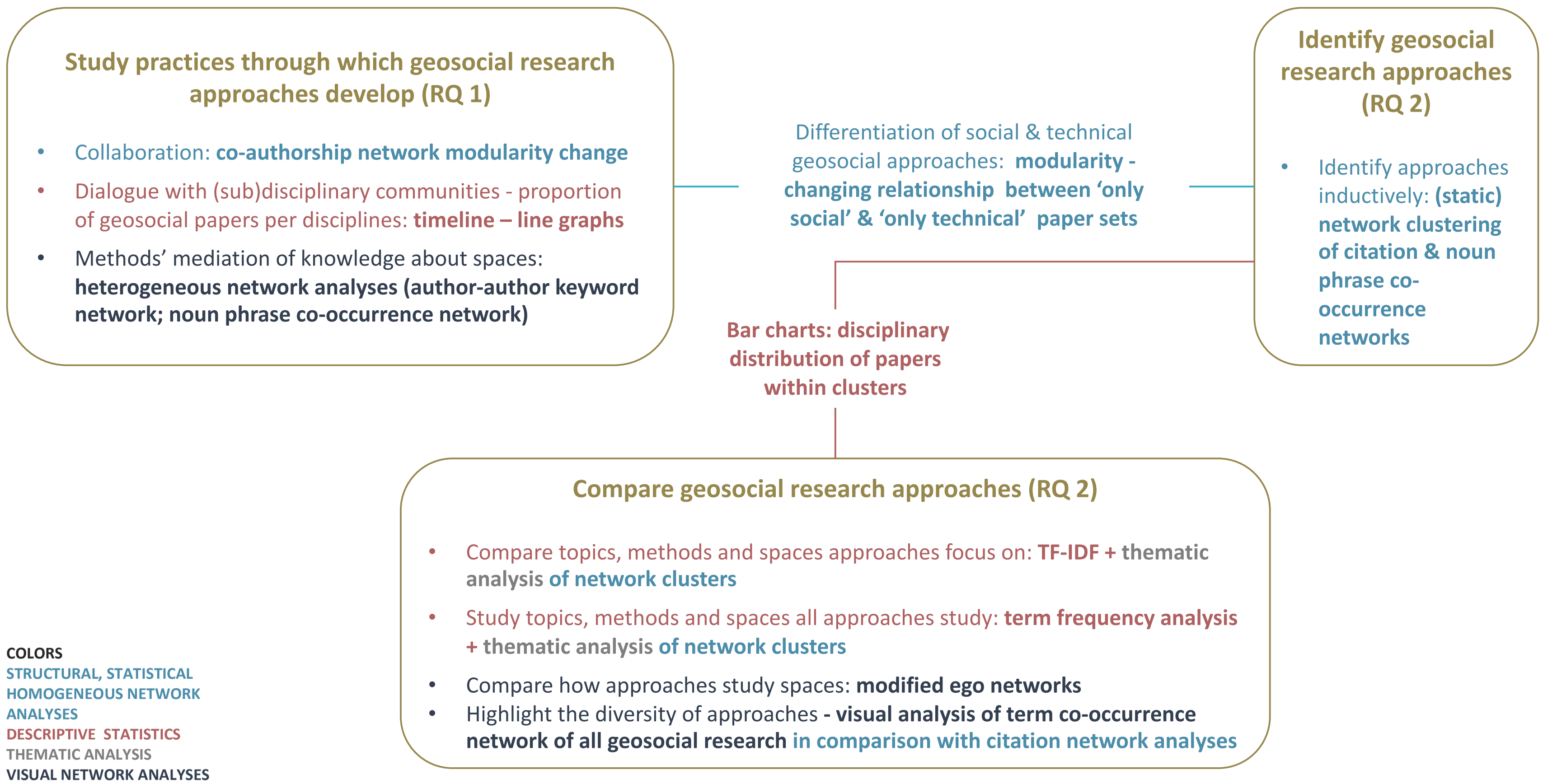


Figure 9.1: Scientometrics and Research Questions / relationality of knowledge

Figure 9.2: How interview analysis and conceptual framework shape scientometrics

9.3.1.1 Structural, Homogeneous Network Analyses

STS scholars have critiqued structural, homogeneous network analysis theories such as ANT for obscuring heterogeneous agencies and the dynamism of scientific research (e.g. Cambrosio, Bourret, et al., 2014). I argued that their affordances 'for STS' are contingent on the interpretative context. Used with interviews, they helped me explore the differentiation of and differences among geosocial research approaches.

I conducted two types of structural, homogeneous network analyses. In Chapters Five and Six, I used network modularity to trace scholars' efforts to differentiate their geosocial research approaches through relational practices, suggested by interviews and by my conceptual framework. Chapter Five's co-authorship analysis used network modularity to trace the development of small scale collaborations - enabled by fewer relations among these tight-knit groups through collaboration using loose theoretical or methodological common ground - theorised through interviews. Network modularity could capture a relational 'configuration' (multiple tight-knit clusters loosely connected to each other) suggested by interviews. Both interviews and the network analysis suggested that geosocial research comprises a collection of small-scale collaborations, and interviews suggested that through setting up their own geosocial laboratories, scholars develop distinct geosocial research approaches.

In Chapter Six, I compared the network modularity of counterfactual author-bibliographic coupling networks which omitted edges among geosocial papers published in 'only social' and 'only computational' journals to the modularity of networks which omitted equal numbers of random edges (and thus were statistically comparable), to trace social scientists' efforts to differentiate their approach from technical geosocial approaches. I studied how scientometric traces of relational practices, which previous findings suggested render geosocial research more diverse, render the bibliographic coupling network of geosocial research more or less cohesive over time. These included two sets of relational practices. Firstly, social scientists' search for and belonging to sub-disciplinary communities who welcome their geosocial research, operationalised as the (decreasing) proportion of edges between 'only social' and 'only computational' journals over time. Secondly, relational practices between social and technical geosocial scholars identified through interviews, such as collaboration amid loose common ground, shared interest in computational methods, social scientists' efforts to modify these methods and critical reading of technical scholarship. I operationalised these through studying how edges between 'only social' and 'only computational' journals impact the network's cohesion or modularity over time.

Similar to studying collaboration through co-authorship network analysis, I operationalised these relational practices as network relations through the practices' assumed scientometric traces. However, in contrast to the co-author network analysis which did not analytically differentiate network edges, the structural network analyses in Chapter Six studied relations I assumed render the author bibliographic coupling network more cohesive - and geosocial research more diverse - over time. Furthermore, in contrast to the co-authorship network analysis which did not make any assumptions about the number or type of geosocial approaches, based on previous findings, Chapter Six's analysis assumed that social and technical geosocial research differentiate over time. However, it did not make a strong assumption about the internal coherence of social and technical approaches or about how the number of geosocial research approaches one could scientometrically identify.

The second type of structural network analyses I performed were static network clustering in Chapters Seven and Eight. I clustered the author-bibliographic coupling networks *G3* (comprising all geosocial papers) and *G4* (comprising 'only social' and 'only computational' geosocial papers) and the term co-occurrence network of all geosocial papers to identify and compare the scientometric traces of geosocial research approaches. Resonating with my conceptual framework which assumed that geosocial research approaches can only be identified in comparison to each other, network clustering helped identify densely connected sets of nodes that are less strongly connected to the rest of the network. Interpreting network clusters' scientometric analyses in light of interviews helped explore their 'sociological relevance' (cf. Cambrosio, Bourret, et al., 2014; Held, Laudel, and Glaser, 2020). I identified social, technical, geographic approaches (also captured through interviews), as well as a fourth approach which comparatively focused studying urbanism and spatial diffusion using computational methods. Comparing the diverse network clusterings helped diversify my narrative about geosocial approaches, illustrate their diversity and my findings' contingency on units of analyses. While the term map highlighted the difference between geosocial research that does and does not use computational methods, clustering *G3* helped illustrate the diverse ways computational methods are used and the joint interest in urban research across approaches, while clustering *G4* helped highlight social approaches' focus on diverse collective practices and societal actors and technical scholarship's relative focus on methods development and the study of spatial scales. Next, I discuss my use of descriptive statistics.

9.3.1.2 Descriptive Statistics

Mixed methods STS literature primarily focuses on network analyses and disregards the potential of descriptive statistics (cf. Moats and Borra, 2018). Used in conjunction with interviews, descriptive statistics helped me identify and compare geosocial research approaches, as well as trace their development through scholars' search for disciplinary and sub-disciplinary communities they could belong to, which valued their geosocial research. Identifying geosocial approaches through studying the disciplinary distribution of papers in clusters of *G3* and *G4* hinged on interpreting disciplinary classification of papers with respect to geosocial research approaches, based on the interview finding that scholars publish in journals associated with their disciplines for research evaluation purposes, belonging and as they seek publication venues which welcome their research. In addition, the TF-IDF analysis of terms in the abstracts and titles of papers in the clusters of *G3* and *G4* - used in conjunction with thematic analysis informed by my conceptual framework - helped explore differences among geosocial research approaches because it captured comparative differences between a relatively large number of papers pertaining to each approach. Finally, using different descriptive statistical normalisation methods helped me plot the proportion of geosocial papers per Broad Disciplinary Categories with respect to three paper sets - all geosocial research, disciplinary journals which publish geosocial research and all disciplinary journals. Combined with interviews, this helped study the rise of social geosocial research and interviewees search for belonging to social scientific (sub)-disciplinary communities who value thier geosocial research. Next, I discuss the visual network analyses' contingency on the interpretative context.

9.3.1.3 Visual (Heterogeneous) Network Analyses

Informed by interviews, I used visual, homogeneous network analysis to compare how social and technical geosocial research study cities, citizens and use network methods (highlighting the former's focus on forms of participation and the latter's focus on computable patterns), as well as to illustrate my methods' contingency on the data analysis infrastructure. I also used visual, heterogeneous network analysis, popular among mixed methods studies of biomedicine. Contrasting my use of visual heterogeneous network analyses with heterogeneous network analyses used to study biomedicine helps illustrate the contingency of the method's affordances 'for STS' on the interpretative context.

Analysts of biomedicine mapped heterogeneous actors pertaining to large scale collaborative practices whose boundaries they could conceptually and computationally delineate. Bourret et al. (2006) and Cambrosio, Bourret, et al. (2014) studied large scale, collaborative biomedical research practices: the becoming of a large institute which conducts biomedical research, clinical work and policy making; and submitting molecules to an international conference which sought to develop a novel nomenclature for them. Knowing the boundaries of the institution and the conference at specific times allowed them to select sets of relevant actors. Based on fieldwork, the authors also empirically understood the relevance of these analytical units for studying shifts in the collective coordination of biomedical research. Shi, Foster, and Evans (2015) studied all papers listed in PUBMED - suited to their exploratory research which aimed to predict the development of biomedical research at large. These studies highlighted the affordances of heterogeneous network analysis to study how diverse actors help coordinate the analytically delineated collaborative research practices.

In contrast, as Chapters Five and Six argued, neither geosocial research nor approaches to it are coordinated or collaborative practices. Defining the boundaries and reflecting on the nature and role of collectives in geosocial research - such as geosocial research approaches or disciplinary communities - was a goal, not a starting point for my project. In addition, PUBMED - a scientometric database which explores biomedical research - tags papers with keywords that define diverse non-human entities, such as molecules and research methods. In contrast, to study geosocial research I could only draw on 'basic scientometric' information, such as the abstract and titles of papers, disciplinary classification, authors and institutions - available through most scientometric databases such as Scopus, Web of Science, Dimensions and Microsoft Academic. Thus, tracing heterogeneous relations requires the time-intensive task of manually categorising terms.

