



UNICAMP

UNIVERSIDADE ESTADUAL DE CAMPINAS
Instituto de Estudos da Linguagem (IEL)
Laboratório de Estudos Avançados em Jornalismo (LABJOR)

GUILHERME CAVALCANTE SILVA

BIG DATA E AS CIÊNCIAS SOCIAIS: UM ESTUDO QUALI-
QUANTITATIVO A PARTIR DA BIG DATA & SOCIETY

CAMPINAS,
2020

GUILHERME CAVALCANTE SILVA

Big Data e as Ciências Sociais: Um estudo quali-quantitativo a partir da
Big Data & Society

Dissertação apresentada ao Laboratório de Estudos Avançados em Jornalismo (Labjor), do Instituto de Estudos em Linguagem (IEL), da Universidade Estadual de Campinas, como parte dos requisitos exigidos para obtenção do título de Mestre em Divulgação Científica e Cultural, na área de Divulgação Científica e Cultural.

Orientador: Diego Jair Vicentin

Este trabalho corresponde à versão final da dissertação defendida pelo aluno Guilherme Cavalcante Silva e orientado pelo prof. Dr. Diego Jair Vicentin.

CAMPINAS,
2020

Agência de fomento e nº de processo: CAPES, 88882.434862/2019-01

Ficha catalográfica
Universidade Estadual de Campinas
Biblioteca do Instituto de Estudos da Linguagem
Leandro dos Santos Nascimento – CRB 8/8343

Si38b Silva, Guilherme Cavalcante, 1994-
Big Data e as Ciências Sociais : um estudo quali-quantitativo a partir da Big Data & Society / Guilherme Cavalcante Silva. – Campinas, SP : [s.n.], 2020.

Orientador: Diego Jair Vicentin.

Dissertação (mestrado) – Universidade Estadual de Campinas, Instituto de Estudos da Linguagem.

1. Big data. 2. Ciências sociais. 3. Big Data & Society. 4. Sul global.
I. Vicentin, Diego Jair. II. Universidade Estadual de Campinas. Instituto de Estudos da Linguagem. III. Título.

Informações para biblioteca digital

Título em outro idioma: Big Data and the Social Sciences: a mixed methods study from Big Data & Society

Palavras-chave em inglês:

Big data

Social sciences

Big Data & Society

Global south

Área de concentração: Divulgação Científica e Cultural

Titulação: Mestre em Divulgação Científica e Cultural

Banca examinadora:

Diego Jair Vicentin [Orientador]

Rafael de Almeida Evangelista

Sheila Khan

Data de defesa: 06-04-2020

Programa de Pós-Graduação: Divulgação Científica e Cultural

Identificação e informações acadêmicas do(a) aluno(a)

- ORCID do autor: <https://orcid.org/0000-0001-5104-778X>

- Currículo Lattes do autor: <http://lattes.cnpq.br/6126304434188396>



UNICAMP

BANCA EXAMINADORA

**Prof. Dr. Diego Jair Vicentin – Presidente
Universidade Estadual de Campinas**

**Prof. Dr. Rafael de Almeida Evangelista
Universidade Estadual de Campinas**

**Profa. Dra. Sheila Pereira Khan
Universidade do Minho (Braga/Portugal)**

**IEL/UNICAMP,
2020**

Ata da defesa, assinada pelos membros da Comissão Examinadora, consta no SIGA/Sistema de Fluxo de Dissertação/Tese e na Secretaria de Pós-Graduação do IEL.

Aos que lutam em prol da universidade pública no Brasil em tempos tão sombrios;

Aos que ousam responder o clamor por novo(s) mundo(s).

Agradecimentos

Primeiramente, agradeço ao orientador desta dissertação, Diego Vicentin, por sua prestatividade, paciência e amizade durante estes dois anos de Mestrado. Em meio às várias mudanças de percurso que esta pesquisa sofreu, seus conselhos foram fundamentais para que eu não entrasse em pânico. Não poderia ter pedido um melhor orientador!

O presente trabalho foi realizado com o apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES - Código de Financiamento 001), por meio de uma bolsa de estudos durante o período integral do Mestrado (24 meses).

Aos membros dos grupos de pesquisa ICTS, GrEGS e Geccom, por fornecerem algumas das discussões mais interessantes que tive a oportunidade de testemunhar e que contribuíram em muito para o rumo que este trabalho tomou.

Ao Sci-Hub e outras plataformas de ativismo pirata. Por um mundo sem barreiras à ciência. À Marta Kanashiro, Rafael Evangelista, Pedro Ferreira, Marko Sinésio, Tristan Torriani e Monica Zoppi-Fontana, docentes que me levaram a enxergar questões urgentes sob diferentes prismas. À Sheila Khan, que topou de última hora o convite para avaliar esta dissertação e se mostrou tão amigável e solícita nos poucos dias de convivência que tivemos em Bremen.

À Alessandra Carnauskas e Andressa Alday, pelo carinho com que tratam os discentes do Labjor e pela competência com que trabalham tão avidamente pelo sucesso do programa.

Ao meu grande mestre Tales Tomaz. Jamais sequer pensaria em cursar um Mestrado não fosse por sua orientação e conselhos. Todo o rumo que minha vida tomou, e que me levou até a Unicamp, se deve muito pelo seu apoio.

Às minhas colegas de carona: Thamires, Patricia, Quézia e Bianca. Os trajetos de ida e volta para a Unicamp, realizados nestes dois anos de Mestrado, foram mais alegres com vocês.

À Max, Bobsin, Vitor, Erick, Somaira, Carol, Érica, Cassiano, Kimberly, Hillary, Bruno, Giselly, Ícaro, Charles, Alessandra, Leonardo, Janaina, Luma, pela amizade e suporte. Todos vocês, com atitudes pequenas ou grandes, me fortaleceram nestes últimos dois anos e me encheram de felicidade. A vida com vocês tem mais sabor.

Aos meus pais: Javã e Nilvane, e meus irmãos: Daniel e Nicolas. Jamais esquecerei o amor, o carinho, os conselhos, as brincadeiras, as broncas e as orações que me trouxeram até aqui. À vocês todo o meu coração. E, principalmente, ao Eterno pela vida, força e habilidades que me concedeu. À Ele seja a honra, glória e louvor.

RESUMO:

Esta dissertação consiste em uma exposição de algumas das principais questões levantadas em uma pesquisa de caráter quali-quantitativo realizada na revista *Big Data & Society* (BD&S). A publicação tornou-se um importante locus de análises do Big Data a partir de perspectivas das Ciências Sociais nos últimos anos, sendo pioneira no fomento de tais esforços. O período da publicação analisado envolve os anos de 2014, ano de início da publicação, 2015 e 2016, cobrindo um total de seis edições e 131 artigos publicados. Tal análise teve por objetivo entender a formação e fundamentos dos discursos sobre o Big Data nas Ciências Sociais, à luz da identificação de uma necessidade hodierna de uma maior compreensão de como as perspectivas em torno do Big Data têm se cristalizado na atualidade (AKOKA et al, 2017; YOUTIE et al, 2016). Por um lado, o levantamento, através de um mapeamento analítico de caráter bibliométrico, identificou a distribuição geopolítica dos atores que informam o debate, bem como as principais referências, palavras-chaves, metodologias utilizadas e formação de ‘cânones’. De outro, buscou identificar como esses autores e comunidades científicas estruturam ali suas prioridades, alimentando a própria organização de tais grupos, através de um relato etnográfico de um dos eventos promovidos pela BD&S. O principal argumento da pesquisa é de que existe uma crescente tensão nos artigos e comunidades científicas ligadas a BD&S existente entre a procura por perspectivas críticas dos dados, alternativas às pesquisas de Big Data relacionadas ao mercado, e a manutenção de categorias e linguagem do próprio modelo que buscam repensar. Um choque entre o esforço por pensar novas formas de habitar o digital para além dos modelos sociais, econômicos e políticos do Ocidente e a realidade de não saber se referir ao esgotamento sem esgotar-se ainda mais no funcionamento de mundo típico do Ocidente ou Norte Global. Primeiramente, o trabalho apresenta um panorama geral sobre a origem e desenvolvimentos históricos do termo Big Data, além de seus desdobramentos nas Ciências Sociais. Segue a isso uma análise das comunidades científicas onde a BD&S circula e como elas se organizam, evocando o conceito de colégios invisíveis (ZUCCALA, 2014; PAISLEY, 1972), apresentando também os dados do mapeamento analítico a fim de apresentar um panorama da pesquisa na BD&S. Ainda nesta etapa, o texto apresenta um relato de experiência da participação do autor no *Data Power Conference 2019*, organizado em Bremen, na Alemanha, e promovido pela BD&S, a fim de se obter um olhar único sobre as expectativas, interações sociais e projetos que cercam membros cuja atuação está diretamente relacionada ao funcionamento da BD&S. No capítulo seguinte, apresenta uma análise dos principais textos ‘canônicos’ no contexto da BD&S visando a compreensão de como estes textos impactam os caminhos percorridos nos artigos da publicação. Por fim, a dissertação se encerra ampliando a discussão sobre a tensão entre busca por alternativas fora do Ocidente e retenção de práticas eurocêntricas.

Palavras-chave: Big Data; Ciências Sociais; Big Data & Society; Sul Global.

ABSTRACT:

This thesis explores some of the findings obtained through qualitative and quantitative researches carried out on the journal *Big Data & Society* (BD&S), an important locus of Big Data research from Social Sciences perspectives in recent years, being a pioneer in fostering such efforts. The analysis covers the period from 2014, year of the publication of the first issue of BD&S, to 2016, and a total of six issues and 131 published papers. Its main goal is to understand the formation and foundations of Big Data discourses in Social Sciences considering the growing need for greater understanding of the ways Big Data concepts are established (AKOKA et al, 2017; YOUTIE et al, 2016). On the one hand, the survey identifies, through a bibliometric analytical mapping, the geopolitical distribution of the actors that inform the debates, as well as the main references, key themes, methodologies and ‘canons’ contained in the papers. On the other hand, it sought to identify how these authors and scientific communities organize their priorities– and are organized through - in that space, through an ethnographic account of one of the events promoted by BD&S. The main argument of the research is that the papers and scientific communities linked to BD&S display an increasing tension between the search for critical data perspectives, along with alternatives to the predominant market-related Big Data researches, and the maintenance of categories and language of the same models they seek to rethink. A clash between trying to think in new ways of inhabiting the digital beyond social, economic and political models of the West and the reality of not knowing how to refer to the exhaustion of these models without language and frame of thought of the typical world functioning of the West or Global North. The first chapter presents an overview of the origin and historical developments of Big Data, in addition to its advances in Social Sciences. This is followed by an analysis of the constitution and organization of the scientific communities where BD&S circulates, evoking the concept of “invisible college” (ZUCCALA, 2014; PAISLEY, 1972), while also presenting data from the analytical mapping to offer an overview of the research in BD&S. Still at this stage, the text presents an ethnographic account of the author's participation in the Data Power Conference 2019, organized in Bremen, Germany, and promoted by BD&S, in order to obtain a unique look at the expectations, social interactions, and perspectives that surround BD&S environment. In the following chapter, the work proposes an analysis of the main ‘canonical’ texts in BD&S to understand how these texts impact the paths taken in the papers published in BD&S. Finally, we arrive at a discussion on the tension between the search for alternatives outside the West and the retention of Eurocentric practices.

Keywords: Big Data; Social Sciences; Big Data & Society; Global South.