In part due to the differences in analytical unit and data availability, in contrast to the above previous mixed methods STS studies which used heterogeneous network analysis to explore heterogeneous coordination processes, my research highlighted the potential of heterogeneous network analysis to study the mediated nature of knowledge. Chapter

Eight used visual, heterogeneous network analyses to explore how methods mediate knowledge about spaces - core to the development of geosocial research approaches. Informed by prior findings and my conceptual framework which suggested that comparing geosocial research which uses specific computational methods is a relevant unit of analysis and the popularity of machine learning (ML), I studied heterogeneous networks of geosocial papers which use social network analysis (SNA) or ML. I conducted two types of heterogeneous network analyses. Using author-keyword heterogeneous networks, I compared how scholars position the topics ML and SNA-related geosocial research explore with respect to previous research, and whether they enable a new set of scholars to conduct geosocial research. In addition, studying the noun phrase co-occurrence networks of ML and SNA-related geosocial papers (where I classified terms as related to 'methods', 'spaces' or 'other') helped explore how methods mediate knowledge about spaces in more detail. I found that ML-related geosocial papers refer to scale related spatial categories - also shared across papers, and thus potentially coordinating such research - more often than SNA-related geosocial papers. The latter, in turn tend to study specific and different locations.

Altogether, answering Research Question 3.2, I conceptually contribute to literature which aims to link STS and scientometrics by highlighting the importance of assessing methods' affordances 'for STS' in light of the interpretative context. Empirically, I contribute by showing how the diverse combinations of STS concepts, interview analysis and scientometric data analysis can inform a single case study. Many existing studies that combine STS or sociology of science and scientometrics use either statistical or visual analyses, but rarely a combination of both. In addition, most advocate for the use of (heterogeneous) network analysis and disregard the opportunities associated with descriptive statistics. Given the importance of evaluating methods in the interpretative context, using diverse mixed methods solutions as part of a single case study helps reflect on the strengths and weaknesses of diverse computational methods 'for STS'. Combined with interviews, structural, homogeneous network analysis and descriptive statistics helped me study the development of geosocial research approaches; TF-IDF helped me compare approaches; and visual network analyses helped study both.

9.3.2 Composing Mixed Methods Knowledge: Surprise and Certainty

This section discusses how I explored diverse combinations of STS, interviews and scientometrics to ensure that findings obtained with the diverse methods related but did not fully converge. This was important for two main reasons. Firstly, each method affords a partial perspective on scientific practice. Mixed methods projects can aim to align the analytical units methods capture or let each method enact the research object in unique ways. I found it important to experiment with both of these approaches to explore how STS and scientometrics *could* 'co-compose' knowledge. Secondly, many contemporary challenges, such as the climate crisis require knowledge that incorporates diverse ways of knowing. As outlined below, during the writing process my interview and scientometric analyses increasingly converged, which I felt posed risks to creating knowledge and writing narratives that acknowledge uncertainty and are open to new perspectives on the research object(s) (in this case, geosocial research or methods mixing).

On some occasions I attempted to align interview and scientometric units of analyses. In these instances, methods mixing helped study scientific practice in depth and on larger scales. In other instances I used the methods inductively. The outcomes of these findings helped highlight the partial perspective methods offer and generate 'surprising' findings scientometrically. However, I argue that mixed-methods *research practice* has benefits that differ from mixed-methods *findings*. Mixed methods research practice itself can help create methodological dialogues. For example, as I discuss below, calculation acts themselves can inform qualitative analysis even if the scientometric analytical units do not capture research practice. These instances may require the analyst(s) to actively remember the discrepancy between scientometric and practiced units and relations.

Throughout the project, I developed the argument thorough iterating scientometric and interview analyses. Over the course of my research and the thesis writing, in many instances, scientometric and interview findings converged into coherent narratives. For example, I found it surprising that using the Leiden community detection algorithm - in Chapter Six's scientometric analysis which traced the differentiation of 'only social' and 'only computational' geosocial papers, as well as the clustering of the author-bibliographic coupling network *G4* in Chapter Seven - immediately resonated with interview findings. This resonance evoked mixed feelings in me. I was content about the development of a coherent narrative, given the University of Nottingham's requirement to create a monograph. At the same time, I worried that my use of scientometrics could lead to confirmation bias, and obscure other, relevant knowledge about geosocial research. Although my scientometric analyses allowed me to differentiate among approaches to geosocial research, I increasingly felt the need to highlight the diversity of these approaches (akin to critiques of regional thinking in the social sciences (cf. Elgaard Jensen, 2020) and show that the differences I found among them are not absolute, but rather are contingent on my data analysis infrastructure and analytical choices. To this end, for example, I explored how two methods - machine learning and social network analyses - mediate knowledge about spaces, and complemented the author-bibliographic coupling network analysis with noun phrase co-occurrence network analysis in Chapter Eight. Next, I discuss the eight ways I aligned scientometric and interview analytical units, depicted by figure 9.3.

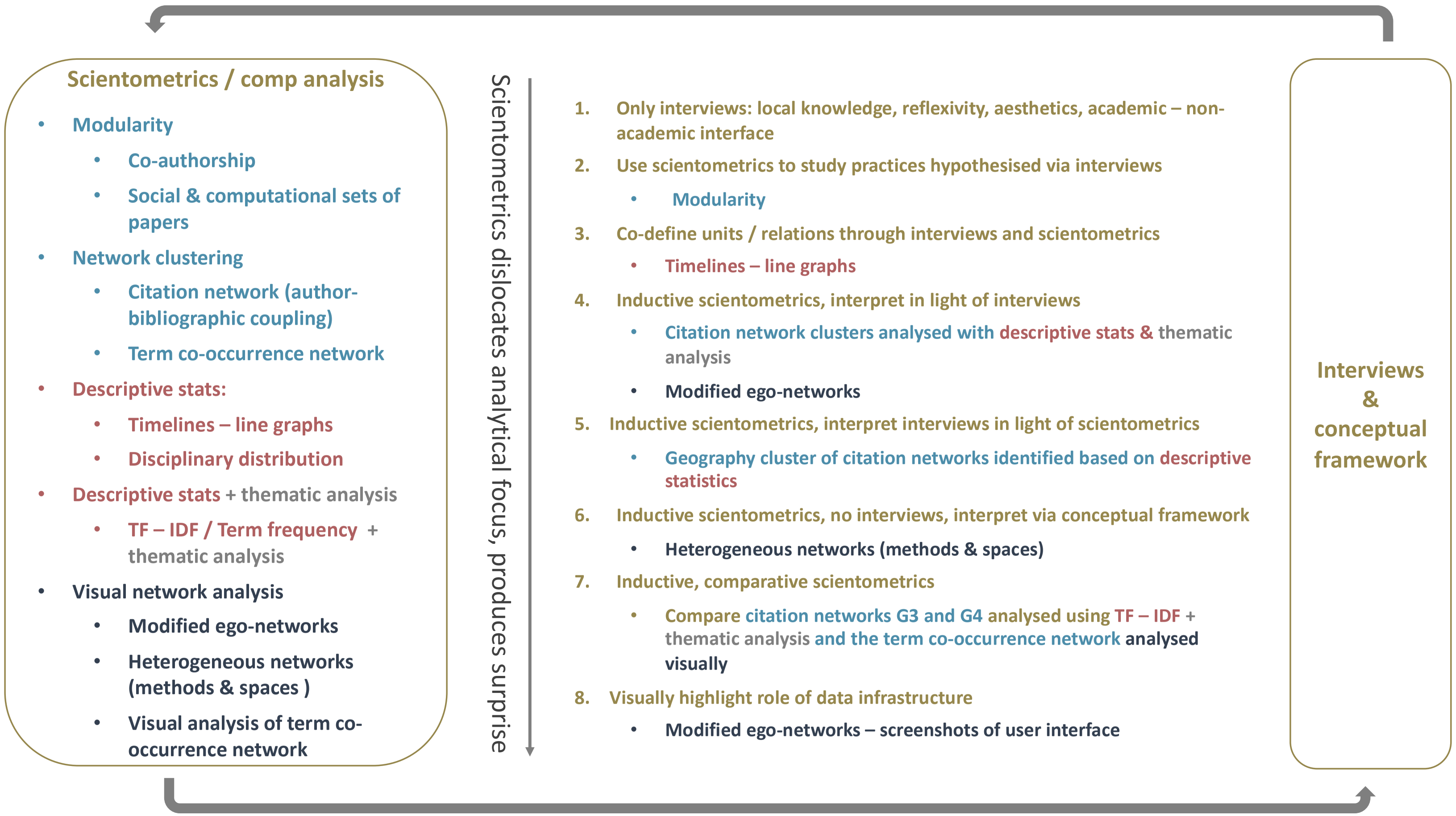


Figure 9.3: How scientometrics shapes interview analysis

Point 1 depicts examples where I used interviews largely independent of scientometrics. These include my analysis of interviewees' narratives about the need to combine computational data analysis and socio-spatial interpretation, as well as their aesthetic interests, concurrent academic and non-academic activities, reflexivity, and use of local knowledge to study specific locations.

Points 2-3 illustrate two main ways I aimed to align the analytical units and relations captured with scientometrics and interviews. As the second point depicts and section 9.3.1.1 explained, the co-authorship network analysis in Chapter Five and the analysis of the changing relationship between 'only social' and 'only computational' geosocial papers in Chapter Six explored relational practices first hypothesised through interviews. In addition, in both cases, scientometrics also affected my interpretation of the interview findings in subtle ways. Scientometrics did not change my interview analysis, but shaped the narrative's focus by prompting me to think about interviewees' narratives more in terms of the units and relations highlighted through network analyses. The co-authorship network analysis highlighted the importance of the concurrent development of small collaborations or geosocial laboratories. The citation network analysis prompted me to reflect on social scientists' efforts to differentiate their geosocial research from computational approaches in depth.

As the third point depicts and section 9.3.1.2 discussed, I plotted the proportion (fraction) of geosocial papers per Broad Disciplinary Categories over time with respect to three paper sets: all geosocial papers, all papers published in disciplinary journals, and all papers in journals which publish geosocial research. To calculate the proportions, I normalised geosocial paper count in each Broad Disciplinary Category: I divided it by the count of papers in the three sets. This calculation prompted me to reflect on what the count of paper sets in the fraction's denominator captured. More specifically, the practice of normalising data (dividing with respect to diverse counts) prompted me to better reflect on interviewees' relations to diverse research collectives and communities. Through iterating their analyses with interviews, I developed the argument that social geosocial research is on the rise and that social geosocial scholars are seeking (sub)-disciplinary communities who value their geosocial research and do not belong to the geosocial research collective. It helped me argue that latter is a collection of approaches, rather than a coordinated community. However, whilst iterating interview analysis and the line graphs, I had to remind myself not to equate the count of the paper sets in the denominator with 'sociologically' meaningful collectives - that sets of papers published in specific journals do not (necessarily) correspond to research communities. This shows that (digital) counting practices which can quickly highlight diverse types of relations - such as through diverse normalisation methods - can prompt qualitative reflection about the nature of research practice even if computational units of analyses do not correspond to units that capture scientific practice.