LISTA DE FIGURAS:

Figura 1 - Mapa geopolítico da balança global do Big Data.....	30
Figura 2 - Nuvem de palavras com o sobrenome dos autores mais citados nos artigos da revista BD&S.....	44
Figura 3 - Ligações entre o uso de autores dentro de um mesmo artigo.....	46
Figura 4 - Tendência de aparição dos autores mais referenciados no decorrer do período 2014-2016.....	48
Figura 5 - Nuvem de palavras com os termos mais utilizados nas palavras-chave dos artigos publicados entre 2014 e 2016 na BD&S.....	51
Figura 6 - Origem acadêmica dos trabalhos publicados na BD&S entre 2014 e 2016.....	54
Figura 7 - Palestra inicial da Data Power Conference 2019.....	63
Figura 8 - Interação social que ocorreu na noite do primeiro dia de evento.....	69
Figura 9 - Região do Überseestadt onde foi realizado o tour sobre o passado colonial de Bremen.....	70
Figura 10 - Número de artigos sobre “Big Data”, nos campos de “Comunicação”, “Artes” e “Humanidades e Ciências Sociais”, listados na base de dados da Web of Science no período entre 2000 e 2019.....	78
Figura 11 - Número de referências ao artigo “The Coming Crisis of Empirical Sociology”..	97
Figura 12 - Mapa que critica o exagero dado em mapas ao tamanho da Europa (I).....	114
Figura 13 - Mapa que critica o exagero dado em mapas ao tamanho da Europa (II).....	114
Figura 14 - Distribuição geográfica de membros do Comitê Editorial por filiação institucional.....	116
Figura 15 - Distribuição de artigos por escolha do objeto de estudo ou contexto de análise em relação a contextos geopolíticos.....	119

LISTA DE TABELAS:

Tabela 1 - Obras mais citadas nos artigos da BD&S entre 2014 e 2016.....	47
Tabela 2 - Anos de publicação dos materiais referenciados nos artigos da BD&S.....	49
Tabela 3 - Classificação organizada com base nos modos em que os artigos da análise se apropriam da obra de boyd e Crawford.....	80
Tabela 4 - Classificação organizada com base nos modos em que os artigos da análise se apropriam da obra de Kitchin.....	86
Tabela 5 - Classificação organizada com base nos modos em que os artigos da análise se apropriam da obra de Mayer-Schönberger e Cukier.....	92
Tabela 6 - Classificação organizada com base nos modos em que os artigos da análise se apropriam da obra de Savage e Burrows.....	100
Tabela 7 - Distribuição geográfica dos autores dos artigos das seis primeiras edições da BD&S por filiação institucional.....	116
Tabela 8 – Distribuição geográfica das submissões feitas à BD&S desde o início da publicação até julho de 2019.....	120

LISTA DE ABREVIATURAS E SIGLAS:

AAG - *Association of American Geographers*

BD&S – Big Data & Society

DOAJ - Directory of Open Access Journals

ERC – European Research Council

EUA – Estados Unidos da América

Geoss - Global Earth Observation System of Systems

ICTS – Informação, Comunicação, Tecnologia e Sociedade

IDS - *Indigenous Data Sovereignty*

IEEE – Institute of Electrical and Electronics Engineers

Labjor – Laboratório Avançado de Estudos em Jornalismo

Lavits – Rede Latino-Americana de Estudos sobre Vigilância, Tecnologia e Sociedade

MIT – Massachusetts Institute of Technology

Nasa – National Aeronautics and Space Administration

NIH - National Institutes of Health

SMSociety - *International Conference on Social Media & Society*

SPI – Society for the Philosophy of Information

SSCI - Social Sciences Citation Index

TICs – Tecnologias de Informação e Comunicação

WoS – Web of Science

ZeMKI - Zentrum für Medien-, Kommunikations- und Informationsforschung

Sumário

INTRODUÇÃO	13
1. Situando o terreno: percursos da pesquisa e questões metodológicas.....	19
1.1. O incômodo.....	19
1.2 O assunto.....	22
1.3 O campo.....	26
1.4 Metodologia.....	31
2. Quem está olhando? Mapeamento analítico e as comunidades que compõem a Big Data & Society.....	35
2.1 Apresentação.....	35
2.2 BD&S: informações gerais.....	36
2.3 O meio social da revista científica.....	39
2.4 BD&S: colégios invisíveis e cânones sob um olhar quantitativo	43
2.5 BD&S e sua comunidade científica: um relato de experiência.....	56
2.6 Impressões e discussão	71
3. Olhares para onde? Uma análise a partir dos cânones da BD&S.....	74
3.1 Apresentação.....	74
3.2 boyd & Crawford.....	76
3.3 Rob Kitchin	83
3.4. Viktor Mayer-Schönberger e Kenneth Cukier	89
3.5. Mike Savage e Roger Burrows	96
3.6 Discussão.....	104
4. Esgotamento e tensão: o olhar em prol de alternativas ao Sul na BD&S.....	105
4.1 Apresentação.....	105
4.2 O olhar do ‘conquistador’ e as velhas categorias	107
4.3 BD&S e o olhar para o Sul.....	115
4.4 Discussão.....	122
5. CONSIDERAÇÕES FINAIS.....	125
REFERÊNCIAS.....	127
ANEXO - Tabela com termos utilizados nas palavras-chave da BD&S entre 2014 e 2016.....	143

INTRODUÇÃO

Pode ser que, quando já não sabemos o que fazer, tenhamos começado o nosso verdadeiro trabalho e, quando já não sabemos para onde ir, chegamos à nossa verdadeira jornada. A mente que não está perplexa não é empregada. A corrente interrompida é aquela que canta (BERRY, 2005, p. 97).

Esta dissertação é fruto de uma inquietação acerca de uma certa narrativa que aparecia em muitas das leituras que fiz sobre as ciências dos dados e o Big Data¹. Tais leituras que figuravam noticiários populares como o *The New York Times*², *Washington Post*³, *The Guardian*⁴ e *Süddeutsche Zeitung*⁵, entre tantos outros, cercavam tais assuntos de superlativos e de uma áurea de solução para os problemas da humanidade. Adjetivos como rápido, eficaz, amplo, verídico, verificável e objetivo descreviam para os leigos o que era o Big Data e o que a chegada de sua ‘era’ ofereceria para empresas, consumidores e instituições públicas. Big Data era aí uma mina infinita de recursos que, se sabiamente utilizados, geraria um melhor capitalismo e uma democracia mais sólida. Obviamente, era fácil reconhecer que tais perspectivas eram primariamente associadas ao mercado, empolgado com a possibilidade de potencializar lucros com extrações de dados mais amplas e análises preditivas aperfeiçoadas. Tal discurso se proliferou também na literatura acadêmica sobre o assunto,

¹ Embora reconheça que alguns autores têm dado preferência a uma tradução do termo para “grandes dados” (e.g. CARREIRA, 2016), ou para o uso do termo no plural “os big data” (e.g. TOMAZ; SILVA, 2018), escolhi utilizar aqui o termo no singular “o Big Data” como uma forma de ressaltar o fato de conceber a expressão não tanto como expressando uma mera superlativação na quantidade dos dados, mas como um certo tipo de fenômeno sociotécnico que informa compreensões diversas do real. Elaboro melhor essa asserção no capítulo 1.

² “How Big Data became so big”. Disponível em <<https://www.nytimes.com/2012/08/12/business/how-big-data-became-so-big-unboxed.html?mtref=www.google.com&gwh=618CCD0E9C6F903C77A70CAB133D6CFA&gwt=pay>> Acesso em: 20/02/2020.

³ “Big Data as the next public good”. Disponível em: <https://www.washingtonpost.com/news/worldpost/wp/2018/05/02/big-data/?utm_term=.e813a6127e6b> Acesso em: 20/02/2020.

⁴ “Big Data: what is it and how can it help?”. Disponível em: <<https://www.theguardian.com/news/datablog/2012/oct/26/big-data-what-is-it-examples>> Acesso em: 20/02/2020.

⁵ “Big Data: Wenn Daten sprechen” Disponível em: <<https://www.sueddeutsche.de/digital/big-data-wenn-daten-sprechen-1.1562758>> Acesso em: 20/02/2020.

especialmente partindo de Ciências Empresariais e da Informação (cf. YOUTIE et al, 2017), que viam o fenômeno a partir das lentes do capitalismo ocidental.

Todavia, o incômodo se ampliou a partir do momento em que identifiquei que fragmentos de tal narrativa apareciam em trabalhos de pesquisadores e acadêmicos pertencentes às mais variadas áreas de estudos dentro das Ciências Sociais, estas, supostamente, comprometidas com um pensamento mais crítico em relação às instâncias sociais, políticas e econômicas. A partir deste incômodo, se materializou a proposta de pesquisa desta dissertação, que é a de entender quais olhares informam as pesquisas em Big Data nas Ciências Sociais. De que perspectivas, lugares e referências partem o que se entende por Big Data nesta área do saber?

A resposta a tais questionamentos permite que se compreendam quais os fundamentos que sustentam tais discussões, quais caminhos têm se consolidado e se projetado para o futuro desta interface de pesquisa e quais vícios têm se incrustado neste entremeio. Evidentemente, uma pesquisa de Mestrado seria curta para o tamanho desta tarefa. Sendo assim, este projeto optou por reduzir o escopo de sua análise a um total de 131 artigos nas seis primeiras edições da publicação científica Big Data & Society (BD&S), revista voltada para fomentar os estudos sobre o Big Data a partir de preocupações relacionadas às Ciências Sociais.

A primeira escolha metodológica foi por um mapeamento analítico bibliométrico que permitisse classificar o material, agrupá-lo a partir de categorias de análise como referências utilizadas, afiliação institucional, palavras-chave, tema de pesquisa, tipos de metodologias usadas, entre outros, e, a partir daí, facilitar uma análise do conteúdo do objeto de estudo. Tal mapeamento fornece, portanto, uma visão geral do que se tem publicado na BD&S. Todavia, sentiu-se a necessidade de discutir os achados do levantamento a partir do funcionamento próprio da revista BD&S como reguladora de determinadas comunidades e propostas científicas de pesquisa (VANDERSTRAETEN, 2010) em Big Data e de que forma o funcionamento deste ‘colégio invisível’⁶ (PAISLEY, 1972) contribui para os achados feitos no mapeamento.

⁶ O termo colégio invisível, que será discutido de forma mais extensa no capítulo 2, foi dado para descrever a atuação de grupos de pesquisadores como parte de um conjunto de sistemas sociais, no qual a interação entre eles desempenha um papel essencial para a sua atuação e relevância dentro de determinada área científica.

Sendo assim, a segunda escolha feita aqui foi por se introduzir criticamente alguns dos principais autores e textos que articulam e informam as pesquisas da BD&S, bem como as comunidades de pesquisa que a estruturam. Além disso, o trabalho optou por trazer um relato de experiência exposto a partir da participação do autor em um evento relacionado a BD&S, realizado em setembro de 2019 na cidade de Bremen, na Alemanha. Tal relato, somada à reconstituição da influência de tal colégio invisível em linhas de pesquisa e grupos de estudo sobre Big Data no contexto das Ciências Sociais, permite uma compreensão mais profunda do contexto de onde se produz, onde circula o que é Big Data nas Ciências Sociais e também do modo em que isso influi nas questões que são colocadas e ignoradas aí. Assim, no recorte desta pesquisa, se busca entender não apenas para onde os olhares se dirigem nas investigações sobre Big Data, mas como funcionam estes corpos e de que formas influem na direção do olhar – e, conseqüentemente, o que ele rejeita – das questões de pesquisa. Espera-se também, através desta dissertação, considerar possíveis tensões ou contradições na relação entre os olhares que a revista busca promover e o modo em que ela mesma se situa.

Isso se mostra importante na avaliação do modo como o recorte analisado aqui ignora certas assimetrias no que tange ao Big Data, tanto no reconhecimento de produções acadêmicas outras, na compreensão do que é Big Data, como na própria identificação desta assimetria que se configura especialmente na relação Norte/Sul Global. Afinal, quais as conseqüências disso? Que implicações traria um reconhecimento destes outros lugares de circulação e produção do que é Big Data e da assimetria nesta relação epistemológica?

Ressalta-se aqui, desde já, a importância do termo ‘olhares’ para esta dissertação. Olhares registrados desta forma, no plural. Não pode haver algo como um olhar apenas sobre o que for. Olhares não existem de forma desencarnada. Olhares não contemplam toda uma paisagem. De fato, reconhecer as coisas como tendo apenas um olhar possível é o exemplo de um olhar particularmente autoritário. Esta dissertação reconhece, portanto, que o saber é situado, frágil, temporário e específico. Sendo assim, entende que

os atores existem em muitas e maravilhosas formas. Explicações de um mundo "real", assim, não dependem da lógica da "descoberta", mas de uma relação social de "conversa" carregada de poder. O mundo nem fala por si mesmo, nem desaparece em favor de um senhor decodificador. Os códigos

Apesar das diversas nuances nas definições do termo e seus usos na pesquisa especialmente de caráter cientométrico, o ponto-chave aqui é *destacar a natureza social da atividade científica e de seu alcance*.

do mundo não jazem inertes, apenas à espera de serem lidos. O mundo não é matéria-prima para humanização; todos os ataques ao humanismo, outro ramo do discurso sobre "a morte do sujeito", deixaram isto muito claro. De certa maneira crítica, isso é grosseiramente apontado pela categoria incerta do social ou de agência: o mundo encontrado nos projetos de conhecimento é uma entidade ativa (HARAWAY, 1995, p. 37).

Assim como a deusa Atena, que figura na escultura como a *σκεπτόμενης*⁷ que lança seu olhar sobre a pedra-marco, a fronteira sob a qual Atenas é erguida, nos mostra, o olhar *divisa*, demarca um território a partir do qual as coisas são reunidas (cf. HEIDEGGER, 2013). O olhar *divisa* o mundo e o situa dentro de uma dada moldura, uma certa fundação. Buscar compreender de onde vem e para onde se dirigem os olhares sobre o Big Data dentro das Ciências Sociais é assim uma tentativa de reconstituir para então desconstruir e reconstruir todo um mundo de possibilidades outrora enclausuradas. Desta forma, esta pesquisa deseja contribuir para ampliar os olhares possíveis para se entender as implicações do Big Data nas sociedades.

Para cumprir este objetivo, esta dissertação está dividida da seguinte forma: **no primeiro capítulo**, apresento o histórico da pesquisa e de como ela chegou na proposta atual, adentro de modo mais amplo no assunto do Big Data e seus campos de estudo no contexto das Ciências Sociais, bem como nas questões metodológicas que orientam a dissertação, especialmente em relação à escolha do objeto de pesquisa e do mapeamento.

No **segundo capítulo**, aprofundo as discussões sobre o campo de estudos em Big Data nas Ciências Sociais apresentando como o comitê editorial da revista BD&S, bem como variadas linhas de pesquisa que fomentam sua publicação, se constituem em um ‘colégio invisível’ cujo impacto não pode ser ignorado no entendimento dos caminhos dos estudos em Big Data. O entendimento dos impactos deste funcionamento é trazido à lume através de uma análise qualitativo-quantitativa: a partir de uma reconstituição das ações e histórico destas comunidades científicas, seja em suas ações formais, como publicações científicas, grupos de pesquisa e editais de financiamento de pesquisa, bem como as informais, como eventos científicos. Para tal, o capítulo apresentará também um relato de experiência que tive como participante da 3ª Data Power Conference, uma conferência sobre estudos críticos dos dados, apoiada pela BD&S e realizada na Universidade de Bremen, na Alemanha, em

⁷ O termo é traduzido por aquele(a) que medita ou contempla. No texto alemão, Heidegger (2013) traduz o termo por *die Sinnende*, que literalmente significa aquele(a) que sente.

setembro de 2019. A análise será intercalada com dados do levantamento realizado por meio do mapeamento analítico na BD&S, que também ajudam a identificar o funcionamento da publicação e que implicações isso traz para a pesquisa em Big Data.

O **terceiro capítulo** dá prosseguimento às discussões do capítulo segundo ao discutir as maneiras em que os cânones da pesquisa⁸, os textos que constituem ‘ponto de passagem obrigatório’, fundamentam os artigos das seis primeiras edições da BD&S na construção de suas abordagens. Se achou por bem delimitar esta análise às quatro principais referências utilizadas, as quais também são revisadas e criticamente avaliadas no texto. Uma das conclusões da análise deste capítulo é a de que, tanto as referências como os modos em que estas são apropriadas nos artigos apontam para uma tendência a se tratar o Big Data e seu contexto empresarial-mercadológico sob uma perspectiva crítica, porém sob argumentos característicos exatamente dos modelos liberais de interpretação dos dados. Argumento, então, que tal se dá porque a perspectiva de tais olhares é o das democracias liberais do Ocidente e do que incomoda e perturba as estruturas do capitalismo histórico de países como Estados Unidos e Grã-Bretanha. A partir daqui a discussão toma o rumo que será o enfoque dos capítulos 3 e do apêndice.

No **quarto capítulo** se discute o ‘fora’ do enfoque das pesquisas da BD&S e do funcionamento desta. Afinal, quais assimetrias este funcionamento ajuda a reverberar? Que questões são estas que ficam de fora? O capítulo aprofunda especialmente dados encontrados na análise da revista que apontam para uma ausência representativa do Sul Global tanto no

⁸ É importante aqui ressaltar o que entendo por cânon, já que o significado e alcance do termo é bastante contestado. O termo vem do grego *κανών*, comumente traduzido por regra, padrão ou norma, embora originalmente se refira a uma vara de medição (THE ANALYTICAL, 1852, p. 212). A palavra ganha realce na tradição ocidental a partir dos debates teológicos acerca do desenvolvimento dos livros sagrados a serem incorporados nas Escrituras judaico-cristãs. Diante de controvérsias doutrinárias e da circulação de obras de caráter ‘espúrio’, as igrejas cristãs, por volta do terceiro e quarto séculos, sentiram a necessidade de validar o que chamavam de ‘cânion’, isto é, a lista definitiva de obras sagradas que deveriam ser tratadas como autoritativas para a fé e prática da comunidade cristã (POIRIER, 2014). A partir de então, a ideia ocidental de cânion se tornou a de um conjunto *consensual* de obras fixas, uma referência de autoridade, a partir das quais todo um conjunto de ideias e práticas é avaliado, rejeitado ou aprovado. Boa parte das críticas a esse desenvolvimento do termo gira em torno da maneira em que parece rejeitar as contradições e disputas inerentes à própria concepção destes. Deste ponto de vista, os cânones têm um fundamento tão sólido quanto o dos sistemas que os sustentam. Prefiro aqui entender os cânones como um conjunto de obras (ou valores, ideias e autores) que assumem um papel, de maneira consensual ou não, de referência (dá a ideia de vara de medição) dentro de determinada comunidade, assumindo em relação a ela uma relação de interdependência. Isto é, as comunidades e suas variações internas transformam o que se toma por cânion, assim como o próprio cânion ajuda a moldar a comunidade e dar a ela seus contornos. Nesta dissertação, me volto particularmente para as obras que assumem este papel dentro das comunidades científicas que guiam o funcionamento da BD&S e de seu projeto.

que se refere à participação nos artigos, nas referências como no Comitê Editorial. Porém, se discute aqui especialmente de que forma essa assimetria influi também numa crescente tensão existente entre a procura por ‘estudos críticos sobre os dados’, alternativos às pesquisas de Big Data relacionadas ao mercado, e a manutenção de categorias e linguagem do próprio modelo que buscam repensar. Considerando também como o Big Data, como prática, como política e como um objeto de estudo, circula no Sul, por quais espaços ele circula? Tal pergunta visa a entender melhor o que perpassa o que se encontra ausente no escopo analisado na BD&S.

É importante recordar que este trabalho reconhece as limitações de sua proposta em que pese o fato de pesquisar uma revista científica que continua em publicação e cujos dados não se encontram em bases de dados como a Scopus ou o Web of Science. De qualquer modo, espera-se com esta pesquisa iluminar ao menos o que foi o período ‘formador’ da BD&S e das questões colocadas ali. E tal exercício retrospectivo pode sempre nos ajudar a compreender o que nos foi destinado⁹.

Por fim, nunca é demais recordar que este trabalho não se encontra sozinho no oceano do conhecimento, mas rema em mares há muito percorridos por importantes pensadores que deixaram marcas decisivas por aqui, por meio de revisão bibliográfica, entre os quais menciono, além de autores diretamente presentes no levantamento: Boaventura de Sousa Santos, Francisco de Oliveira, Payal Arora, Rafael Evangelista, Gilles Deleuze, Gilbert Simondon, Donna Haraway e outros pensadores cuja contribuição já é tida como clássica, que colaboraram, ainda que sob as sombras, para com o espírito crítico desta dissertação, além de muitas outros cujos nomes figuram na lista de referências bibliográficas no final.

Com tudo isto em mente, esta dissertação enseja por contribuir com material bibliográfico relevante sobre os estudos em Big Data, instigar novos desenvolvimentos na pesquisa em Big Data nas Ciências Sociais, compreender o funcionamento de um dos principais centros de pesquisa em Big Data no campo, tanto na via formal de uma publicação científica como na informal de eventos científicos, bem como distinguir ao menos algumas das assimetrias e dilemas presentes nestas pesquisas.

⁹ Aqui não no sentido calvinista de predestinação, mas no sentido de “enviar”.

1. Situando o terreno: percursos da pesquisa e questões metodológicas

1.1. O incômodo

Ser capaz de questionar significa ser capaz de esperar, ainda que por toda uma vida. Mas uma era em que o atual é apenas o que segue rápido e pode ser agarrado com as duas mãos toma o questionar como algo ‘alheio à realidade’, como algo que não conta como útil (HEIDEGGER, 2000, p. 221).

Todo ensejo de pesquisa científica nasce, ou deveria nascer, a partir de um incômodo. O incômodo que move esta pesquisa se materializou através de um processo de reflexão em torno das problemáticas que envolvem a sociabilidade digital. *A priori*, minha preocupação recaía sobre o avanço de uma perspectiva tecnicista de comunicação, que perpetuasse uma dicotomia objetivista de sujeito-objeto, implícita em considerações recentes do campo da Inteligência Artificial. Reagir a este incômodo ocupou boa parte de meus desenvolvimentos acadêmicos iniciais e culminou por ser o tema de minha monografia na graduação (cf. SILVA, 2016; TOMAZ; SILVA, 2018) e a proposta que apresentei para ingresso no Programa de Mestrado em Divulgação Científica e Cultural do Labjor.

Foi somente enquanto matriculado no Labjor que as problemáticas associadas à concepção, produção, uso e controle dos dados se mostraram mais e mais importantes em meu percurso acadêmico. Conceitos recentes como ‘datificação’ e Big Data me chamaram a atenção, porém, em um primeiro momento, pensados apenas em relação com a Teoria da Comunicação e os estudos filosóficos da comunicação, áreas com as quais já nutria afinidade. Até ali ainda me aferrava a um certo ‘pensamento disciplinar’ que se fechava a desdobramentos transdisciplinares - algo já questionado entre cientistas sociais há pelo menos três décadas (cf. SANTOS, 1985). De qualquer modo, me chamava a atenção a forma em que ‘a era do Big Data’ era comumente retratada como o momento histórico onde a teoria e o conhecimento epistêmico (como é tradicionalmente conhecido) ficariam obsoletos ante a eficácia, transparência e objetividade de métodos analíticos envolvendo dados. No contexto dos Estudos em Comunicação, tais abordagens ressaltavam uma maior eficácia e ‘redução de ruído’ a medida em que a comunicação, tida aqui como a mera troca de informação, de dados, fosse cada vez mais informatizada (cf. SILVA, 2018).

Minha motivação, a princípio, era apontar uma crítica às ideias de neutralidade e objetividade presentes na proposição de que os dados tornariam a teoria ou o pensamento inúteis (ANDERSON, 2008). Até aí, nenhuma novidade. Apesar de ser um assunto recente na academia - passando agora por sua primeira década de presença na discussão científica, sendo mais recente ainda sua discussão nas Ciências Sociais - já existe mesmo todo um esforço crescente nos estudos sobre Big Data que se preocupa em ressaltar como os dados, bem como seus métodos e infraestruturas, longe de neutros, “são informados por histórias, ideologias e filosofias específicas” (ILIADIS; RUSSO, 2016, p. 2). Foi com este pressuposto que decidi tomar como objeto de análise as seis primeiras edições da revista *Big Data & Society* (BD&S), dedicada a pensar o Big Data a partir de contribuições das Ciências Sociais, majoritariamente. Imaginava que encontraria um cenário onde a maior parte dos cientistas sociais adotariam uma postura questionável ante os dados, de fato fomentando noções de neutralidade dos dados mais afeitas às Ciências Exatas, onde via o Big Data prevalecer como solução e não como problema, dados sendo primeiro aplicados, depois problematizados (CROLL, 2012). Antes mesmo de analisar a revista, já buscava questionar as supostas escolhas metodológicas dos cientistas sociais.

Com o tempo, e com o auxílio de discussões realizadas em algumas das disciplinas cursadas no decorrer do Mestrado, minha postura necessitou ser revisada por ao menos duas razões: a primeira foi o contato inicial com o objeto de estudo. A avidez por englobar as pesquisas ali contidas num pacote que permitiria uma crítica de minha parte caiu por terra ante a diversidade de preocupações e temáticas encontradas ali, bem como ao perceber no material investigado um senso de abertura na pesquisa em Big Data a uma crítica de visões neutras e objetivas dos dados (KITCHIN, 2014). Os diferentes trabalhos com suas distintas metodologias, envolvendo abordagens teóricas, *survey*, estudos de caso, análise de rede social, etnografia, mapeamento de controvérsias etc., derrubaram minha hipótese inicial e direcionaram meu trabalho para a necessidade de revisar meus problemas de pesquisa iniciais e a maneira como iria respondê-los.