Highlighting the importance of sub-disciplinary communities interviewees belong to through combining the line graphs and interviews helped illustrate the diversity of disciplines and de-essentialise the disciplinary categorisation - the Broad Disciplinary Categories - which underpins several scientometric analyses in this thesis. Although, as discussed earlier, the relevance of Broad Disciplinary Categories was supported by interviews and they proved analytically useful, I worried that using them as aggregate units across several scientometric analyses depict disciplines as homogeneous and fixed. Although this tension is present in my scientometric analyses, depicting the proportion of geosocial papers with respect to paper sets that capture disciplinary, sub-disciplinary and 'non-disciplinary' (geosocial research) collectives helped illustrate the heterogeneity of disciplines and the importance of sub-disciplinary communities to which interviewees belong.

Points 4-7 illustrate ways I used scientometrics in a more inductive fashion, to diversify my study of geosocial research approaches by studying analytical units scientometrically that I had not hypothesised through interviews. As section 2.6.3.4 discussed, STS and anthropology scholars have proposed ways to use computational methods to produce 'surprising' insights not obtained through interviews or fieldwork. Cambrosio, Bourret, et al. (2014) and Munk (2019) note that the social scientific relevance of patterns identified with computational data analysis is uncertain but note that computational data analysis has the potential to produce surprising findings. As points 4-7 illustrate, I used scientometrics inductively to produce surprises in various ways.

Point 4 explains two inductive scientometric analyses whose interpretation hinged on interviews and conceptual framework. Firstly, I interpreted the findings of the TF-IDF based comparison of citation network clusters in light of the thematic analysis informed by the conceptual framework and interview findings about participants' reflexivity. The TF-IDF analysis, in turn - which, as section 3.4.6 explained, highlights relative differences among units compared - highlighted differences in the spatial units geosocial research approaches study and the methods they use, which enriched my interview analysis about the difference among approaches. In addition, the TF-IDF analysis prompted me to reflect on the contingency of my findings on the comparison units. This then prompted me to compare my analysis of geosocial research approaches using the two citation networks *G3* and G4 depicted by point 7, which better highlighted the diversity of geosocial research approaches. Secondly, I interpreted the modified ego-network analyses of the terms of 'city', 'citizen' and 'network' in light of interview findings about participants' reflexivity. The findings of the modified ego-network analyses mainly aligned with my interview findings.

In contrast to interpreting scientometrics in light of interviews, as point 5 depicts, I identified a third approach to geosocial research - geographic geosocial research - by inductively using scientometrics, and reflected on interview quotes in light of this scientometric finding. Some interviewees' experiences supported the finding that geographic geosocial research is a distinct approach, but the interview data was not extensive enough to hypothesise that geographic geosocial research was a distinct approach. Thus, scientometrics helped produce an insight interviews did not afford.

Point 6 explains a third way I interpreted inductive scientometric findings: I interpreted the heterogeneous network analyses, which explored how methods mediate knowledge about spaces, largely independent of interviews. I compared the heterogeneous networks of SNA and ML-related geosocial research without exploring geosocial research using ML and SNA through interviews. As Chapter Seven explained, this analysis produced a surprising result because it shifted my attention from my original question about how methods mediate knowledge about spaces to asking how spatial units might coordinate geosocial research.

As discussed above, point 7 depicts my efforts to produce comparative, inductive scientometrics analyses. I complemented the analysis of the clusters of the author-bibliographic coupling network G3 with that of the analysis of network *G4* as well as the noun-phrase co-occurrence network of geosocial papers. As Chapter Eight explained, diversifying scientometric units helped diversify knowledge about geosocial research and helped highlight my findings' contingency on my analytical decisions and data analysis infrastructure. These analyses helped highlight the diversity of geosocial research approaches and de-essentialise them.

Finally, point 8 illustrates my efforts to highlight the contingency of my scientometric findings on the data infrastructure by visually highlighting the modified ego-network analysis' contingency on analytical choices using screenshots of the software user interface.

In sum, to benefit from the partial and complementary perspectives interviews, participant observation and scientometrics afford on geosocial research, I combined them both by aligning the analytical units I studied with them, as well as by using them separately, inductively to varying extents. I illustrated that while in some instances, scientometric and interview findings converged, I also produced findings with both methods that I could not have learned without the mixed methods approach.

Altogether, this section outlined the thesis' conceptual and empirical contribution to literature which explores the affordances of scientometric or computational methods 'for STS'. Empirically, I argued that diverse combinations of STS and scientometrics as part of a case study help study diverse relations that comprise knowledge practices and reflect on the strengths and weaknesses of diverse computational methods. Conceptually, I highlighted the importance of assessing the affordances of computational methods 'for STS' in light of the interpretative context. Next, I summarise findings relevant to research methods curriculum development in geography and more broadly in the social sciences.

9.4 Considerations for Curriculum Development

This section summarises six insights related to digital or computational research methods teaching in human geography or the social sciences more broadly based on the thesis' findings. Firstly, my project highlights the importance of providing introductory training for social science students in diverse computational methods - for example, exploratory data analysis, data visualisation, frequentist statistical inference, (agent based) modelling and network analyses. My efforts to combine STS and scientometrics without deciding about 'valid' approaches *a priori* and my interviewees' narratives about the importance of combining computational data analysis and socio-spatial interpretation for geosocial research as well as the importance of loose common ground when collaborating with researchers with complementary backgrounds highlight the value of basic knowledge (akin to interactional expertise Collins, 2004) about diverse computational data analysis methods. Thus, research methods modules could focus on the introduction and limited hands on experience with diverse types of methods.

Secondly, interviewees' aesthetic motivations (discussed in Chapter Four) to conduct geosocial research which combines computational data analysis and socio-spatial interpretation highlights the importance of using data and teaching data analysis methods (e.g. data visualisation methods) that students find aesthetically pleasing. Several interviewees stated that they find geotagged social media data aesthetically pleasing. For example, Anne noted the beauty of the 'chatter' element of social media posts which capture fleeting, everyday experiences. Several interviewees noted the beauty of pictures posted on social media. Thus, in research methods training, 'traditional' data sources like surveys and census can be complemented with web-based data that students encounter in their everyday lives, such as websites (as in controversy mapping) or social media posts. In addition, several interviewees noted the beauty of data visualisations. Thus, research methods training could have a data visualisation element.

Thirdly, informed by my finding about methods' mediation of spatial units, mixed-methods human geography education could place an emphasis on fostering reflexivity about the spatial units computational methods enable, juxtaposing computational data analysis with theories about space (e.g. Shelton et al., 2014). This can build on existing geographic scholarship which reflects on the spatial methods brought forth through diverse mapping techniques including GIS (e.g. Kwan, 2012) and participatory mapping (e.g. Pain, 2004).

Fourthly, similar to Dumit and Nafus (2018) my findings highlight the importance of exploratory data analysis (EDA) for scholars from diverse disciplines and across geosocial research approaches. As Chapter Five argued, interviewees across research groups experimented with data analysis methods in search for data patterns. Thus, EDA should be a core focus of mixed-methods training programs.

Fifthly, I present three recommendations for educational settings which attempt to foster collaboration and dialogues among students from diverse disciplines. Given the relevance of EDA across disciplinary research traditions, students can collaboratively conduct EDA, thus learning about different exploratory questions one can ask. Given the centrality of reflexivity for computational research across research traditions, teaching can also focus on fostering students' reflexivities about how data and analytical decisions shape knowledge (about spaces). Finally, given the benefits of local knowledge across geosocial research approaches, teaching can facilitate articulating the diverse roles local knowledge can play in geographic research, thus fostering discussions about why scholars from diverse backgrounds find it valuable.

Finally, I envision a hackathon-style method that can be organised as a short educational intervention for scholars with computational expertise from diverse disciplinary backgrounds, that puts the process of data analysis, rather than its outcomes at the centre. Oftentimes, hackathons focus on 'problem solving' or the production of outcomes. Focusing on the analysis process could allow participants to articulate their assumptions and analytical decisions and learn about those of scholars with complementary skills. I envision a short, fast paced event informed by four findings of the thesis: diverse interviewees' interest in experimentally searching for data patterns; the fast paced nature of computational work and the time shortage associated with academic research; my finding that reflecting on how analytical decisions and social media platforms shape data and knowledge about spaces is a core part of participants' geosocial research; and the finding that common ground - even in its loose form - is essential for the development of diverse approaches to geosocial research. A short, quick paced data analysis event which focuses on data analysis process rather than outcomes could fit with interviewees' interest to experiment with computational data analysis methods and reflect on the data analysis process. This, in turn can help identify shared values or points of interest as well as differences in approaches, whilst allowing participants to familiarise themselves with complementary approaches, enabling the formation of loose common ground which can foster and enrich the future development of diverse computational research. Next, I discuss how future work can complement this thesis' findings.

9.5 Future Research

This section concludes the thesis by discussing nine main avenues of future research. Firstly, future research could explore the diversity of geosocial research approaches this thesis identified in more detail. The heterogeneous network analyses which explored how computational methods mediate knowledge about spaces started this line of research and illustrated the potential to use scientometrics to identify finer grained differences, rather than broader patterns. For example, future research could cluster the 'technical', 'mixed' and 'geography' clusters of the author-bibliographic coupling network *G3*, and explore the noun phrase co-occurrence network of the same paper set to study the variety of computational data analysis.

Secondly, future research could explore how computational methods travel across disciplinary cultures and communities in more detail. Previous literature highlighted computational methods' ability to travel across disciplines and domains (e.g. Mackenzie and McNally, 2013; Knuuttila and Loettgers, 2014; Marres and Gerlitz, 2016). Similarly, my participants from diverse disciplinary backgrounds were interested in using computational data analysis methods. However, social scientist interviewees' narratives presented in section 5.3 suggest that using computational methods across disciplines requires active work. Several participants claim they altered computational methods to suit their analytical purposes, like Chase, who altered a computational data analysis method suggested by Kevin to fit his research purposes. Future research could explore how, in practice, scholars from diverse disciplines succeed to use computational methods and the challenges such work raises.

Thirdly, future research could follow up the heterogeneous network analysis presented in Chapter Eight by exploring how diverse spacial units coordinate or geosocial research project and geosocial research approaches. As I argued, the heterogeneous network analysis prompted a question about spaces' mediating role, in addition to how computational methods' mediate knowledge about spaces.

Fourthly, future research could explore how scholars who conduct geosocial research use their combined computational and social scientific skills in future projects. Given changes in social media platforms' geotagging policies discussed in section 2.4, geotagged social media research may lose popularity. However, my participants discussed their (aesthetic) interest in combining computational data analysis and social scientific research. Thus, future research could explore how scholars use the skills they developed for geosocial research for other research projects.

Fifthly, future research could explore if and how geosocial research changes the relationship among disciplinary or sub-disciplinary communities. As the Introduction outlined, in recent years university programs which specialise in exploring the interface between computational data analysis and social science became more popular. I argued that interviewees primarily associate themselves with disciplinary or sub-disciplinary communities, and develop geosocial research through dialogues with these communities. Chapter Six argued that social scientists are actively developing ways to use computational methods in line with their disciplinary heritage, and technical scholars' narratives discussed in Chapter Five suggested that computational scholars also seek to apply computational methods to new topics. These suggest that the shared interest in computational methods may shape the boundaries among disciplines in the long term.

Sixthly, future research could 'qualitatively' explore the shared author citations between the paper sets associated with 'social' and 'technical' geosocial research in Chapter Six, and the clusters of the author-bibliographic coupling networks *G3* and *G4* analysed in Chapters Seven and Eight. Studying the scholarship of jointly-cited authors could provide hints about (changes in) the nature of common ground between diverse approaches to geosocial research and benefit from computational methods' affordances to flexibly explore data at different levels of aggregation (cf. Munk and Elgaard Jensen, 2014; cf. Elgaard Jensen, 2020; cf. Shelton et al., 2014).

Seventhly, follow-up studies could explore the variability of the scientometric findings and the scientometric field as a function of scientometric data providers. As section 3.2.2 discussed, the Web of Science database this thesis used over represents papers from the natural sciences and engineering, and under represents arts, humanities and social sciences. Given that geosocial research sits at the intersection of computational and social scientific disciplines, future research could reproduce the scientometric analyses in this thesis using the Scopus database - or where necessary, equivalent analyses using on Scopus' disciplinary classification system - given Scopus' enhanced coverage of social sciences and humanities, or Microsoft Academic to benefit from open access scientometric data.

Eightly, future research could focus on exploring how diverse communities' situated knowledge can be brought into and negotiated through computational methods in geosocial research and projects which combine STS and scientometrics (e.g. the participatory scientometrics by Marres and de Rijcke, 2020). Such work would have a more normative starting point than this project. Computational research is trusted by many - 'big data' analyses are seen as solutions in business and across scientific disciplines. Such methods are often used to identify statistical regularities and numerically optimised outcomes. However, many contemporary societal challenges - such as the climate crisis and social sustainability - require knowledge that accounts for diverse groups' perspectives and the specificities of local communities and ecologies (e.g. Latour, 2013; Haraway, 2016). The interpretative social sciences, such as human geography and STS developed many methods to produce such knowledge. Thus, combining computational and interpretative scientific schools of thought - an including local knowledge into computational analyses - may become a core part of future politics.

Finally, future research could use scientometric data analyses and visualisations as discussion prompts with geosocial scholars. It could develop more and less ambiguous data visualisations to explore how these elicit participants' experiences. Anderson et al. (2009) use ambiguous data visualisations to elicit the everyday experiences of technology users. Like the interviewees of Cambrosio, Bourret, et al. (2014), my participants have experience with computational data analysis methods similar to the ones I used. Future research could explore how scholars from diverse disciplines engage with data visualisations, and how they help elicit participants' reflections about their data practices. In addition, future research could explore the use of ambiguous visualisations as discussion prompts to contradict scholars' expectations about the nature of data visualisations, shaped by their computational research.

Bibliography

Aksnes, D., T. Olsen, and P. Seglen (1999). "Inadequacy of a journalbased research field delineation: Incomplete recovery of Norvegian microbiology articles in ISI's Microbiology field". In: *Proceedings of Seventh Conference of the International Society for Scientometrics and Informetrics. Colima, Mexico.* Pp. 1–10.

Anderson, Ken et al. (2009). "Numbers Have Qualities Too: Experiences with Ethno-Mining". In: *EPIC 2009*, pp. 123–140. ISBN: 0-9799094-2-4.

Antonijevic, Smiljana (2015). *Among Digital Humanists: An Ethnographic Study of Digital Kowledge Production.* Palgrave Macmillan. ISBN: 978-1-349-55335-8.

Balaban, Corina (2018). "Mobility as homelessness". In: *Learning and Teaching* 11.2, pp. 30–50. DOI: 10.3167/latiss.2018.110203.

Barabási, Albert-László (2012). "Chapter 1: Introduction". In: *Network Science.*

Barnes, Trevor J and Matthew W. Wilson (2014). "Big Data, social physics, and spatial analysis: The early years". In: *Big Data & Society* April–June, pp. 1–14. DOI: 10.1177/2053951714535365.

Barry, Andrew and Georgina Born (2013). "Interdisciplinarity: Reconfigurations of the Social and Natural Sciences". In: *Interdisciplinarity: Reconfigurations of the Social and Natural Sciences.* Ed. by Andrew Barry and Georgina Born. London: Routledge, pp. 1–56.

Bartlett, Andrew et al. (2018). "The locus of legitimate interpretation in Big Data sciences: Lessons for computational social science from -omic biology and high-energy physics". In: *Big Data & Society* January -, pp. 1–15. DOI: 10.1177/2053951718768831.

Bastian, M., S. Heymann, and M. Jacomy (2009). "Gephi: An Open Source Software for Exploring and Manipulating Networks". In: *Third International ICWSM Conference*, pp. 361–362. URL: https://www.aaai.org/ocs/index.php/ICWSM/09/paper/view/154.

Bates, Jo, Yu-Wei Lin, and Paula Goodale (2016). "Data journeys: Capturing the socio-material constitution of data objects and flows". In: *Big Data & Society* July–Decem, pp. 1–12. DOI: `10.1177/2053951716654502`.

Bateson, Gregory (1972). *Steps to an ecology of mind*. Jason Aronson Inc.

Beaulieu, A. and E. Simakova (2006). "Textured connectivity: An ethnographic approach to understanding the timescape of hyperlinks". In: *Cybermetrics* 10.1. DOI: `http://www.cindoc.csic.es/cybermetrics/vol10iss1.html`.

Beaulieu, Anne (2010). "Research Note: From co-location to co-presence: Shifts in the use of ethnography for the study of knowledge". In: *Social Studies of Science* 40.3, pp. 453–470. DOI: `10.1177/0306312709359219`.

Beaulieu, Anne, Andrea Scharnhorst, and Paul Wouters (2007). "Not Another Case Study: A Middle-Range Interrogation of Ethnographic Case Studies in the Exploration of E-science". In: *Science, Technology, & Human Values* 32.6, pp. 672–692. DOI: `https://doi.org/10.1177/0162243907306188`.

Becher, Tony (1990). "The Counter-Culture of Specialisation". In: *European Journal of Education* 25.3, pp. 333–346. DOI: `10.2307/1503322`. URL: `https://www.jstor.org/stable/1503322`.

Benton, Joshua (2019). *Twitter is removing precise-location tagging on tweets — a small win for privacy but a small loss for journalists and researchers.*

Bergmann, Luke, Eric Sheppard, and Paul S. Plummer (2009). "Capitalism beyond harmonious equilibrium: Mathematics as if human agency mattered". In: *Environment and Planning A* 41.2, pp. 265–283. ISSN: 0308518X. DOI: `10.1068/a411`.

Birkbak, Andreas, Morten Krogh Petersen, and Torben Elgaard Jensen (2015). "Critical Proximity as a Methodological Move in Techno-Anthropology". In: *Techné: Research in Philosophy and Technology* 19.2, pp. 266–290. DOI: `10.5840/techne201591138`.

Borges Rey, Eddy (2017). "Towards an epistemology of data journalism in the devolved nations of the UK: Changes and continuities in materiality, performativity and reflexivity". In: *Journalism*. URL: `https://doi.org/10.1177/1464884917693864`.

Bornmann, Lutz and Daniel Hans-Dieter (2008). "What do citation counts measure? A review of studies on citing behavior". In: *Journal of Documentation* 64.1, pp. 45–80. DOI: `10.1108/00220410810844150`.

Bornmann, Lutz and Rüdiger Mutz (2015). "Growth rates of modern science: A bibliometric analysis based on the number of publications and cited references". In: *Journal of the Association for Information Science & Technology* 66.11, pp. 2215–2222. URL: `https://doi.org/10.1002/asi.23329`.

Bourret, Pascale et al. (2006). “A New Clinical Collective for French Cancer Genetics. A Heterogeneous Mapping Analysis”. In: *Science, Technology, & Human Values* 31.4, pp. 431–464. DOI: 10.1177/0162243906287545.

Bowker, Geoffrey (2010). “A Plea for Pleats”. In: *Deleuzian Intersections: Science, Technology, Anthropology.* Ed. by Casper Bruun Jensen and Kjetil Rödje. New York: Berghahn, pp. 123–39.

Bowker, Geoffrey C. (2005). *Memory Practices in the Sciences.* MIT Press.

Boy, John D. and Justus Uitermark (2017). “Reassembling the city through Instagram”. In: *Transactions of the Institute of British Geographers* 42.4, pp. 612–624. DOI: 10.1111/tran.12185.

Braun, V and V Clarke (2012). “Thematic Analysis”. In: *APA handbook of research methods in psychology, Vol 2: Research designs: Quantitative, qualitative, neuropsychological, and biological.* Ed. by H. Cooper et al., pp. 57–71. DOI: 10.1037/13620-004.

Brindley, P., J. Goulding, and M.L. Wilson (2014). “A data driven approach to mapping urban neighbourhoods”. In: *Proceedings of the 22nd ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems (Dallas, Texas, USA, November 4-7th)*, pp. 437–440.

Bryman, Alan (2006). “Integrating quantitative and qualitative research: How is it done?” In: *Qualitative Research* 6.1, pp. 97–113. ISSN: 14687941. DOI: 10.1177/1468794106058877.

Burri, Regula Valérie (2008). “Doing Distinctions: Boundary Work and Symbolic Capital in Radiology”. In: *Social Studies of Science* 38.1, pp. 35–62.

Callon, Michel et al. (1983). “From translations to problematic networks: An introduction to co-word analysis”. In: *Social Science Information* 22.2, pp. 191–235.

Cambrosio, A., P. Keating, and A. Mogoutov (2004). “Mapping collaborative work and innovation in biomedicine: A computer-assisted analysis of antibody reagent workshops.” In: *Social Studies of Science* 34.3, pp. 325–364. DOI: 10.1177/0306312704043767.

Cambrosio, Alberto, Pascale Bourret, et al. (2014). “Big Data and the Collective Turn in Biomedicine How Should We Analyze Post-genomic Practices?” In: *TECNOSCIENZA Italian Journal of Science & Technology Studies* 5.1, pp. 11–42.

Clarivate (2020). *Research Areas (Categories / Classification).* URL: https://images.webofknowledge.com/images/help/WOS/hp_research_areas_easca.html.

Clauset, Aaron, M. E. J. Newman, and Cristopher Moore (2004). "Finding community structure in very large networks". In: *Physical Review E* 70.6, p. 066111. DOI: `https://doi.org/10.1103/PhysRevE.70.066111066111`.

Collins, Harry (2004). "Interactional expertise as a third kind of knowledge". In: *Phenomenology and the Cognitive Sciences* 3.2, pp. 125–143.

Crampton, Jeremy W (2010). *Mapping: A Critical Introduction to Cartography and GIS.* Wiley-Blakcwell. ISBN: 978-1-405-12172-9.

– (2011). "Cartographic calculations of territory". In: *Progress in Human Geography* 35.1, pp. 92–103. ISSN: 0309-1325. DOI: `10.1177/0309132509358474`.

Cronin, B (1982). "Norms and functions in citation: the view of journal editors and referees in psychology". In: *Social Science Information Studies* 2, pp. 65–78.

Csardi, G. (n.d.[a]). *Generate random graphs according to the Erdos-Renyi model.*

– (n.d.[b]). *R Igraph Modularity Function.* URL: `https://igraph.org/r/doc/modularity.igraph.html`.

Csardi, G. and T. Nepusz (2006). "The igraph software package for complex network research". In: *Interjournal, Complex Systems*, p. 1695.