Afinal, o que faria com toda a base de dados e análise feita com o material da BD&S? A resposta viria juntamente com uma mudança de postura ante minhas pretensões iniciais. Se antes me imaginava numa posição de ‘cérebro das humanidades’, um ‘carrasco’ da objetividade, foi a partir da leitura de uma obra aparentemente desconexa dos meus

trabalhos, no contexto das reuniões do grupo de pesquisa ICTS (Informação, Comunicação, Tecnologia e Sociedade), no primeiro semestre de 2018, que comecei a repensar minha posição como pesquisador. Devo a Pignarre e Stengers, naquele momento, o despertar do meu ‘sono dogmático’. O papel dos *sounders of the depths*¹⁰ e das trajetórias do aprendizado, evocados com o grito por um outro mundo que Seattle deixou, me alertou acerca do risco que é apontar que um outro mundo é possível a partir de uma posição profética, ou nas palavras deles, de um “teorista”, amparado por jargões como “nós temos que” ou “esse é o caminho” (PIGNARRE; STENGERS, 2011, p. 9) . Afinal, como bem ilustram, mesmo quem se coloca na proa do barco para avistar o cenário ainda está no barco, ainda está em perigo. Não que o perigo de discursos hegemônicos que retiram seu poder de uma definição de neutralidade instrumental associada aos dados seja irrelevante. Os próprios apresentam críticas a tal afirmativa. Todavia, sabia que precisava ter atenção ao lugar de fala, ao corpo do meu próprio olhar, situando bem o meu chão assim como o de outros que pretendia visitar.

O próximo passo, então, envolveu a decisão de gastar mais energia na descrição, no mapeamento, do que o material me revelava. Desde então, os incômodos se multiplicaram. Largar a meta de ser o profeta enviado para denunciar os pecados da neutralidade possibilitou que atentasse para os diversos modos de construção do que é Big Data, bem como dos caminhos e tendências que se configuravam nas pesquisas sobre Big Data no material analisado. Somente a partir daí, o incômodo inicial, potencializado em distintas direções, se materializaria em problemas de pesquisa. Olhar apenas para a presença da neutralidade ou objetividade na pesquisa em Big Data nas Ciências Sociais se me apresentou como um objetivo pobre, mesmo entendendo que este e tantos outros vícios poderiam estar presentes nestes estudos.

O desenrolar da pesquisa, a partir de comentários feitos em reuniões de orientação e participação em congressos científicos, o qual destaco especialmente o VI Congresso Internacional Lavits, me levou ainda a um outro tipo de reflexão que não fazia parte de minhas preocupações iniciais: o status da BD&S e das comunidades científicas ali

¹⁰ Pensadores e filósofos são repensados aqui por Pignarre e Stengers (2011) como sendo não aqueles que se colocam de fora do caos do real para pensá-lo mais claramente e propor fundamentos que o orientem em direção a um eterno presente. Pensar é se colocar no caos, não traçar ou escolher direções, como colocam Pignarre e Stengers (2011). Um pensador crítico é aquele que arria o seu prumo em direção ao oceano para medir o tamanho do estrago e poder apenas alertar intuitivamente o tamanho do perigo que nos aguarda.

representadas e seu modo de funcionamento. Afinal, percebi que tentar considerar os olhares que informam as pesquisas em Big Data nas Ciências Sociais envolvia questões que vão muito além do que um mapeamento bibliométrico poderia oferecer. Não considerar isso acabaria por colocar a BD&S em uma posição que acabaria por reforçar as assimetrias que o próprio mapeamento apontou. Afinal, o mapeamento diz pouco, sozinho, sobre as lógicas do ‘colégio invisível’ que constitui a revista e a maior parte das colaborações ali contidas.

Desta forma, a etapa descritiva do mapeamento se torna agora informada pela constituição das comunidades que formam a BD&S. Minha participação em um congresso promovido pela BD&S e que se apresentou como um ponto de encontro entre os diversos grupos que compõem o Comitê Editorial da publicação, bem como autores que ali publicavam ou intencionavam publicar seus artigos, acabou por se tornar uma importante fonte de informação sobre como o funcionamento das comunidades que identifiquei no levantamento analítico acabam por definir os olhares, e aquilo que fica de fora deles.

O que se segue é, portanto, uma tentativa de fornecer uma mirada introspectiva, para o próprio fazer científico das Ciências Sociais em relação ao Big Data e as implicações sociais dos dados. Este é o incômodo que orientou esta pesquisa. Não se espera aqui ofertar respostas definitivas sobre o tema. Mas este incômodo, assim como muitos outros que desde já moveram inúmeros outros pesquisadores, pode certamente proporcionar futuramente olhares renovados. Que outras mentes, e mesmo a minha posteriormente, continuem a habitar e construir sobre esse solo.

1.2 O assunto

Desde o fim do século passado, o nível de interesse para com o que se convencionou chamar de Big Data tem crescido. Embora o uso do termo regresse até os primórdios da computação, foi somente nos últimos trinta anos que a expressão passou a se referir a uma conjuntura sociotécnica na qual os algoritmos e o fluxo de dados passaram a exercer um papel todo abrangente no contexto sociopolítico global. Big Data se tornou uma espécie de *catchphrase*, termo da vez para englobar questões que abrangem desde o crescimento contínuo das bases de dados, bem como a progressiva importância de análise de dados não somente para cientistas da computação ou engenheiros, mas também para

cientistas de todos os campos, além é claro do realce das ‘inúmeras’ possibilidades de lucro e capitalização que aquecem os sonhos de sucesso do mercado¹¹ (TOMAZ; SILVA, 2018).

Os primeiros usos da expressão Big Data são usualmente atribuídos ao jornalista e escritor estadunidense Erik Larson, quando em julho de 1989, escreveu uma crítica em um artigo publicado na Harper’s Magazine acerca da quantidade de spams que recebia em sua caixa de entrada do e-mail, reclamando dos usos que os “guardiões do Big Data” estavam a dar aos seus dados¹². Steve Lohr (2013), escrevendo para o New York Times, argumenta, todavia, que a conotação atual do termo surge no Vale do Silício nos anos 1990. Segundo Lohr, cientistas como John Mashey já utilizavam o termo no contexto de conversas informais e reuniões de trabalho de empresas do ramo tecnológico¹³. “O termo Big Data [...] provavelmente se originou de conversas de almoço na Silicon Graphics Inc. na metade dos anos 90, tendo em John Mashey uma figura proeminente”, assim concluiu Francis Diebold (2012) um artigo sobre a origem e desenvolvimento do termo Big Data.

O artigo “*Managing Big Data for Scientific Visualization*”, de Cox e Ellsworth (1997) é, possivelmente, o primeiro relato acadêmico apresentando o termo, sendo ambos à época funcionários da Intel e do Centro de Pesquisa Ames, da Nasa. É a primeira tentativa de definir o que chamam de “problema do Big Data”, classificando-o em objetos de big data, amplos demais para processamento em apenas uma máquina, e coleta de big data, abrangendo a qualidade das bases de dados e da análise destes. O contexto, tanto do artigo de Cox e Ellsworth (1997) e das apresentações de Mashey, porém, é o mesmo. Resolução de problemas relacionados à aplicações empresariais ou governamentais. Pode-se dizer, portanto, que *o contexto genealógico do Big Data é o da cultura empresarial do Vale do Silício*.

Recentemente, a Google lançou oficialmente o Data Transfer Project, englobando também gigantes da internet como Facebook, Twitter e Microsoft, com o

¹¹ Basta perceber o tom otimista acerca do investimento em análise de dados por parte de grandes companhias presente em relatórios e pesquisas de taxa de sucesso empresarial. Ver, por exemplo: <<https://www.statista.com/statistics/742935/worldwide-survey-corporate-big-data-initiatives-and-success-rate/>>. Acesso em: 20/02/2020.

¹² O artigo original pode ser acessado através do link: <<https://harpers.org/archive/1989/07/what-sort-of-car-rt-sort-am-i-junk-mail-and-the-search-for-self/>>. Acesso em: 20/02/2020.

¹³ Há mesmo um link que leva a uma apresentação de slides preparada e apresentada por Mashey em uma reunião de negócios datada de 1998. Disponível em: <https://static.usenix.org/event/usenix99/invited_talks/mashey.pdf> Acesso em: 20/02/2020.

propósito de reunir sob uma única plataforma toda a base de dados de usuários dos serviços das quatro companhias¹⁴. No documento oficial deste que já é o maior projeto de portabilidade de dados do qual se tem notícia, não deixam de aparecer termos como “velocidade”, “eficiência”, “tamanho” e “transparência” como resultados naturais de tal iniciativa¹⁵. Não apenas nas iniciativas privadas, como também a nível governamental projetos envolvendo big data como o Global Earth Observation System of Systems (Geoss)¹⁶, nos Estados Unidos, trabalham sob o fundamento da melhora na ‘capacidade’, ‘eficácia’ e ‘transparência’ no manejo de desafios nacionais (KIM et al., 2014).

De fato, definições do Big Data e do que ele representa, em linhas gerais, têm sido elaboradas em torno do que ficou conhecido como os três Vs (LANEY, 2001): velocidade, variedade e volume. Na verdade, faltaria espaço para compartilhar as *v-words* já utilizadas para dar conta do que seria o Big Data: versatilidade, volatilidade, virtuosidade, vitalidade, vigor, virilidade, valor, veracidade, variabilidade, entre tantas outras (KITCHIN; MCARDLE, 2016; UPRICHARD, 2013). Todas aparecem em contraste com uma situação anterior, na qual, por exemplo, as informações obtidas sobre hábitos de vida, dados de geolocalização e informações de transações financeiras por parte de consumidores poderiam ser obtidos apenas em pequenas amostras, geralmente feitas por entrevistas, e que formavam pequenas bases de dados. Com a possibilidade de personalização, a ampliação escalar das bases de dados e um escopo de amostra quase ‘universal’, todos os superlativos possíveis descreveriam o que se apresentava para o nosso tempo. Em todo o caso, “a expressão passou a se referir a uma conjuntura sociotécnica” na qual a análise e o fluxo de [grandes] dados “passaram a exercer um papel todo abrangente no contexto sociopolítico global” (SILVA, 2018, p. 109). Todavia, é importante lembrar que chamar os dados de grandes, complexos ou variados ainda diz pouco sobre o que são os dados. Nas palavras de boyd¹⁷ e Crawford (2012, 663), “Big Data é, de diversos modos, um termo pobre”.

¹⁴ “Google, Facebook, Microsoft, and Twitter partner for ambitious new data project”. Disponível em: <<https://www.theverge.com/2018/7/20/17589246/data-transfer-project-google-facebook-microsoft-twitter>> Acesso em: 20/02/2020.

¹⁵ “Data Transfer Project: from theory to practice”. Disponível em: <<https://services.google.com/fh/files/blogs/data-transfer-project-google-whitepaper-v4.pdf>> Acesso em: 20/02/2020.

¹⁶ “Geoss Portal”. Disponível em: <<http://www.geoportal.org/>> Acesso em: 20/02/2020.

¹⁷ boyd entrou na Justiça estadunidense em 2011 e ganhou o direito de poder grafar seu nome e sobrenome em minúsculo. Mais detalhes sobre os motivos da mudança estão em: <http://www.danah.org/name.html>

Cornelius Puschmann e Jean Burgess (2014), num estudo sobre as metáforas construídas em torno do Big Data na mídia apontaram para uma tendência de posicionamento do Big Data como instaurador de uma nova era, capazes de romper com antigas molduras de pensamento, ética e crenças, bem como de uma força natural a ser controlada e usufruída pelo ser humano (PUSCHMANN; BURGESS, 2014) (MAYER-SCHÖNBERGER; CUKIER, 2013). Outros chegam a ver o Big Data ou lógicas regidas pelos dados como uma revolução epistemológica capaz de derrubar concepções tradicionais do conhecimento, como a famigerada relação causa-efeito, já que a primeira lidaria com informações brutas, repetíveis em diversos modelos (DUBOIS et al., 2000). O fim da teoria chegou, já foi dito (ANDERSON, 2008). Isso levou boyd e Crawford (2012), avaliando todos estes aspectos do fervor sobre o Big Data, a apontarem, juntamente com elementos técnicos e analíticos, para uma certa mitologia¹⁸ como sendo também constituinte do Big Data.

Apesar de fortemente atrelado, no imaginário popular, à computação, o termo *data* já estava ligado às ciências desde a Grécia Antiga. Em sua raiz etimológica, a expressão se refere simplesmente a algo que é dado, um presente (PUSCHMANN; BURGESS, 2014). No contexto científico, o termo passou a cunhar, literalmente, algo que se encontra dado, que está dado à mão, disponível, a “substância a ser analisada e sondada à qual é assumida uma essência prévia” (SILVA, 2018, p. 109). Este é exatamente o que Rafael Capurro (2006, p. 4) apontou como o caráter paradigmático da ciência matemática, não a aplicação de regras matemáticas à processos naturais, mas “a projeção a priori das entidades que esta descobre”. Como Melanie Swan (2015, p. 469) aponta com relação a esse tipo de compreensão, “dado é o fato” que provê a “base para o pensamento ou cálculo”.