Currie, Morgan et al. (2016). "The conundrum of police officer-involved homicides: Counter-data in Los Angeles County". In: *Big Data & Society* July–Decem, pp. 1–14. DOI: `10.1177/2053951716663566`.

D'Ignazio, Catherine and Lauren F. Klein (2020). *Data Feminism.* MIT Press.

Dalton, Craig M, Linnet Taylor, and Jim Thatcher (2016). "Critical Data Studies: A dialog on data and space". In: *Big Data & Society* January -, pp. 1–9. DOI: `10.1177/2053951716648346`.

Dalton, Craig M and Jim Thatcher (2015). "Inflated granularity: Spatial "Big Data" and geodemographics". In: *Big Data & Society* July–Decem, pp. 1–15. DOI: `10.1177/2053951715601144`.

Daston, Lorraine and Peter Galison (1992). "The Image of Objectivity". In: *Representations* 40.1, pp. 81–128.

De Rijcke, S., P. Wouters, et al. (2016). "Evaluation practices and effects of indicator use - A literature review". In: *Research Evaluation* 25.2, pp. 161–169.

De Rijcke, Sarah, T. Holtrop, et al. (2019). "Evaluative Inquiry: Engaging Reaearch Evakuation Analytically and Strategically". In: *fteval Journal for Research and Technology Policy Evaluation* 48, pp. 176–182. DOI: `10.22163/fteval.2019.386`.

DeLyser, Dydia and Daniel Sui (2012). “Crossing the qualitative-quantitative divide II: Inventive approaches to big data, mobile methods and rhythmanalysis”. In: *Progress in Human Geography* 37.2, pp. 293–305.

Dourish, Paul and Edgar Gómez Cruz (2018). “Datafication and data fiction: Narrating data and narrating with data”. In: *Big Data & Society* July–Decem, pp. 1–10. DOI: 10.1177/2053951718784083.

Dumit, Joseph and Dawn Nafus (2018). “The other ninety per cent: thinking with data science, creating data studies”. In: *Ethnography for a data-saturated world*. Ed. by Hannah Knox and Dawn Nafus. Manchester University Press.

Edge, David and Roy MacLeod (1986). “Editorial”. In: *Social Studies of Science* 16.1, pp. 3–8.

Edwards, Paul N. (2010). *A Vast Machine: Computer Models, Climate Data, and the Politics of Global Warming*. MIT Press.

Elgaard Jensen, Torben (1999). “The construction of scientific facts in social science - A case study in how data are combined in an anthropoloical text”. In: *Nordisk Psykologi* 51.4, pp. 260–278. DOI: https://doi.org/10.1080/00291463.1999.11863955.

– (2011). “The Map is the Territory - Cartographies, Categories and Categorizers in the Danish user driven innovation programs”. In: *Cartographies of Innovation new sites new actors*, pp. 1–7.

– (2012). “Intervention by Invitation: New Concerns and New Versions of the User in STS”. In: *Science & Technology Studies* 25.1, pp. 13–36. DOI: https://doi.org/10.23987/sts.55279.

– (2013). “Techno Anthropology - a new move in Science and Technology Studies”. In: *STS Encounters* 05.01, pp. 1–22.

– (2017). “Materiality, Performance and the Making of Professional Identity”. In: *Professional Identity and Social Work*. Ed. by S. A. Webb. 1st. Routledge. Chap. Vol 1. Pp. 51–61.

– (2019). “Exploring the knowledge practices of psychology: Reflections on a field study”. In: *Theory & Psychology* 29.4, pp. 466–483. DOI: 10.1177/095935431985.

– (2020). “Exploring the Trading Zones of Digital STS”. In: *STS Encounters - DASTS working paper series* 11.1, pp. 1–29.

Elgaard Jensen, Torben et al. (2019). “Identifying notions of environment in obesity research using a mixed methods approach”. In: *Obesity Reviews* 20.4, pp. 621–630. DOI: 10.1111/obr.12807.

Engeström, Yrjö, Reijo Miettinen, and Raija-Leena Punamäki (1999). *Perspectives on Activity Theory*. Cambridge Universtiy Press.

Erikson, Emily (2013). "Formalist and Relationalist Theory in Social Network Analysis". In: *Sociological Theory* 31.3, pp. 219–242.

ESRC (2013). *International Benchmarking Review of UK Human Geography*. Tech. rep. London: ESRC/RGS/AHRC.

Farr, Robert M. (1991). "The long past and the short history of social psychology". In: *European Journal of Social Psychology*. DOI: 10.1002/ejsp.2420210502.

Fortun, Kim et al. (2016). "Pushback: Critical data designers and pollution politics". In: *Big Data & Society* July–Decem, pp. 1–14. DOI: 10.1177/2053951716668903.

Fortunato, S et al. (2018). "Science of science". In: *Science* 359.6379, eaao0185. DOI: 10.1126/science.aao0185.

Fortunato, Santo (2010). "Community detection in graphs". In: *Physics Reports* 486.3-5, pp. 75–174. ISSN: 03701573. DOI: 10.1016/j.physrep.2009.11.002.

Freedman, David A. (2009). "Snow on cholera". In: *Statistical Models: Theory And Practice Paperback*. Cambrdge University Press. Chap. 1.3, pp. 6–9.

– (2010). *Statistical Models and Causal Inference - A Dialogue with the Social Sciences*. Ed. by David Collier, Jasjeet S. Sekhon, and Philip B. Stark.

Fry, Jenny (2006). "Coordination and Control of Research Practice across Scientific Fields: Implications for a Differentiated E-Science". In: *New Infrastructures for Knowledge Production: Understanding E-Science*. Ed. by Christine Hine. Information Science Publishing, pp. 167–. ISBN: 1-59140-718-4.

Galam, Serge (2012). *Sociophysics: A Physicist's Modeling of Psycho-political Phenomena*. Springer.

Galison, Peter (1997). "Three laboratories". In: *Social Research* 64.3, pp. 1127–1155.

– (2011). "Computer Simulations and the Trading Zone". In: *From Science to Computational Science*. Ed. by Gabriele Gramelsberger. Diaphanes: Diaphanes, pp. 118–157.

Garfield, E. (1962). "Can citation indexing be automated?" In: *Essays of an Information Scientist* 1, pp. 84–90.

Garnett, Emma (2016). "Developing a feeling for error: Practices of monitoring and modelling air pollution data". In: *Big Data & Society* July–Decem, pp. 1–12. DOI: 10.1177/2053951716658061.

El-Ghoroury, N. H. et al. (2012). "Stress, coping, and barriers to wellness among psychology graduate students". In: *Training and Education in Professional Psychology* 6, pp. 122–134. DOI: 10.1037/a0028768.

Gibson, James J. (1979). *The Ecological Approach To Visual Perception*. Boston: Houghton Mifflin.

Gingras, Y. and B. Gemme (2006). "L'Emprise du champ scientifique sur le champ universitaire et ses effets". In: *Actes de la Recherche en Sciences Sociales* 164, pp. 51–60.

Gitelman, Lisa and Virginia Jackson (2013). "Introduction". In: *"Raw Data" Is an Oxymoron*. Ed. by Lisa Gitelman. MIT Press, Cambridge, MA, USA, London, England., pp. 1–14.

Glaser, Jochen and Grit Laudel (2015). "A Bibliometric Reconstruction of Research Trails for Qualitative Investigations of Scientific Innovation". In: *Historical Social Research* 3, pp. 299–330. DOI: 10.12759/hsr.40.2015.3.299-330.

González-Bailón, Sandra (2013). "Commentary: Big data and the fabric of human geography". In: *Dialogues in Human Geography* 3.3, pp. 292–296. ISSN: 2043-8206. DOI: 10.1177/2043820613515379. URL: http://dhg.sagepub.com/cgi/content/abstract/3/3/292.

Goulden, Murray et al. (2017). "Wild interdisciplinarity: ethnography and computer science". In: *International Journal of Social Research Methodology* 20.2, pp. 137–150. DOI: 10.1080/13645579.2016.1152022.

Graham, Mark and Cristopher Foster (2016). "Geographies of Information Inequality in Sub-Saharan Africa". In: *The African Technopolitan* 5, pp. 78–85.

Graham, Mark and T. Shelton (2013). "Geography and the future of big data, big data and the future of geography". In: *Dialogues in Human Geography* 3.3, pp. 255–261. ISSN: 2043-8206. DOI: 10.1177/2043820613513121. URL: http://dhg.sagepub.com/lookup/doi/10.1177/2043820613513121.

Graham, Mark and Matthew Zook (2013). "Augmented realities and uneven geographies: exploring the geolinguistic contours of the web". In: *Environment and Planning A* 45, pp. 77–99.

Greene, Jennifer C., Valerie J. Carcelli, and Wendy F. Graham (1989). "Toward a Conceptual Framework for Mixed-Method Evaluation Designs Jennifer C . Greene ; Valerie J . Caracelli ; Wendy F . Graham". In: *Educationl Evaluation and Policy Analysis* 11.3, pp. 255–274.

Grimmer, Justin (2015). "We Are All Social Scientists Now: How Big Data, Machine Learning, and Causal Inference Work Together". In: *Political Science & Politics* 48.1, pp. 80–83. DOI: https://doi.org/10.1017/S1049096514001784.

Gross, P. L. K. and E. M. Gross (1927). "College libraries and chemical education". In: *Science* 66, pp. 385–389.

Haklay, Mordechai (Muki) (2012). "'Nobody wants to do council estates' – digital divide, spatial justice and outliers". In: *Presentation at the Annual Meeting of the Association of American Geographers.* URL: https://povesham.wordpress.com/2012/03/05/nobody-wants-to-do-council-estates-digital-divide-spatial-justice-and-outliers-aag-2012/.

Hanna, Alex et al. (2020). "Towards a critical race methodology in algorithmic fairness". In: *FAT* 2020 - Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, pp. 501–512. DOI: 10.1145/3351095.3372826.

Haraway, Donna J. (2016). *Staying with the Trouble: Making Kin in the Chthulucene.* Duke University Press.

Harvey, David (1969). *Explanation in Geography.* Hodder & Stoughton Educational. ISBN: ISBN-10: 0713154640.

Held, Matthias, Grit Laudel, and Jochen Glaser (2020). "Topic Reconstruction from Networks of Papers may not be possible if only one Algorithm is applied to only one Data Model". In: *Lockdown Bibliometrics: Papers not submitted to the STI Conference 2020 in Aarhus*, pp. 18–26.

Hicks, Diana et al. (2015). "The Leiden Manifesto for research metrics". In: *Nature* 520, pp. 429–431.

Hine, Christine (2006). "Computerization Movements and Scientific Disciplines: The Reflexive Potential of New Technologies". In: *New Infrastructures for Knowledge Production: Understanding E-Science.* Ed. by Christine Hine. Information Science Publishing, pp. 26–47.

Hochman, Nadev and Lev Manovich (2013). "Zooming into an Instagram City: Reading the local through social media". In: *First Monday* 18.7.

Hood, William W. and C. S. Wilson (2001). "The literature of bibliometrics, scientometrics and informetrics". In: *Scientometrics* 52.2, pp. 291–314.

Hossenfelder, Sabine (2018). *Lost in Math: How Beauty Leads Physics Astray.* Basic Books, Hachette Book Group.

Huang, C., A. Notten, and N. Rasters (2011). "Nanoscience and technology publications and patents: A review of social science studies and search strategies". In: *The Journal of Technology Transfer* 36.2, pp. 145–172.