Obviamente, o falatório acerca dos dados não se resume ao campo das Ciências da Computação. A presença do Big Data como tema de publicações, edições especiais de revistas que perpassam desde economia e administração até saúde pública, ilustra sua abrangência na contemporaneidade. O que dizer, então, de sua presença nos estudos das Ciências Sociais? Como essas questões têm sido discutidas ali? Afinal, por que a análise se

¹⁸ As autoras entendem mitologia aqui como um certo tipo de “fé largamente difundida de que os dados ofereceram uma forma superior de inteligência e conhecimento” (BOYD; CRAWFORD, 2012, p. 663). O artigo de boyd e Crawford, considerado seminal nos estudos em Big Data, será pormenorizadamente tratado no capítulo 3.

volta para as perspectivas sobre Big Data a partir de cientistas sociais e não de engenheiros ou cientistas dos dados?

1.3 O campo

A relação entre as Ciências Sociais e os dados não é recente. Na verdade, boa parte da atividade da Sociologia no século XX foi ancorada em metodologias que lidavam com dados e estatísticas – de fato, métodos quantitativos de análise já apareciam inclusive em autores clássicos como Émile Durkheim (RAFTERY, 2001) em um esforço de consolidação das ciências da sociedade como ciências *verificáveis* e passíveis de análise estatística objetiva, status que reivindicavam as ciências exatas e naturais (RODRÍGUEZ, 2012). À medida em que o interesse por *surveys* e dados de grupos populacionais aumentava, crescia também a confiança no uso de bancos de dados e de estatísticas para um entendimento ‘mais acurado’ das dinâmicas sociais. O interesse pelo uso de dados nas análises sociais era também parte de projetos de poder dos Estados-nação à época, que viam na análise estatística social uma oportunidade de ‘aperfeiçoar’ a nação e seus membros dentro do sistema de fluxo do capital (MARTIN, 2001)¹⁹.

Isso também é acompanhado de uma compreensão crescente do social em termos mecanicistas e objetificantes, na onda do movimento transdisciplinar da cibernética, a partir dos trabalhos de Norbert Wiener, Claude Shannon e Warren Weaver, bem como do impacto das Conferências Macy, entre o final da década de 1940 e o começo dos anos 50²⁰. Sendo assim, se apresentava como um dever vital para as Ciências Sociais a reunião do maior número de dados sobre os grupos sociais a fim de se fornecer uma maior compreensão e previsibilidade do funcionamento do ‘organismo social’.

¹⁹ No capítulo “O nascimento da medicina social” do clássico “Microfísica do Poder”, o filósofo francês Michel Foucault relaciona o advento da análise social estatística às políticas mercantilistas dos primeiros Estados modernos a partir do século XVI. Dados sobre natalidade, mortalidade e nível de saúde da população, por exemplo, emergem em um contexto de uma “política mercantilista [que] consistia essencialmente em majorar a produção da população, a quantidade de população ativa, a produção de cada indivíduo ativo e, a partir daí, estabelecer fluxos que possibilitem a entrada [...] da maior quantidade possível de moeda” (FOUCAULT, 1998, p. 82).

²⁰ É importante lembrar que um dos elementos centrais da teoria cibernética se firmava na ideia de que a fim de se minimizar os efeitos da entropia, isto é, da mudança de um estado de ordem para um de desordem nos organismos, era necessária uma maior informatização. Isso passava por uma troca de informações e feedbacks cada vez mais eficiente, com o menor ruído possível, para a maior eficácia de todo o sistema mecânico. Para mais detalhes sobre a influência da cibernética nas ciências, ver SILVA, 2016.

Tal engajamento, entretanto, resultou numa apropriação acrítica do papel da estatística e dos dados na análise sociológica, como bem atestou Labovitz (1972) ao usar a linguagem de vacas sagradas como metáfora para a crença presente entre cientistas sociais de seu tempo de que as Ciências Sociais somente poderiam operar sua práxis em conjunto com análises quantitativas:

Com relação ao uso da estatística, existem duas vacas sagradas na sociologia: (1) o tratamento de premissas como sendo invioláveis, e (2) o uso de estatísticas como fim em si mesmas. Estas são classificadas como vacas sagradas porque parecem ser tratadas como acima de qualquer repreensão e são aceitas por muitos sociólogos sem qualquer pensamento crítico ou debate (LABOVITZ, 1972, p. 17).

Com isto se pretende afirmar que o afã sobre uso de estatísticas derivadas e dedução causal para teste de hipóteses como estandartes do entendimento ‘real’ do funcionamento do social cegou boa parte dos sociólogos para a análise das próprias categorias que informavam seus estudos quantitativos, como raça, gênero, orientação sexual, entre outras (MCFARLAND et al, 2015). Foi nesse contexto, ao menos na sociologia *mainstream*²¹, que novas possibilidades de análises de dados, potencializadas com o Big Data, se apresentaram aos cientistas sociais.

Enquanto a preocupação em geral tem recaído sobre o impacto do Big Data na eficiência, alcance e apuração das análises sociais (c.f. MADSEN et al, 2016; SAVAGE; BURROWS, 2007), o questionamento político acerca da sociologia de dados tem recebido menor atenção, como elucida Deborah Lupton (2015, p. 101):

Os dados digitais que são transmitidos em números por tecnologias digitais são tanto os produtos de dispositivos sócio técnicos como eles mesmos são os próprios dispositivos, possuindo sua agência e poder próprios. Não existe algo como dado ‘bruto’.

Portanto, o levantamento proposto nesta dissertação, e mesmo o próprio ensejo investigativo que se põe aqui, de olhar para a pesquisa em Big Data nas Ciências Sociais, se

²¹ Me refiro aqui especialmente às escolas sociológicas britânicas e estadunidenses. A primeira historicamente enraizada sob análises quantitativas e a segunda bastante ligada a pesquisas de nível militar e mercadológico, que envolviam também larga utilização de estatísticas e bases de dados (WILLIAMS et al, 2008; CROMPTON, 2008; MCFARLAND et al, 2015).

coloca no intermeio de uma longa – e complicada – relação entre o uso e análise de dados por parte de cientistas sociais e como isso influenciou na própria construção das Ciências Sociais nos últimos anos. A discussão de fundo que se põe é a de quais estruturas compõem os dados? De onde estes vêm? Tal relação fica ainda mais latente com o avanço do Big Data e novos ferramentais que possibilitam análises mais complexas como as de rede (WAGNER, 2008), ao mesmo tempo em que relações de exploração e vigilância atreladas aos dados também ficam patentes (cf. ZUBOFF, 2019).

É neste contexto de tensão que se localiza o escopo deste trabalho e o levantamento quantitativo apresentado nas próximas páginas. Tensão esta que se apresenta peculiar às Ciências Sociais (ILIADIS; RUSSO, 2016), apesar dos estudos em Big Data mobilizarem um amplo espectro de esforços disciplinares e transdisciplinares. Basta notar a quantidade de projetos de pesquisa²² e eventos acadêmicos²³ que giram em torno do tema em campos tão diversos como Inteligência Artificial, Administração, Geografia e Saúde Pública em relação com o Big Data.

Apesar da relação histórica entre dados e estudos sociológicos, até o momento, poucas foram as tentativas de se mapear os estudos em Big Data nas Ciências Sociais. Os levantamentos têm contemplado questões como o estado da pesquisa empírica em Big Data (WIENHOFEN et al, 2015), definições do Big Data dentro dos estudos em Administração (YLIJOKI; PORRAS, 2016) e das chamadas Humanidades Digitais (KAPLAN, 2015). Porém é com os esforços de Jan Youtie, Alan L. Porter e Ying Huang (2017) que surge o primeiro levantamento voltado especificamente para a pesquisa em Big Data nas Ciências Sociais. O objetivo dos pesquisadores era averiguar a distribuição de interesses de pesquisa e de subliteraturas dentro da pesquisa em Big Data nas Ciências Sociais a partir de um extrato de 488 artigos retirados do acervo da Web of Science (WoS). Uma das contribuições do mapeamento de Youtie et al (2017) é que este reafirma a relativa novidade da pesquisa em Big Data na área, tendo esta alcançado o total de 600 artigos apenas a partir de 2012, quando

²² Cito, como exemplo, iniciativas transdisciplinares e transterritoriais em torno do Big Data como o *Data Diplomacy*, que envolve instituições acadêmicas de Nova Zelândia, Austrália e Inglaterra, que investiga o papel da diplomacia com relação ao acesso e compartilhamento de dados; e o *Data Institute*, sediado na Universidade de São Francisco, EUA, e dedicado a formar cientistas de dados de todas as partes do globo.

²³ Em sites de divulgação de chamadas de trabalho para eventos científicos, a temática do Big Data aparece entre as mais presentes. Ver, por exemplo: <<https://www.papercrowd.com/conferences/search>> Acesso em: 20/02/2020.

o tema recebe um *boom* nas Ciências Sociais. Este ano marca a publicação, identificam Youtie et al (2017), daquele que se mostrou um dos cânones do campo até o momento da publicação do levantamento, o artigo “*Critical Questions for Big Data*”, publicado por danah boyd e Kate Crawford (2012) na revista *Information, Communication & Society*.

Além de agrupar as referências mais citadas na pesquisa em Big Data nas Ciências Sociais, que incluem boyd e Crawford (2012), trabalhos como o de Lazer et al (2014) e o livro “*Big Data*”, de Mayer-Schönberger e Cukier (2013), Youtie et al (2017) identificaram também temas que funcionam como articuladores centrais de pesquisa, como: Internet e Sociedade, Estudos em Privacidade, Sociologia da Ciência, Big Data e Medicina, Geolocalização, Tomada de Decisão, Impactos Empresariais e Analítica/Software.

Outro mapeamento analítico a contribuir para uma maior compreensão da pesquisa em Big Data nas Ciências Sociais foi o de Akoka et al (2017), que tratou de investigar o terreno por onde circularam as pesquisas em Big Data entre os anos de 2013 e 2016 dentro da base de dados do ScienceDirect. Chama a atenção, além do número de artigos levantados, um total de 1843, o tamanho dos estudos em Big Data – com 24 áreas distintas representadas - bem como a presença marcante da Ciência da Computação e das Engenharias entre as publicações, alcançando 935 artigos dentre o total (50,73%), mais do que metade de todas as outras 22 disciplinas somadas, seguidos de Administração e ciências de decisão, e Ciências Sociais. Isso ajuda a explicar o fato de aspectos técnicos e performativos [qualidade, eficiência, velocidade etc.] aparecerem entre os termos mais pesquisados, ainda segundo Akoka et al (2017).

Além da relação histórica das Ciências Sociais com a estatística e da identificação das primeiras abordagens sobre Big Data no campo a se ter em conta na pesquisa, é preciso reconhecer também a importância de uma geopolítica da produção científica sobre o Big Data para melhor compreender o campo do Big Data. Apesar da tradição acadêmica de países europeus como Inglaterra, Alemanha e França, por exemplo, bem como os Estados Unidos, é importante discorrer sobre a importância de movimentos como o da China²⁴, que ao menos

²⁴ Um exemplo que pode ser colocado neste sentido é a importante participação de universidades chinesas no fomento de discussões sobre Big Data a nível global. Isso é perceptível, por exemplo, na quantidade de periódicos transdisciplinares internacionais sob liderança de pesquisadores e instituições chinesas, tais como o *International Journal of Big Data Intelligence* (National Chung Cheng University), *Big Data and Cognitive Computing* (Huazhong University of Science and Technology), *Big Data Research* (Zhejiang University) e o *Open Journal of Big Data* (Xi'An Jiaotong Liverpool University).

na área de Finanças e Administração tem se equiparado aos Estados Unidos em produção acadêmica em Big Data²⁵ - e não apenas no âmbito acadêmico, vale lembrar²⁶.

Apesar de não se limitar à esfera científico-acadêmica, o mapa da Figura 1 compreende um esforço de expressar a balança de poderes da geopolítica da produção, uso e impacto do Big Data no globo. Percebe-se, desde já, uma forte tendência de concentração euroamericana, bem como a expressividade do Big Data em locais como China, Índia, Coreia do Sul e Japão. Por outro lado, América Latina e África aparecem em um nível de quase inexpressividade na balança do Big Data. Esta assimetria é um dos questionamentos que perpassa toda esta dissertação, afinal não se trata de uma problemática menor. Ela impacta conceitos, teorias e métodos que aparecem no horizonte do que é o Big Data atualmente (MILAN; TRERÉ, 2019). Tal questão, também constatada no levantamento feito aqui, será discutida posteriormente, a partir do entendimento do funcionamento das comunidades científicas que constituem a BD&S.