Hutchins, Edwin (1995). *Cognition in the wild.* MIT Press.

IC2S2 (2019). *5th International Conference on Computational Social Science*. URL: `https://2019.ic2s2.org`.

Iliadis, Andrew and Federica Russo (2016). "Critical data studies: An introduction". In: *Big Data & Society* July–Decem, pp. 1–7.

Jackson, D (2009). "An international profile of industry-relevant competencies and skill gaps in modern graduates". In: *International Journal of Management Education* 8.3. DOI: `10.3794/ijme.83.288`.

Jebb, Andrew T., Scott Parrigon, and Sang Eun Woo (2017). "Exploratory data analysis as a foundation of inductive research". In: *Human Resource Management Review* 27.2, pp. 265–276. ISSN: 10534822. DOI: `10.1016/j.hrmr.2016.08.003`. URL: `http://dx.doi.org/10.1016/j.hrmr.2016.08.003`.

Jessop, Bob, Neil Brenner, and Martin Jones (2008). "Theorozing sociospatial relations". In: *Environment and Planning D Society and Space* 26, pp. 389–401. DOI: `10.1068/d9107`.

Jia, Jianfeng et al. (2016). "Towards Interactive Analytics and Visualization on One Billion Tweets". In: *SIGSPATIAL'16 October 31 - November 03, 2016, Burlingame, CA, USA*. DOI: `http://dx.doi.org/10.1145/2996913.2996923`.

Johnston, Ron (2003). "Geography: A different sort of discipline?" In: *Transactions of the Institute of British Geographers* 28.2, pp. 133–141. ISSN: 00202754. DOI: `10.1111/1475-5661.00083`.

Kelty, Christopher M. (2008). "Geeks and Recursive Publics". In: *Two Bits: The Cultural Significance of Free Software*. Duke University Press, pp. 27–63.

Kennedy, Helen, Thomas Poell, and Jose van Dijck (2015). "Data and agency". In: *Big Data & Society* 2.2, pp. 1–7. DOI: `10.1177/2053951715621569`.

Kertesz, Janos and Balazs Vedres (2018). "Europe's First PhD Program in Network Science". In: *Network Science In Education*. Springer, Cham, pp. 87–97.

Kitchin, Rob (2013). "Big data and human geography Opportunities, challenges and risks". In: *Dialogues in Human Geography* 3.3, pp. 262–267.

Klein, Julie Thompson (1990). *Interdisciplinarity: History, Theory, and Practice*. Wayne State University Press. ISBN: ISBN 0-8143-2088-0.

Knorr-Cetina, Karin (1982). "Scientific Communities or Transepistemic Arenas of Research" A Critique of Quasi-Economic Models of Science". In: *Social Studies of Science*, pp. 101–130.

– (1999). *Epistemic Cultures: How The Sciences Make Knowledge*. Harvard University Press.

Knox, Hannah and Dawn Nafus (2018). *Ethnography for a data-saturated world*. Manchester University Press.

Knuuttila, Tarja and Andrea Loettgers (2014). "Magnets, Spins, and Neurons: The Dissemination of Model Templates Across Disciplines". In: *The Monist* 97.3, pp. 80–300.

Kramer, A., J. Guillory, and J. Hancock (2014). "Experimental evidence of massive-scale emotional contagion through social networks". In: *Proceedings of the National Academy of Sciences* 111.24, pp. 8788–8790. URL: `http://www.pnas.org/content/111/24/8788.full.pdf+html`.

Kwan, Mei-Po (2012). "The Uncertain Geographic Context Problem". In: *Annals of the Association of American Geographers* 102.5, pp. 958–968.

Lam, Alice (2010). "From 'Ivory Tower Traditionalists' to 'Entrepreneurial Scientists'? Academic Scientists in Fuzzy University–Industry Boundaries". In: *Social Studies of Science* 40.2, pp. 307–340. DOI: `10.1177/0306312709349963`.

Landström, Catharina et al. (2011). "Coproducing Flood Risk Knowledge: Redistributing Expertise in Critical 'Participatory Modelling'". In: *Environment and Planning A: Economy and Space* 43.7, pp. 1617–1633.

Latour, Bruno (1987). *Science in action: How to follow scientists and engineers through society*. Cambridge, MA: Harvard University Press.

– (1988). *The Pasteurization of France*. Harvard University Press.

– (2005a). *Critical Distance or Critical Proximity? A dialogue in Honor of Donna Haraway*. Tech. rep. URL: `http://www.bruno-latour.fr/node/248.html`.

– (2005b). *Reassembling the Social: An introduction into actor-network theory*. Oxford University Press.

– (2010). "An Attempt at a "Compositionist Manifesto"". In: *New Literary History* 41, pp. 471–490.

– (2013). *An Inquiry into the Modes of Existence*. Harvard University Press, Cambridge, Mass.

Latour, Bruno, Pablo Jensen, et al. (2012). "'The whole is always smaller than its parts' - a digital test of Gabriel Tarde's monads". In: *The British Journal of Socioology* 63.4.

Latour, Bruno, Philippe Mauguin, and Geneviève Teil (1992). "A Note on Socio-Technical Graphs". In: *Social Studies of Science* 22.1, pp. 33–57.

Latour, Bruno and S Woolgar (1986). *Laboratory Life: The Construction of Scientific Facts*. 2nd editio. Princeton, NJ: Princeton University Press.

Laudel, Grit and Jochen Gläser (2007). "Interviewing Scientists". In: *Science, Technology & Innovation Studies* 3.2, pp. 91–111.

Laumann, E., P. Marsden, and D. Prensky (1983). "The boundary specification problem in network analysis". In: *Applied Network Analysis*. Ed. by Ronald Burt and Michael Minor. Beverly Hills, CA: Sage, pp. 18–34.

Law, J, E Ruppert, and M Savage (2011). *The Double Social Life of Methods*. Tech. rep. CRESC, Open University.

Lazer, David et al. (2009). "Computational Social Science". In: *Science* 323.5915, pp. 721–723. DOI: 10.1126/science.1167742.

Lenoir, Timothy (1979). "Quantitative Foundations for the Sociology of Science: On Linking Blockmodeling with Co-Citation Analysis". In: *Social Studies of Science* 9.4, pp. 455–480.

Leonelli, Sabina (2016). *Data-Centric Biology: A Philosophical Study*. University of Chicago Press.

Lépinay, Vincent Antonin (2011). *Codes of Finance Engineering Derivatives in a Global Bank*. Princeton University Press.

Leszczynski, Agnieszka and Jeremy W Crampton (2016). "Introduction: Spatial Big Data and everyday life". In: *Big Data & Society* July–Decem, pp. 1–6.

Levin, Nadine (2014). "Multivariate statistics and the enactment of metabolic complexity". In: *Social Studies of Science* 44.4, pp. 555–578. DOI: 10.1177/0306312714524845.

Leydesdorff, Loet (1989). "The relations between qualitative theory and scientometric methods in science and technology studies: Introduction to the topical issue". In: *Scientometrics* 15.5-6, pp. 333–347.

– (2008). "On the normalization and visualization of author co-citation data: Salton's Cosine versus the Jaccard index". In: *Journal of the American Society for Information Science and Technology* 59.1, pp. 77–85. DOI: https://doi.org/10.1002/asi.20732.

Leydesdorff, Loet, Ismael Ràfols, and Staša Milojević (2020). "Bridging the divide between qualitative and quantitative science studies". In: *Quantitative Science Studies* 1.3, pp. 918–926. DOI: 10.1162/qss{_}e{_}00061.

Lezaun, Javier, Noortje Marres, and Manuel Tironi (2017). "Experiments in Participation". In: *The Handbook of Science and Technology Studies*. Ed. by Ulrike Felt et al. The MIT Press, pp. 195–221.

Loukissas, Yanni Alexander (2016). "A place for Big Data: Close and distant readings of accessions data from the Arnold Arboretum". In: *Big Data & Society* July–Decem, pp. 1–20. DOI: 10.1177/2053951716661365.

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

Lury, Celia and Nina Wakeford (2012). "Introduction: a perpetual inventory". In: *Inventive Methods: The happening of the social*. Ed. by Celia Lury and Nina Wakeford. Routledge, pp. 1–24. ISBN: 978-0-415-57481-5.

Lynch, Michael (1993). *Scientific practice and ordinary action Ethnomethodology and social studies of science*. Cambrdge University Press. ISBN: 0-521-43152-2.

– (2000). "Against Reflexivity as an Academic Virtue and Source of Privileged Knowledge". In: *Theory, Culture & Society* 17.3, pp. 26–54.

MacKenzie, Adrian (2017). *Machine Learners: Archaeology of a Data Practice*. Cambridge, Massachusetts London, England: The MIT Press.

Mackenzie, Adrian and Ruth McNally (2013). "Living Multiples: How Large-scale Scientific Data-mining Pursues Identity and Differences". In: *Theory, Culture & Society* 30.4, pp. 72–91. DOI: 10.1177/0263276413476558.

MacRoberts, M. H. and B. R. MacRoberts (1986). "Quantitative measures of communication in science: A study of the formal level". In: *Social Studies of Science* 16.1, pp. 151–172.

MacRoberts, Michael H and Barbara R MacRoberts (1989). "Problems of Citation Analysis: A Critical Review". In: *Journal of the American Society for Information Science* 40.5, pp. 342–349.

Madden, Marguerite and Amy Ross (2009). "Genocide and GIScience: Integrating Personal Narratives and Geographic Information Science to Study Human Rights". In: *The Professional Geographer* 61.4, pp. 508–526. ISSN: 0033-0124. DOI: 10.1080/00330120903163480. URL: http://www.tandfonline.com/doi/abs/10.1080/00330120903163480.

Madsen, Koed Anders (2015). "Between technical features and analytic capabilities: Charting a relational affordance space for digital social analytics". In: *Big Data & Society* 2.1, pp. 1–15. DOI: 10.1177/2053951714568727.

Malik, Momin M. et al. (2015). "Population Bias in Geotagged Tweets". In: *Standards and Practices in Large-Scale Social Media Research: Papers from the 2015 ICWSM Workshop*, pp. 18–27.

Maniglier, Patrice (2019). "Problem and Structure: Bachelard, Deleuze and Transdisciplinarity". In: *Theory, Culture & Society*, pp. 1–21. DOI: 10.1177/0263276419878245.

Manovich, L. et al. (2014). "The Exceptional and the Everyday: 144 Hours in Kiev". In: *2014 IEEE International Conference on Big Data*, pp. 72–79.

Marcus, George E. (1995). "Ethnography in/of the world system: The Emergence of Multi-Sited Ethnography". In: *Annual Review of Anthropology* 24, pp. 95–117.

Marres, Noortje (2017a). *Digital Sociology: The Reinvention of Social Research.* Polity Press.

– (2017b). “Does digital sociology have problems?” In: *Digital Sociology: The Reinvention of Social Research.* Polity Press, pp. 174–194. ISBN: 978-0-7456-8479-6.

Marres, Noortje and Sarah de Rijcke (2020). “From indicators to indicating interdisciplinarity: A participatory mapping methodology for research communities in-the-making”. In: *Quantitative Science Studies* 1.3, pp. 1041–1055. DOI: 10.1162/qss{_}a{_}00062.

Marres, Noortje and Carolin Gerlitz (2016). “Interface Methods: Renegotiating Relations between Digital Social Research, STS and Sociology”. In: *The Sociological Review* 64.1, pp. 21–46. DOI: 10.1111/1467-954X.12314. URL: https://doi.org/10.1111/1467-954X.12314.