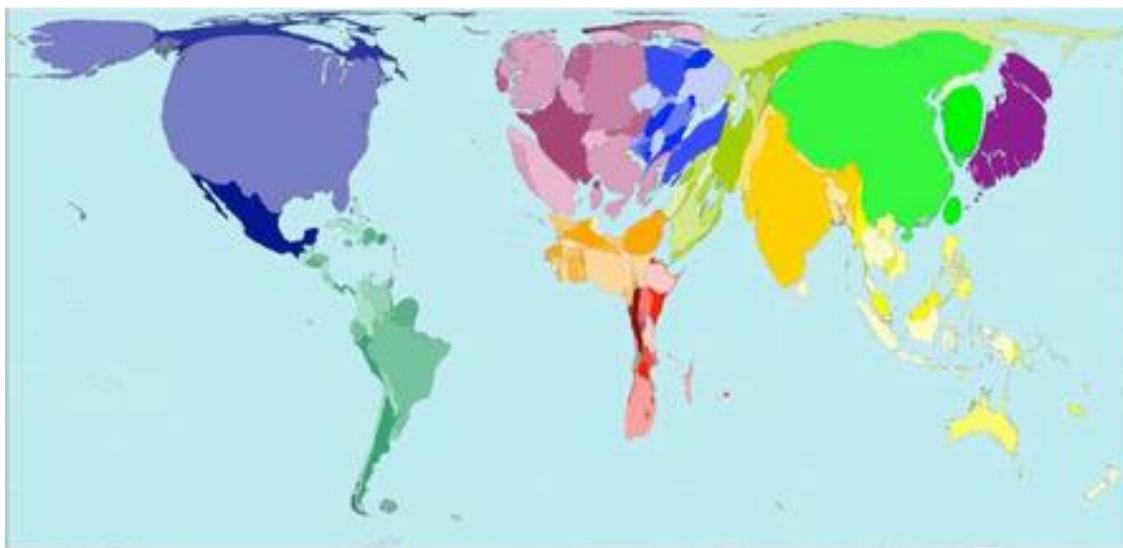


Figura 1: Mapa geopolítico da balança global do Big Data. Elaborado pela cientista computacional Stéphane Grumbach, do Institut National de Recherche en Informatique et en Automatique (Inria/França)²⁷

²⁵ Infelizmente, não existe ainda um mapeamento geral envolvendo toda a produção acadêmica em Big Data que contemple questões geopolíticas. O mapeamento desta dissertação contempla aspectos geopolíticos, porém apenas no que se refere ao campo das Ciências Sociais conforme uma única publicação, a *Big Data & Society*.

²⁶ “China and US compete to dominate big data”. Disponível em: <<https://www.ft.com/content/e33a6994-447e-11e8-93cf-67ac3a6482fd>> Acesso em: 20/02/2020.

²⁷ “Big Data? The Global Imbalance”. Disponível em: <<http://www.in2p3.fr/actions/formation/Info13/lift-27sept-121001065613-phapp01.pdf>> Acesso em: 20/02/2020.

O ponto-chave aqui é ter um panorama geral dos lugares de circulação da pesquisa em Big Data, o que, conseqüentemente, impacta os diversos modos em que este é entendido. Como visto, as principais constatações da fase primária deste trabalho foram a: i) emergência do Big Data em um contexto de, por um lado, progressiva confiança dentro das Ciências Sociais no uso de métodos estatísticos para compreensão da sociedade e, por outro, de um crescente reconhecimento, por parte também de grupos de cientistas sociais, de dados e estatísticas como entidades que imbricam relações de poder que não podem ser resumidas à uma despreziosa neutralidade no seu uso; ii) a interdisciplinaridade do campo de estudos do Big Data, se é que pode-se falar de um ‘único’ campo ou mesmo da existência de um ‘campo’ do Big Data; e, mesmo com a amplitude dos esforços em torno do tema, iii) a concentração de produção de literatura científica nos campos técnicos da Ciência da Computação e das Engenharias, bem como iv) a centralização euroamericana na produção, uso e concepção do Big Data. Tais constatações são importantes no sentido de situar a pesquisa em Big Data nas Ciências Sociais dentro de um terreno maior, terreno este ocupado por certos tipos de discurso e instituições. Para delinear mais claramente esta interface relativamente pouco explorada entre Ciências Sociais e Big Data será necessário primeiramente introduzir o objeto de estudo desta dissertação, a revista BD&S, pioneira no fomento da investigação do Big Data entre os cientistas sociais e um conhecido lugar de circulação destes estudos, bem como os grupos e lugares pelos quais ela é constituída. Isso será feito no capítulo 2 desta dissertação. Antes é importante clarificar alguns detalhes metodológicos.

1.4 Metodologia

Caminhando nas tensões entre perspectivas qualitativas e quantitativas do Big Data nas Ciências Sociais, esta dissertação adota uma estratégia metodológica quali-quantitativa. Esta estratégia, comumente nomeada de ‘*mixed methods*’, ganhou proeminência a partir do final do século 20 com o entendimento de que “a crescente complexidade dos problemas de pesquisa, a percepção de que as pesquisas qualitativas são um meio legítimo de investigação e um ensejo por parte de pesquisadores qualitativos em ampliar o alcance de suas descobertas” (TIMANS; WOUTERS; HEILBRON, 2019, p. 198) clamavam por uma abordagem metodológica que contemplasse tanto abordagens qualitativas, como entrevistas

ou relatos etnográficos, como quantitativas, como coleta e análise de grandezas numéricas. Aqui nesta dissertação, particularmente, o uso da metodologia quali-quantitativa se justifica por explorar conjuntamente visões distintas de um mesmo fenômeno: o funcionamento e atuação da BD&S na pesquisa em Big Data nas Ciências Sociais (cf TEDDLIE; TASHAKKORI, 2009).

Além da presença óbvia de revisão bibliográfica sobre o Big Data e seus meandros nas diversas ciências, o que foi feito neste capítulo introdutório, o trabalho é informado por um mapeamento analítico de cunho bibliométrico realizado sobre a revista BD&S. O mapeamento, que se mostrou uma abordagem interessante para obter um panorama geral do objeto estudado e, a partir daí, levantar diversas questões mais específicas sobre novas formas de interpretar dada literatura (PIRYANI et al., 2016; LUCHETTA, 2018), foi realizado de forma manual²⁸ em seis das treze edições da publicação (até abril de 2020). Esta metodologia tem sido amplamente utilizada em contextos de revisão bibliográfica (JEREZ-ROIG et al., 2014; CONFORTO et al., 2011).

A primeira etapa do levantamento consistiu em definir as principais variáveis a serem analisadas. Foram elas: autores, título dos artigos, palavras-chave, país de origem da pesquisa (da instituição a que se filiam os autores), áreas de pesquisa dos autores, objetos de análise, principais referências utilizadas nos artigos, tipo de pesquisa (teórica, estudo de caso, etnografia, entrevista etc) e dossiês temáticos publicados no período.

A partir daí me dediquei a investigar todos os artigos das edições supracitadas a fim de coletar estas informações. Tais esforços envolveram *a leitura de todo o material*. No caso das referências, este levantamento contempla apenas as que aparecem mais de uma vez nos artigos, constituindo-se, assim, das referências mais importantes para a elaboração dos artigos na BD&S. O resultado deste esforço foi compilado e representado visualmente com o auxílio do software open-source de text-mining Voyant Tools²⁹. Os gráficos, tabelas e índices serão apresentados e discutidos no próximo capítulo.

²⁸ É preciso esclarecer aqui que, por manual, nomeio um processo que não contou com o auxílio de dados indexados em databases como Web of Science ou Scopus, e nem de analíticas ou ferramentas próprias da revista. Todo o processo que se seguiu desde a criação das categorias de classificação, levantamento de dados até o tratamento dos dados foi realizado por mim. Mesmo o uso mais comum do termo manual se aplica ao esforço realizado aqui, já que mesmo páginas de bloquinhos de anotações serviram de instrumentos no decorrer da pesquisa.

²⁹ Disponível em: <<https://voyant-tools.org/>> Acesso em: 20/02/2020.

Todavia, a pesquisa não se encerra na descrição de um ‘estado da arte’ da pesquisa na BD&S. É a proposta desta dissertação entrar mais a fundo nos informes do ‘pequeno’ banco de dados levantado e ampliar as questões para entender como se construiu o que aparece ali. Se a etapa quantitativa visava contemplar a questão “quais são os principais temas, referências e enfoques da pesquisa em Big Data no âmbito das Ciências Sociais?”, a etapa qualitativa pergunta pelo “por quê?”, “como?”, “em que sentido?” e “para quem?”.

Em certo sentido, e não no de uma simples filiação metodológica, o trabalho busca fazer uma análise do conteúdo, apenas no senso de que busca fundamentos e modos de construção de tendências e padrões presentes na literatura avaliada (KRIPPENDORF, 1989). Para tais considerações, é preciso lembrar que o próprio mapeamento provê auxílio bibliográfico para uma análise mais crítica, a partir das referências que informam os artigos da BD&S. Porém, acima de tudo, se entende aqui que as respostas a tais questionamentos dependem de um acesso a informações disponíveis no envolvimento de campo com o objeto de estudo. Para isso, a pesquisa traz um relato de experiência que tive ao participar da 3ª edição da Data Power Conference, um evento organizado por pesquisadores ligados a departamentos de instituições como a Universidade de Bremen, na Alemanha; de Sheffield, no Reino Unido; e de Carleton, no Canadá. A conferência, realizada em Bremen, na Alemanha, forneceu uma oportunidade única para o andamento desta dissertação visto que é um ponto de encontro de vários dos membros do Comitê Editorial da BD&S, além de ter sido ativamente promovida por esta – inclusive com a publicação iminente dos anais do evento em uma edição futura da revista.

O relato de experiência, como parte constitutiva do método etnográfico (GUBER, 2001), fornece a possibilidade de uma análise mais complexa do funcionamento de determinado grupo ou contexto social a partir das relações e interações cotidianas constituídas ali. Nele, a experiência vivida pelo pesquisador em dado contexto é transposta para a análise de determinado fenômeno. A relação entre a experiência e o texto é bem exposta por Strathern (2014, p. 345-6):

Um dos elementos que torna o trabalho de campo desafiador é ele ser realizado tendo em mente uma atividade muito diferente: a escrita. [...] as ideias e as narrativas que conferiam sentido à experiência de campo cotidiana têm de ser rearranjadas para fazer sentido no contexto dos argumentos e das análises dirigidos a outro público. Em vez de ser uma

atividade derivada ou residual, como se pode pensar de um relatório ou reportagem, a escrita etnográfica cria um segundo campo. A relação entre esses dois campos, portanto, pode ser descrita como "complexa", no sentido de que cada um deles constitui uma ordem de envolvimento que habita ou toca parcialmente, mas não abrange a outra.

A experiência aqui transposta reúne ainda mais destes vários ‘universos’ em contato. Afinal, além da minha atuação como pesquisador, este relato apresenta detalhes de um evento científico do qual participei apresentando um trabalho, outro interesse específico, e encontrando inúmeros outros pesquisadores com os mais diversos interesses para estarem em Bremen. Sendo assim, é importante ressaltar que não estive na *Data Power Conference* realizando um trabalho etnográfico no sentido formal. Minha experiência como participante da conferência é aqui rememorada não tanto através de conversas gravadas, entrevistas ou visitas previamente direcionadas, mas da memória de conteúdos de palestras e impressões sobre a reação de ouvintes, de conversas de corredores e confissões feitas por participantes do evento em diálogos informais, da maneira como fui recebido e ‘inserido’ no meio dos participantes da conferência, bem como dos diversos eventos sociais externos ao evento.

Algumas das impressões foram anotadas em um caderno de notas que utilizei durante todo o evento, especialmente no que se refere às sessões de trabalho e palestras. No mais, meu maior aliado é a memória das experiências vividas ali. Algumas fotos tiradas durante o evento serão apresentadas a seguir, assim como registros de pequenas conversas informais que marcaram meus dias em Bremen e ajudam a entender melhor o cenário ainda pouco explorado - em um sentido acadêmico, ao menos - da cadeia de relações que alimenta o funcionamento da BD&S e deste considerável espaço de debates sobre Big Data nas Ciências Sociais. Além disso, a relevância do relato e das informações etnográficas se dá em apresentar uma análise do objeto que extrapola - e acaba por informar - o que é encontrado pela via quantitativa (e.g. SILVA, 2008).

2. Quem está olhando? Mapeamento analítico e as comunidades que compõem a Big Data & Society

2.1 Apresentação

O relacionamento entre Big Data e Ciências Sociais têm começado a ganhar espaço na literatura científica recente. Nos últimos dois anos, por exemplo, duas obras foram publicadas tendo como temática o “uso de Big Data nas várias disciplinas das Ciências Sociais” (CHEN; YUN, 2018, p. 3). Como os próprios nomes dos trabalhos indicam³⁰, a perspectiva é a do uso de metodologias quantitativas baseadas em extração e análise de dados para as pesquisas empíricas por parte de cientistas sociais. O foco aí é em fornecer um apanhado geral dos principais métodos e, a partir daí, traçar um passo-a-passo para uma ciência social quantificável (FAZEKAS, 2014; FOSTER et al., 2017).

Todavia, o que fazer de um campo que sempre teve por objeto de análise algo tão complexo e irreduzível como este que se convencionou chamar de humano e um *habitat* tão subjetivo quanto fluído como o social? Que dizer de aspectos qualitativos, epistemológicos, filosóficos, políticos? Tais questionamentos e outros mais refletem o que se constituiu historicamente como um certo espírito crítico das Ciências Sociais que afeta o modo como este campo usualmente tem se posicionado ante os diversos fenômenos da sociedade, inclusive um tão recente quanto o Big Data (DALTON; THATCHER, 2014). Para Dalton e Thatcher (2014) é esta a grande contribuição que as Ciências Humanas e Sociais podem trazer para os estudos em Big Data, um engajamento crítico para com os dados, para além de sua mera aplicação funcional (SILVA, 2018).

Diante disso, constitui-se o objetivo deste levantamento entender quais têm sido as contribuições das Ciências Sociais para o estudo em Big Data. É neste campo de circulações heterogêneas que esta dissertação busca transitar. Para isso, o trabalho se decidiu pelo recorte de uma revista científica, ‘fechando as portas’ para outros territórios de circulação do Big Data como a produção bibliográfica (livros)³¹, veículos de imprensa³² e

³⁰ “Big data and social science: a practical guide to methods and tools”, de Foster et al. (2017); e “Big Data in Computational Social Science and Humanities”, de Chen (2018).

³¹ Como os já citados anteriormente, além de outros que constam na bibliografia desta dissertação.

³² Como, por exemplo, o famoso artigo de Chris Anderson, publicado na revista Wired em 2008, cujo polêmico título proclamava o fim da teoria com a chegada do Big Data. “The End of Theory”. Disponível em: <<https://www.wired.com/2008/06/pb-theory/>> Acesso em: 20/02/2020. Outras amostras interessantes que

blogs na web³³, devido ao ensejo de compreender quais são e como se formam as tendências, fundamentos e lugares de fala que caracterizam as tratativas *acadêmicas* sobre Big Data nas Ciências Sociais.

O recorte a partir do qual esta pesquisa se constrói é o da BD&S. A seguir, exponho uma breve introdução com informações gerais da revista, seguida de uma reflexão sobre as comunidades que constituem as suas bases e a importância disto para o entendimento do local em que a revista se propõe a ocupar. Dados do levantamento realizado na BD&S serão apresentados à medida em que as discussões avançam. Por fim, este capítulo trará também um relato de experiência de minha participação em um dos eventos ligados à BD&S e uma discussão sobre o funcionamento da BD&S e como este ‘fecha o cerco’ em torno da proposição e do modo de abordagem de certas questões. Isso se revela também nos textos principais que informam essa comunidade, tema do capítulo 3. Se espera argumentar ainda que ao mesmo tempo em que a BD&S se coloca em posição crítica em relação aos projetos tradicionais de Big Data e suas consequências sócio-políticas, especialmente para contextos marginais globais, como ilustra o evento aqui descrito, o referencial de discussão, o funcionamento desta comunidade científica e sua base epistemológica continuam sendo a do contexto geopolítico do Norte global. Esta tensão será discutida de modo mais aprofundado no quarto capítulo.

2.2 BD&S: informações gerais

A revista BD&S se constituiu como um dos principais canais de divulgação, interação, desenvolvimento e consolidação dos estudos sobre Big Data nas Ciências Sociais. Ela teve sua primeira edição publicada em abril de 2014 com a proposta de ser uma

apresentam a amplitude do tema nos noticiários são as reportagens de Dave Ryan, publicada em 2015 no Huffington Post, sobre a interface entre Big Data e esportes: “Sports: Where Big Data Finally Makes Sense”. Disponível em: <https://www.huffingtonpost.com/dave-ryan/sports-where-big-data-fin_b_8553884.html> Acesso em: 20/02/2020; e de Jennifer Schuessler, lançada em 2017, no New York Times, sobre o encontro entre Big Data e literatura: “Reading by the Numbers: When Big Data Meets Literature”. Disponível em: <<https://www.nytimes.com/2017/10/30/arts/franco-moretti-stanford-literary-lab-big-data.html>> Acesso em: 20/02/2020.

³³ Textos recorrentes nesta dissertação, por exemplo, foram frutos de discussões em blogs, como o intrigante “Big Data is our generation’s civil rights issue, and we don’t know it”, publicado em 2012 por Alistair Croll. Disponível em: <<http://solveforinteresting.com/big-data-is-our-generations-civil-rights-issue-and-we-dont-know-it/>> Acesso em: 20/02/2020.

publicação pioneira na exploração do *Big Data* a partir das contribuições das Ciências Sociais³⁴. Seu objetivo era “analisar as práticas de Big Data, ou envolver propostas empíricas [...] que também reflitam nas consequências de como as sociedades [com/através do *Big Data*] são representadas [epistemologias], compreendidas [ontologias] e governadas [políticas]”³⁵, como aponta o editorial. O tom de ênfase crítica da BD&S tem relação com o círculo maior de pesquisa do Big Data apontado na seção 1.3. Como visto, as discussões sobre Big Data englobam os mais diversos campos, incluindo aí áreas tão distintas quanto Medicina e Engenharia, por exemplo. Enquanto a pesquisa sobre Big Data mais publicada e popular nos meios acadêmicos e não-acadêmicos é a que parte das pesquisas comerciais, envolvendo soluções e práticas mais eficientes de Big Data para o mercado, a BD&S, que se pode chamar, nesse contexto, de uma revista periférica ante as discussões mais comuns do Big Data, construiu suas bases fazendo frente a essa realidade.

Esse ‘choque’, que ajuda a entender a própria formação dos campos que se reúnem na BD&S, é perceptível, por exemplo, na própria formação acadêmica dos que publicam na revista e nas instituições mais importantes dentro destas comunidades. De um lado a pesquisa em Big Data nos campos da Engenharia, Computação e Administração tende a se concentrar em universidades de maior porte financeiro e maior capacidade de captação de recursos a partir de empresas *high-tec* e órgãos governamentais, como MIT e Harvard, com maior preocupação em desenvolvimentos técnicos para o aperfeiçoamento dos sistemas

³⁴ É importante, neste momento, ressaltar a existência de outras publicações científicas que se propõem, como mote principal, a investigar o *Big Data*. Em levantamento breve, constatei a presença de doze revistas acadêmicas, incluindo *Big Data & Society*, a lidarem, como proposta editorial, com o Big Data. São elas [entre colchetes, seus principais enfoques]: *IEEE Transactions on Big Data* [Engenharia e Ciência da Computação], *Big Data* [Pesquisa e inovações em big data - interdisciplinar], *Journal of Big Data* [Ciência da Computação e Computação cognitiva], *Big Data Research* [Interdisciplinar - com destaque para publicações na área de Ciência da Computação], *Big Data & Cognitive Computing* [Computação cognitiva], *Frontiers in Big Data* [Data mining e cybersecurity], *International Journal of Big Data Intelligence* [Vigilância e gerenciamento de big data], *Big Data Analytics* [Computação cognitiva], *International Journal of Big Data and Analytics in Healthcare* [Pesquisa em Saúde], *Open Journal of Big Data* [Progresso e inovação no/com o uso dos big data - interdisciplinar], e *Big Data Analytics for Healthcare* [Pesquisa em Saúde]. Várias outras publicações que tratam de temáticas relacionadas e/ou similares, como o *International Journal of Data Science and Analytics*, *Computational Statistics & Data Analysis (CSDA)*, *Data Science Journal*, e *Journal of Digital Humanities*, são citadas aqui, porém permanecem fora do escopo do trabalho. A escolha pela análise da *Big Data & Society* se deve pela singularidade de sua proposta editorial, ao menos explícita, voltada às Ciências Sociais, que é o lócus onde a pesquisa busca situar-se.

³⁵ About the Journal. *Big Data & Society*. Disponível em: <<http://bigdatasoc.blogspot.com.br/p/big-data-and-society.html>> Acesso em: 20/02/2020.

de dados³⁶. Do outro, a BD&S emerge ao lado de comunidades de cientistas sociais que pesquisam dados, ainda que de maneira periférica aos desenvolvimentos mais ‘tradicionais’, congregando centros de pesquisa menos tradicionais como as universidades de Carleton, Cardiff e Sheffield, além de pesquisadores e textos que também caminham nas ‘margens, oferecendo uma visão crítica dos caminhos comuns da pesquisa em dados³⁷. Isso é perceptível também pela presença de nomes como Rob Kitchin, Lev Manovich, David Lyon, Bruno Latour, Tarleton Gillespie e Luciano Floridi, conhecidos em círculos das Ciências Sociais e Humanidades Digitais por fornecerem uma visão crítica em relação a diferentes aspectos do desenvolvimento tecnológico, digital e da datificação na contemporaneidade.

Até novembro de 2019, 266 textos, entre artigos, comentários e editoriais, foram publicados pela BD&S, espalhados por dez edições, com frequência de publicação semestral. Para efeito de delimitação do corpus de análise, se optou por mapear um total de 131 artigos, publicados nas seis primeiras edições, entre os anos de 2014 e 2016. Tal escolha se deu por meio de análise prévia do material, que constatou ser este um período de maturação das discussões sobre Big Data na revista, devido à presença de abordagens epistemológicas/metodológicas, nas primeiras edições, e à publicação dos primeiros dossiês temáticos, que permitiram a identificação de múltiplas perspectivas - estudos em vigilância, meio-ambiente, arte urbana etc. - sobre a questão do Big Data nas Ciências Sociais.

Outro fator que corroborou para a delimitação deste escopo de análise se deu pelo tipo de trabalho que envolvia a análise, já que todos os dados levantados no estudo partiram de classificação e análise manual, sendo a quantidade total de artigos um número elevado para um projeto de dissertação de Mestrado. Atualmente, a revista tem se consolidado como referência na interface Big Data-Ciências Sociais, especialmente para perspectivas críticas sobre os dados, publicando temas especiais a partir de conferências internacionais apoiadas pelo seu Comitê Editorial.

O portal da revista informa ser esta indexada nas bases de dados da Analytics: Social Sciences Citation Index (SSCI), Directory of Open Access Journals (DOAJ), Google Scholar e Scopus, apesar desta última não fornecer nenhum registro da BD&S em seu sistema de buscas. Não obstante o reduzido número de bases de indexação, parte do renome da revista

³⁶ Os temas da quinta edição da “Big Data Conference”, realizada em Harvard, ilustram essa preocupação. Disponível em: <<https://cmsa.fas.harvard.edu/2019-big-data/>> Acesso em: 20/02/2020.

³⁷ Ver nota 58.

em comparação com outras publicações que tomam o Big Data por objeto de estudo se deve ao fato de ser parte da rede de publicações SAGE, uma das gigantes do mercado de publicações acadêmicas ao lado de Elsevier, do grupo Springer Nature, entre outros. Ter o nome vinculado às publicações destes grupos, portanto, traria um senso de maior reconhecimento entre os pares e de maior prestígio na ‘casta’ acadêmica (BURRIS, 2004).

Além de ser uma publicação da rede SAGE, a BD&S é atrelada primariamente a pesquisadores ligados a importantes grupos de pesquisa da Europa, Canadá e Estados Unidos que abordam estudos críticos sobre dados, vigilância e sociedade. Entre eles destacam-se o *Data Justice Lab*, abrigado na Universidade de Cardiff, no Reino Unido; o *Institute for Data Science*, da Universidade de Carleton, no Canadá; e o *Data & Society Research Institute*, sediado em Nova Iorque, nos Estados Unidos, entre outros. O impacto das agendas de pesquisa e do funcionamento destes grupos na BD&S é o foco deste capítulo, porém, inicialmente, é preciso entender de que modo se dá o funcionamento deste universo acadêmico e em que tecido social se dá a publicação de uma revista científica e sua circulação.

2.3 O meio social da revista científica

Falar em algo como tecido social da academia³⁸ ou ‘funcionamento social’ do meio acadêmico – ou mesmo o que se convencionou chamar de sociologia da ciência - de pronto chama a atenção para uma questão fundamental: o desenrolar da produção científica e da sua circulação tanto na sociedade de modo geral, como entre os seus próprios pares, não se dá, como se imaginava numa visão idílica da ciência, na fagulha de genialidade e na “geração autônoma de ideias” (CAPPELL; GUTERBOCK, 1992, p. 266) por parte dos heróis acadêmicos. Essa ideia foi predominante nas reflexões sobre ciência até, pelo menos, a década de 1970. Parecia aí que o fazer científico se dava de um modo quase cíclico, orientado por inúmeras teses e antíteses bem como pelo ciclo quase ‘natural’ das mudanças de paradigmas³⁹.