Marres, Noortje and David Moats (2015). “Mapping Controversies with Social Media: The Case for Symmetry”. In: *Social Media + Society*, pp. 1–17. DOI: 10.1177/2056305115604176.

Martin, P. Y. (2004). “Gender As Social Institution”. In: *Social Forces* 82.4, pp. 1249–1273. URL: https://doi.org/10.1353/sof.2004.0081.

Massey, Doreen (1994). *Space, Place and Gender.* Minneapolis: University of Minnesota Press.

Mayhew, Robert J (2011). “Geography’s Genealogies”. In: *The SAGE Handbook of Geographical Knowledge.* Ed. by John A Agnew and David N Livingstone. SAGE Publications Inc., pp. 21–38.

McAllister, James W. (2002). “Recent work on aesthetics of science”. In: *International Studies in the Philosophy of Science* 16.1, pp. 7–11. DOI: 10.1080/02698590120118783.

Meeteren, Michiel van et al. (2016). “Pacifying Babel’s Tower: A scientometric analysis of polycentricity in urban research”. In: *Urban Studies* 53.6, pp. 1278–1298.

Mellon, Jonathan and Christopher Prosser (2017). “Twitter and Facebook are not representative of the general population: Political attitudes and demographics of British social media users”. In: *Research & Politics* 4.3, pp. 1–9. URL: https://doi.org/10.1177/2053168017720008.

Metzler, K. et al. (2016). *Who is doing computational social science? Trends in big data research (White paper).* Tech. rep. London, UK: SAGE. DOI: 10.4135/wp160926.. URL: https://us.sagepub.com/sites/default/%0Afiles/CompSocSci.pdf.

Miller, Daniel and Jolynna Sinanan (2017). *Visualising Facebook.* UCL Press.

Millington, James D A and John Wainwright (2017). "Mixed qualitative-simulation methods: Understanding geography through thick and thin". In: *Progress in Human Geography* 41.1, pp. 68–88. DOI: 10.1177/0309132515627021.

Moats, David and Erik Borra (2018). "Quali-quantitative methods beyond networks: Studying information diffusion on Twitter with the Modulation Sequencer". In: *Big Data & Society* January–Ju, pp. 1–17. DOI: 10.1177/2053951718772137.

Mol, Annemarie (2002). *The Body Multiple: Ontology in Medical Practice*. Durham: Duke University Press.

Mongeon, Philippe and Adèle Paul-Hus (2016). "The journal coverage of Web of Science and Scopus: a comparative analysis". In: *Scientometrics* 106, pp. 213–228.

Morstatter, Fred et al. (2013). "Is the Sample Good Enough? Comparing Data from Twitter's Streaming API with Twitter's Firehose". In: *Proceedings of the Seventh International AAAI Conference on Weblogs and Social Media*, pp. 400–408.

Muller, Michael et al. (2019). "How Data Science Workers Work with Data: Discovery, Capture, Curation, Design, Creation". In: *Systems Proceedings (CHI'19)*. Glasgow, Scotland UK. DOI: https://doi.org/10.1145/3290605.3300356.

Munk, Anders Kristian (2019). "Four Styles of Quali-Quantitative Analysis Making sense of the new Nordic food movement on the web". In: *Nordicom Review* 40.1, pp. 159–176. DOI: 10.2478/nor-2019-0020.

Munk, Anders Kristian and Torben Elgaard Jensen (2014). "Revisiting the Histories of Mapping: Is there a Future for a Cartographic Ethnology?" In: *Ethnologia Europaea* 44.2, pp. 31–47.

Munoz-Ecija, Teresa et al. (2013). "Nanoscience and nanotechnology in scopus: Journal identification and visualization". In: *Proceedings of ISSI 2013 - 14th International Society of Scientometrics and Informetrics Conference* 2, pp. 2061–2063.

Navon, Daniel and Uri Shwed (2012). "The chromosome 22q11.2 deletion: From the unification of biomedical fields to a new kind of genetic condition". In: *Social Science & Medicine* 75, pp. 1633–1641. URL: http://dx.doi.org/10.1016/j.socscimed.2012.06.023.

Neff, G et al. (2017). "Critique and Contribute: A Practice-Based Framework for Improving Critical Data Studies and Data Science". In: *Big Data* 5.2, pp. 85–97. DOI: doi:10.1089/big.2016.0050..

Nelson, Nicole C. (2018). *Model Behavior. Animal experiments, complexity and the gentics of psychiatric disorders*. The University of Chicago Press.

Nepusz, T. and G. Csardi (n.d.). *Community structure via greedy optimization of modularity.*

Newman, M E J (2001). "The structure of scientific collaboration networks". In: *Proceedings of the National Academy of Sciences of the Unitedeoleol States of America* 98.2, pp. 404–409. ISSN: 0027-8424. DOI: 10.1073/pnas.021544898.

Nikitopoulos, Panagiotis et al. (2016). "BigCAB: Distributed Hot Spot Analysis over Big Spatio-temporal Data using Apache Spark (GIS Cup)". In: *SIGSPATIAL'16, October 31-November 03, 2016, Burlingame, CA, USA*. DOI: http://dx.doi.org/10.1145/2996913.3004065.

O'Neil, Cathy (2016). *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy.* Crown Publishing Group.

Osborne, T (2013). "Inter that Discipline!" In: *Interdisciplinarity: Reconfigurations of the Social and Natural Sciences.* Ed. by Andrew Barry and Georgina Born. Routledge, pp. 82–98.

Pain, Rachel (2004). "Social geography: participatory research". In: *Progress in Human Geography* 28.5, pp. 652–663.

Parker-Jenkins, Marie (2018). "Problematising ethnography and case study: reflections on using ethnographic techniques and researcher positioning". In: *Ethnography and Education* 13.1, pp. 18–33. DOI: 10.1080/17457823.2016.1253028.

Pentland, Alex (2015). *Social Physics.* Penguin Random House.

Pléh, Csaba (2010). *A lélektan története.* Osiris Kiadó.

Pound, Eleri (2018). *Centre for Data Analytics and Society welcomes its second cohort of students.*

Reevy, G.M. and G. Deason (2014). "Predictors of depression, stress, and anxiety among non-tenure track faculty". In: *Frontiers in Psychology* 5. URL: http://dx.doi.org/10.3389/fpsyg.2014.%2000701..

Rheinberger, Hans-Jorg (1997). *Toward a History of Epistemic Things. Synthesizing Proteins in the Test Tube.* Stanford, California: Stanford University Press.

Rieder, Bernhard (2017). "Scrutinizing an algorithmic technique: the Bayes classifier as interested reading of reality". In: *Information, Communication & Society* 20.1, pp. 100–117. DOI: 10.1080/1369118X.2016.1181195.

Robinson-García, Nicolás et al. (2014). "New data, new possibilities: Exploring the insides of Altmetric.com". In: *El profesional de la información* 23.4, pp. 359–366.

Rotolo, Daniele, Diana Hicks, and Ben Martin (2015). "What Is an Emerging Technology?"

Ruppert, E., J. Law, and M. Savage (2013). "Reassembling social science methods:The challenge of digital devices." In: *Theory Culture and Society* 30.4, pp. 22–46.

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

Salton, G and C. Buckley (1988). "Term-weighting approaches in automatic text retrieval". In: *Information Processing & Management* 24, pp. 513–523.

Savage, Mike and Roger Burrows (2007). "The Coming Crisis of Empirical Sociology". In: *Sociology* 41.5, pp. 885–899.

Schönbauer, Sarah Maria (2020). "'From Bench to Stage': How Life Scientists' Leisure Groups Build Collective Self-Care". In: *Science as Culture* 29.4, pp. 524–545. DOI: 10.1080/09505431.2020.1713077.

Schreiber, Maria (2017). "Audiences, Aesthetics and Affordances: Analysing Practices of Visual Communication on Social Media". In: *Digital Culture & Society* 3.2, pp. 143–163. DOI: https://doi.org/10.25969/mediarep/13519.

Shelton, Taylor (2016). "Spatialities of data: Mapping social media 'beyond the geotag'". In: *GeoJournal*.

Shelton, Taylor et al. (2014). "Mapping the data shadows of Hurricane Sandy: Uncovering the sociospatial dimensions of 'big data'". In: *Geoforum* 52, pp. 167–179.

Sheppard, Eric (2005). "Knowledge Production through Critical GIS: Genealogy and Prospects". In: *Cartographica: The International Journal for Geographic Information and Geovisualization* 40.4, pp. 5–21. DOI: 10.3138/GH27-1847-QP71-7TP7.

Shi, Feng, Jacob G. Foster, and James A. Evans (2015). "Weaving the fabric of science: Dynamic network models of science's unfolding structure". In: *Social Networks* 43, pp. 73–85. DOI: https://doi.org/10.1016/j.socnet.2015.02.006.

Simini, Filippo et al. (2012). "A universal model for mobility and migration patterns." In: *Nature*, pp. 8–12. ISSN: 1476-4687. DOI: 10.1038/nature10856. URL: http://www.ncbi.nlm.nih.gov/pubmed/22367540.

Sinatra, Roberta et al. (2016). "Quantifying the evolution of individual scientific impact". In: *Science* 354.6312. DOI: 10.1126/science.aaf5239.

Sismondo, Sergio (2010). "The Prehistory of Science and Technology Studies". In: *An Introduction to Science and Technology Studies*. Second Edi. Wiley-Blackwell, pp. 1–11.

Sloan, Luke and Jeffrey Morgan (2015). "Who Tweets with Their Location? Understanding the Relationship between Demographic Characteristics and the Use of Geoservices and Geotagging on Twitter". In: *PLOS One* 10.11. DOI: 10.1371/journal.pone.0142209. URL: https://doi.org/10.1371/journal.pone.0142209.

Small, H. G. (1977). "A co-citation model of a scientific specialty: A longitudinal study of collagen research". In: *Social Studies of Science* 7, pp. 139–166.

– (1978). "Cited documents as concept symbols". In: *Social Studies of Science* 8.3, pp. 327–340.

Smith-Doerr, Laurel (2017). "Doing, Exploring, and Reflecting on Methods". In: *The Handbook of Science and Technology Studies*. Ed. by Ulrike Felt et al. The MIT Press, pp. 27–30.

Starke, Robert E. (1978). "The Case Study Method in Social Inquiry". In: *Educational Researcher*.

Strathern, M. (2004). "Introduction: In Crisis Mode - a Comment on Interculturality". In: *Commons and Borderlands: Working Papers on Interdisciplinarity, Accountability and the Flow of Knowledge*. Sean Kingston Publishing, pp. 1–14. ISBN: 0-9545572-2-0.

Sui, D. and D. DeLyser (2012). "Crossing the qualitative-quantitative chasm I: Hybrid geographies, the spatial turn, and volunteered geographic information (VGI)". In: *Progress in Human Geography* 36.1, pp. 111–124. ISSN: 0309-1325. DOI: 10.1177/0309132510392164.

Sui, Daniel and Michael F. Goodchild (2011). "The convergence of GIS and social media: Challenges for GIScience". In: *International Journal of Geographical Information Science* 25.11, pp. 1737–1748. DOI: 10.1080/13658816.2011.604636.

Sunderland, N. et al. (2012). "What does it feel like to live here? Exploring sensory ethnography as a collaborative methodology for investigating social determinants of health in place". In: *Health & Place* 17, pp. 1056–1067.

Taylor, Alex S, Sian Lindley, et al. (2014). "Data and life on the street". In: *Big Data & Society* 1.2, pp. 1–7. DOI: 10.1177/2053951714539278.

Taylor, Linnet and C. Richter (2015). "Big Data & Urban Governance". In: *Geographies of urban governance*. Springer, pp. 175–191.

Teachout, D. J. (2004). "Incentives and barriers for potential music teacher education doctoral students". In: *Journal of Research in Music Education* 52.3, pp. 234–247.

Teil, Geneviève and Bruno Latour (1995). "The Hume Machine: Can Association Networks do more than Formal Rules". In: *Stanford Humanities Review* 4.2, pp. 47–65.

Timans, Rob, Paul Wouters, and Johan Heilbron (2019). "Mixed methods research: what it is and what it could be". In: *Theory and Society* 48.2, pp. 193–216. ISSN: 15737853. DOI: 10.1007/s11186-019-09345-5.

Traag, V. A., L. Waltman, and N. J. van Eck (2019). "From Louvain to Leiden: guaranteeing well-connected communities". In: *Scientific Reports* 9.5233, pp. 1–12. DOI: https://doi.org/10.1038/s41598-019-41695-z.

Tuan, Yi-Fu (2001). *Space and Place: The Perspective of Experience.* University of Minnesota Press.

Tytherleigh, M.Y. et al. (2005). "Occupational stress inUK higher education institutions: a comparative study of all staff categories." In: *Higher Education Research & Development* 24.1, pp. 41–61.

University of Groningen (2019).

University of Groningen (2019). URL: https://www.rug.nl/gmw/education/minor/datawise.

Van Eck, Nees Jan and Ludo Waltman (2010). "Software survey: VOSviewer, a computer program for bibliometric mapping". In: *Scientometrics* 84, pp. 523–538. DOI: 10.1007/s11192-009-0146-3.

– (n.d.). *Text mining and visualization using VOSviewer.* Tech. rep. URL: https://arxiv.org/pdf/1109.2058.pdf.

Van Eck, Nees Jan and Ludo Waltmann (2020). *VOSviewer Manual: Manual for VOSviewer version 1.6.15.* Tech. rep. CWTS, Leiden University, pp. 1–53.

Velden, Theresa and Carl Lagoze (2013). "The Extraction of Community Structures from Publication Networks to Support Ethnographic Observations of Field Differences in Scientific Communication". In: *Journal of the American Society for iIformation Science and Technology* 64.12, pp. 2405–2427. DOI: 10.1002/asi.22929.

Venturini, Tommaso, Liliana Bounegru, et al. (2018). "A reality check(list) for digital methods". In: *New Media & Society* 20.11, pp. 1–23. DOI: 1.o0r.g1/107.171/1774/6114461444841881787669923 6.

Venturini, Tommaso, Anders Kristian Munk, and Mathieu Jacomy (2019). "Actor-Network versus Network Analysis versus Digital Networks Are We Talking about the Same Networks?" In: *digitalSTS: A Field Guide For Science & Technology Studies.* Ed. by J. Vertesi et al. Princeton University Press, pp. 510–523.

Vertesi, J. et al., eds. (2019a). *digitalSTS: A Field Guide for Science & Technology Studies.* Princeton University Press.

"Introduction" (2019b). In: *digitalSTS: A Field Guide For Science & Technology Studies.* Ed. by J. Vertesi et al. Princeton University Press, pp. 1–10.

"Preface: The digitalSTS Community" (2019c). In: *digitalSTS: A Field Guide For Science & Technology Studies.* Ed. by J. Vertesi et al. Princeton University Press, pp. ix–xi.

Wajid, Faizan and Hanan Samet (2016). “CrimeStand: Spatial Tracking of Criminal Activity”. In: *SIGSPATIAL’16 October 31 - November 03, 2016, Burlingame, CA, USA*. DOI: `http://dx.doi.org/10.1145/2996913.2997006`.

Waltman, Ludi and Nees Jan van Eck (2012). “A New Methodology for Constructing a Publication-Level Classification System of Science”. In: *JOURNAL OF THE AMERICAN SOCIETY FOR INFORMATION SCIENCE AND TECHNOLOGY* 63.12, pp. 2378–2392. DOI: `10.1002/asi.22748`.

Waltman, Ludo, Kevin W. Boyack, et al. (2020). “A principled methodology for comparing relatedness measures for clustering publications”. In: *Quantitative Science Studies* 1.2, pp. 691–713. DOI: `https://doi.org/10.1162/qss{\_}a{\_}00035`.

Waltman, Ludo, Vincent Larivière, et al. (2020). “Opening science: The rebirth of a scholarly journal”. In: *Quantitative Science Studies* 1.1, pp. 1–3. URL: `https://doi.org/10.1162/qss_e_00025`.

Wasserman, Stanley and Katherine Faust (1994). “Social Network Analysis in the Social and Behavioural Sciences”. In: *Social Network Analysis*. Cambridge Universtiy Press, pp. 3–27. ISBN: 978-0-521-38707-1.

Weisz, George, Alberto Cambrosio, and Jean-Philippe Cointet (2017). “Mapping Global Health: A network analysis of a heterogeneous publication domain”. In: *BioSocieties* 12.4, pp. 520–542.

Wen, Bei et al. (2017). “Mapping Science Through Bibliometric Triangulation: An Experimental Approach Applied to Water Research”. In: *JOURNAL OF THE ASSOCIATION FOR INFORMATION SCIENCE AND TECHNOLOGY* 68.3, pp. 724–738. DOI: `10.1002/asi.23696`.

Weszkalnys, Gisa and Andrew Barry (2013). “Multiple Environments: Accountability, integration and ontology”. In: *Interdisciplinarity: Reconfigurations of the social and natural sciences*. Ed. by Andrew Barry and Georgina Born. Routledge, pp. 178–208.

White, Howard D. (2011). “Scientific and Scholarly Networks”. In: *The SAGE Handbook of Social Network Analysis*. Ed. by John Scott and Peter J. Carrington. SAGE Publications Inc., pp. 271–285.

Winthereik, Brit Ross (2019). “Is ANT’s radical empiricism ethnographic?” In: *The Routledge Companion to Actor-Network Theory*, pp. 24–33. DOI: `10.4324/9781315111667-4`.

Wouters, P. (2014). “The Citation: From Culture to Infrastructure”. In: *Beyond Bibliometrics: Harnessing Multidimensional Indicators of Scholarly Impact*. Ed. by Blaise Cronin and Cassidy R Sugimoto. The MIT Press, pp. 47–66.

Wouters, Paul and Anne Beaulieu (2006). "Imagining E-Science beyond Computation". In: *New Infrastructures for Knowledge Production: Understanding E-Science*. Ed. by Christine Hine. Information Science Publishing, pp. 48–70.

Wouters, Paul, Anne Beaulieu, et al. (2012). *Virtual Knowledge: Experimenting in the Humanities and the Social Sciences*. Ed. by Paul Wouters et al. The MIT Press. ISBN: 9780262517911.

Wyatt, Sally, Staša Milojević, et al. (2017). "Intellectual and Practical Contributions of Scientometrics to STS". In: *The Handbook of Science and Technology Studies*. Ed. by Ulrike Felt et al. Fourth Edi. The MIT Press Cambridge, Massachusetts London, England, pp. 87–112.

Wyatt, Sally, Andrea Scharnhorst, et al. (2013). "Introduction to Virtual Knowledge". In: *Virtual Knowledge: Experimenting in the Humanities and the Social Sciences*. Ed. by Paul Wouters et al. The MIT Press, Cambridge, Massachusetts; London, England, pp. 1–23.

Wylie, Lindsey E. et al. (2018). "Four decades of the journal Law and Human Behavior: a content analysis". In: *Scientometrics* 115.2, pp. 655–693. DOI: `https://doi.org/10.1007/s11192-018-2685-y`.

Zitt, Michel (2015). "Meso-level retrieval: IR-bibliometrics interplay and hybrid citation-words methods in scientific fields delineation". In: *Scientometrics* 102.3, pp. 2223–2245. ISSN: 15882861. DOI: `10.1007/s11192-014-1482-5`.

Appendix A

Interview Questions

I used the prompts below for the semi structured interviews I conducted.

1. Motivations, Introduction
   (a) Why did you start to use geotagged social media data for your research? How did you come to this field?
   (b) What does geotagged social media research allow you to do that you could not do otherwise?
2. Interests and Challenges
   (a) What is the most exciting and interesting aspect of geotagged social media research?
   (b) What is the most surprising aspect of geotagged social media research?
   (c) Why do you work with the specific type of social media data?
   (d) What is the biggest challenge of social media research for you? Can you recount specific challenges? Were these resolvable and if so how did you resolve them?
   (e) What would you regard as the most 'successful' aspect of your geotagged social media research and why do you consider it successful?
3. Epistemic / Uncertainty
   (a) Why did you choose the methods and theoretical constructs you chose and theoretical constructs? How useful have they been? Can you recall alternatives you considered? Is there anything that you would do differently?

   (b) How did your interaction with or opinion of geotagged social media data research change over time?
   (c) Data cleaning: Did you omit any data from the analyses? How do you think it affects your analysis?

4. Data Access and Sharing
   (a) How (and with whom) did you access the geotagged social media data?
   (b) Are you sharing your data, or data analysis tools with anyone else, and is anyone sharing data with you? If so, why? Are you planning to work with people you share data with?

5. Collaboration
   (a) Have you started to collaborate with someone as part of doing geotagged social media data analysis?
   (b) How do you collaborate as part of geotagged social media data analysis? What skills do you and co-authors contribute?
   (c) Were there any disagreements between the authors that you can remember? What were these about? How did you resolve them?
   (d) Collaborators outside of academia: Do you, or do you wish to work with policy makers, or other actors outside of academia as part of geotagged social media analysis? Do you perceive or anticipate any challenges or tensions?
   (e) Which other research groups' work do you like?

6. Knowledge of Spaces
   (a) Do you know the places you study using geotagged social media data? Have you visited them?
   (b) How does familiarity with places (or lack of thereof) affect geotagged social media data practices in your opinion?
   (c) If relevant based on paper: How did you collaborate with local actors or stakeholders? Could or did the use of geotagged social media data help you to collaborate with local stakeholders?
   (d) How did social media research shape your engagement with the spaces you studied?

7. 'Imaginaries': If you could...

(a) have a new collaborator, a person with what skills would you like to work with?

(b) what new or additional data would like to use?

(c) what new data analysis tool would you like to try?

(d) get funding to do anything, what aspect of cities / landscapes / places would you study

8. Funding

   (a) How does your use of geotagged social media data relate to funding arrangements? Who funds your work? What does the funding body have a preference for?

9. Ethics and RRI

   (a) Have you considered the different ways in which your use of geotagged social media data might raise ethical or moral issues?

10. Future of geotagged social media research

   (a) What do you think is the future of geotagged social media data analysis?

Appendix B

Python and R code

The Python and R scripts used in this thesis are available at the following Zenodo Repository:

`http://doi.org/10.5281/zenodo.5138104`

and the following Github repository:

Description

`https://github.com/juditvarga/juditvarga-PhD_scripts/tree/main`

Scripts

`https://github.com/juditvarga/juditvarga-PhD_scripts/tree/master`