³⁸ O termo, recorrente neste trabalho, é usado aqui no sentido tradicional evocado desde Platão, associado ao lócus de desenvolvimento da sabedoria. Aqui é utilizado em um sentido mais formal para designar o lócus de acumulação e circulação do conhecimento produzido nas mais diversas instituições científicas.

³⁹ Aqui uma referência clara a influente obra de Thomas Kuhn (1962), “A Estrutura das Revoluções Científicas”. Outras críticas similares também foram feitas a uma certa ‘naturalização’ do fazer científico que

Todavia, boa parte dos esforços da sociologia da ciência a partir da década de 1970 consistiu em desafiar essa ‘naturalização’ do funcionamento científico e apontar para a sua existência dentro de uma rede concebida como afetando e sendo afetada por pesquisadores e suas pesquisas, desde a própria concepção do que é relevante como tema de pesquisa até os valores e premiações que são atrelados ao que pode ser defendido academicamente falando. Como aponta Paisley (1972, p. 5), “é este sistema [...] que concede prêmios Nobeis, enfatiza as prioridades de descoberta, estabelece grandes fundações privadas e mantém as universidades”⁴⁰.

É neste senso que Pierre Bourdieu aponta para o fazer científico como um campo, nem completamente amarrado ao meio sociopolítico onde se situam as universidades e seus acadêmicos, nem um espaço cujo funcionamento se dá de maneira autômata, repleto de indivíduos com pouco nexos entre si apenas seguindo o chamado maior da ciência rumo ao progresso, restando apenas o olhar sobre os ‘produtos’ da ciência em si, os achados científicos.

Em outras palavras, é preciso escapar à alternativa da “ciência pura”, totalmente livre de qualquer necessidade social, e da “ciência escrava”, sujeita a todas as demandas político-econômicas. O campo científico é um mundo social [...]. Isso significa que só compreendemos, verdadeiramente, o que diz ou faz um agente engajado num campo se estamos em condições de nos referirmos à posição que ele ocupa nesse campo, se sabemos “de onde ele fala” (BOURDIEU, 2004, p. 21, 23)

No contexto do reconhecimento da interpelação do pesquisador e sua pesquisa no emaranhado de relações que informam a maneira como este se relaciona com o conhecimento que produz, um dos esforços da sociologia da ciência se voltou para tentar compreender o funcionamento do que se convencionou chamar de colégios invisíveis. Esta expressão tem sido definida como a interrelação entre a especialidade do pesquisador, o papel que ele ocupa como ator social no contexto acadêmico e o valor que sua contribuição possui

Kuhn parece seguir. Já leituras distintas veem Kuhn como um dos primeiros anti-naturalistas em sua leitura sobre a ciência (cf. BANKS, 1983; LALUMIA, 1991).

⁴⁰ Cabe a ressalva aqui, apesar do peso do parágrafo seguinte, ao fato de que a visão de Paisley aqui é a de um sistema científico que, apesar de afetado constantemente por políticas de Estado, consegue absorver tais políticas de formas refratadas, ora ampliando a força do status quo, ora fornecendo espaços de questionamento ao sistema estatal. O importante aqui é notar, mais uma vez, a relativa autonomia deste sistema social científico.

dentro do ambiente em que esta circula (ZUCCALA, 2004), uma comunidade informal de pesquisadores de diferentes instituições de ensino ou mesmo de países distintos (PRICE, 1976), bem como um contexto social que envolve cientistas trabalhando em questões similares e que os afeta seja por comunicação direta e pessoal, seja pela influência indireta de referências-chave da área (BURT; DOREIAN, 1982).

Os colégios invisíveis funcionam, portanto, como sistemas voluntários ou involuntários de afiliação, formais ou informais, que atuam como referência modulando⁴¹ a atividade de todo o ecossistema científico, desde a liderança de comitês editoriais de revistas científicas, passando por projetos de pesquisa e seus financiamentos, sistema de motivação e recompensa que estrutura os objetivos do cientista, sistemas de patentes e disputas comerciais, até alcançar a formação dos cânones de determinada área e do que deve ser considerado relevante (PAISLEY, 1972). Enquanto o arranjo dos colégios invisíveis e suas comunidades científicas particulares é visto por alguns como aquilo que move o progresso do conhecimento e da pesquisa criativa (e.g. WAGNER, 2008), é apreendido por outros como tendo um caráter clubístico, voltado para o favorecimento dos membros que se conformam aos conceitos e prioridades vigentes no grupo (PAISLEY, 1972).

Em resumo, no estudo do funcionamento destas comunidades está em jogo o reconhecimento de que

cada escolha científica, a escolha de uma área de pesquisa, de metodologias, de revista para se publicar um artigo, é um investimento político estratégico, direcionado objetivamente, pelo menos, à maximização do lucro estritamente científico, isto é, do potencial de reconhecimento do pesquisador ante outros colegas-competidores (BOURDIEU, 1975, p. 22).

Mas, afinal, de que forma tais comunidades se fortalecem? Que indivíduos ou forças a impelem a se recrudescer? As respostas aqui são multifacetadas. Pierre Bourdieu (2004) usa a metáfora do jogo como uma ilustração do funcionamento do campo científico, onde enquanto os jogadores contribuem para transformações na própria forma como o jogo é disputado, eles mesmos são subordinados às regras e princípios que regem o certame. Uma

⁴¹ Destaco aqui o uso desta palavra ao invés de controlar ou manipular. Os colégios invisíveis atuam, a meu ver, no sentido de moldar, dar forma, ao ecossistema científico. Toda forma aqui entendida como provisória e em constante mudança.

das características que são fundamentais para os jogadores, leia-se pesquisadores, adentrarem a partida e, desta forma, para que esta se consolide, é a capacidade de “antecipar as tendências”, isto é, saber transitar no meio, reconhecer quem já está no jogo há muito tempo e de dominar “as leis imanentes do campo, leis não escritas que são inscritas na realidade em estado de tendências e de ter o que se chama em rugby [...] de o sentido do jogo” (BOURDIEU, 2004, p. 27).

O fortalecimento dos colégios invisíveis se dá, portanto, por meio da retroalimentação de seus sistemas através de um conjunto de ‘regras invisíveis’ e de nomes que ajudam a modular todo o sistema, dando sua consistência. Esse fenômeno que chamo de canonização da ciência⁴², de transmutação de autores e textos, ou mesmo de ‘jogadores com mais experiência e títulos conquistados naquela determinada equipe’, em ‘pontos de passagem obrigatório’, e que é fundamental para dar forma a uma dada comunidade científica, é perceptível, por exemplo, em estudos cientométricos. Como exemplo, dados do norte-americano National Institutes of Health (NIH), na virada do milênio, revelaram que um conjunto de 33 cientistas da área de Biomedicina de grande produtividade constituíam uma rede centralizada que contabilizava até 83% de todos os projetos financiados em centros de pesquisas estadunidenses, além de receberem o maior número de citações em artigos científicos da área (ZUCCALA, 2004).

A co-citação repetida do trabalho de outros autores é um dos mecanismos através dos quais pesquisadores criam e sustentam laços que facilitam a rejeição de ‘informação irrelevante’. Tais laços constituem colégios invisíveis (ZUCCALA, 2004, p. 32).

Apesar das motivações para as citações serem inúmeras, não são poucas as contribuições que apontam para as citações aos cânones como uma forma de reconhecimento social, de adentramento naquele habitat específico ocupado por tal autor. De um domínio, afinal, das regras que sustentam aquela comunidade científica e que parecem “administrar a ciência para nós” (GUIMARÃES; HAYASHI, 2014, p. 218)

⁴² Isto é, o processo que ocorre dentro do seio de uma ou da relação entre uma ou mais áreas do conhecimento em torno da consolidação de certas obras e autores como pontos de passagem obrigatórios para a própria compreensão e elaboração da área. Como afirmo na nota 8, este processo deve ser visto também como o da própria consolidação da área.

2.4 BD&S: colégios invisíveis e cânones sob um olhar quantitativo

Tal não seria diferente numa publicação científica como a BD&S, que, como aponta Vanderstraeten (2010, p. 561), funciona como um instrumento que permite a organização das disciplinas científicas e referenda aquilo que tais disciplinas tomam por conhecimento ‘certificado’. Funcionando como articuladora dos colégios invisíveis que tomam por preocupação comum os estudos em Big Data a partir de perspectivas críticas, em especial de cientistas sociais, é de se esperar também que a BD&S apresente também seus pontos de passagem por onde os princípios deste ‘jogo’ devem circular.

Figura 2: Nuvem de palavras com o sobrenome dos autores mais citados nos artigos da revista BD&S publicados entre 2014 e 2016.

O mapeamento de caráter bibliométrico feito na BD&S procurou, no processo de contribuir para a proposta de compreender o funcionamento dos colégios invisíveis que constituem a pesquisa na BD&S trazer informações bibliométricas que contribuíssem para um quadro mais amplo. Apesar de reconhecer os limites de tal levantamento na compreensão das motivações internas dos pesquisadores bem como no quadro maior das teias sociais que formam dada área científica, como outros o fizeram (e.g. EDGE, 1979; LIEVROUW, 1990), creio que os resultados deste, somados às informações adquiridas no campo, constituem a melhor possibilidade para a constituição de um ‘mapa de navegação’ (ZUCCALA, 2004) da pesquisa em Big Data entre cientistas sociais feita na BD&S.

A Figura 2 apresenta algum dos nomes dos autores mais citados na publicação durante o período de 2014 e 2016. Mais do que um relato bibliométrico, tal ilustração aponta para aqueles que funcionam como referências dentro dos colégios invisíveis que são interpelados na BD&S (BURT; DOREIAN, 1982). Nomes como Rob Kitchin, autor mais referenciado ao longo dos artigos da BD&S (57 citações), professor na National University of Ireland Maynooth (Irlanda) e importante nome nos estudos em Geografia, Ciências Sociais e dados nos últimos anos, Kate Crawford e danah boyd, principais pesquisadoras da Microsoft Research (com 45 aparições na revista), seguidas do sociólogo francês Bruno Latour (37 vezes), dos sociólogos britânicos Roger Burrows (27 vezes) e Mike Savage (25 vezes), ao lado do sociólogo norte-americano Geoffrey C. Bowker (25 vezes) e do sociólogo australiano Adrian Mackenzie (25 vezes) se apresentam como exercendo uma posição de proeminência e centralidade dentro das comunidades científicas que se encontram na BD&S.

É interessante notar que todos estes nomes, à exceção de Kate Crawford, aparecem, inclusive, entre os membros dos Comitês Editorial e Consultivo da BD&S⁴³. A interação entre estes autores é ainda mais perceptível quando vista sob o espectro da filiação teórica, como mostra a Figura 3.

⁴³ Informações sobre estes comitês se encontram no site da BD&S. Disponível em: <<https://journals.sagepub.com/editorial-board/BDS>> Acesso em: 20/02/2020.

Vínculo de correlação entre principais autores

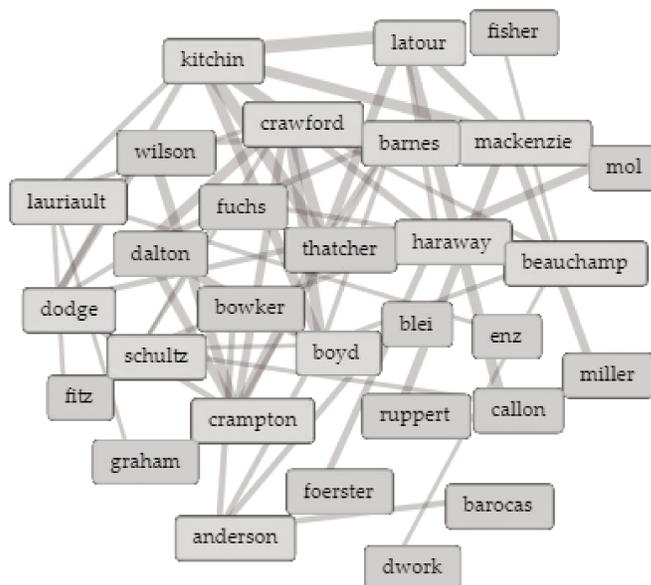


Figura 3: Ligações entre o uso de autores dentro de um mesmo artigo. Linhas mais grossas apresentam autores que aparecem em conjunto dentro de um mesmo artigo com maior frequência, enquanto o inverso, ou mesmo a inexistência de correlação, acontece à medida que a linha é mais fina.

Nota-se, a partir da Figura 3, que o vínculo de correlação⁴⁴ entre os autores mais citados, que se encontram na faixa central-superior do gráfico, é elevado, isto é, não somente aparecem com frequência nos artigos, mas em geral aparecem juntos nos textos da BD&S. Rob Kitchin, por exemplo, aparece associado com frequência elevada exatamente a outros teóricos que fazem parte do mesmo grupo consultivo e editorial da publicação, como Bruno Latour, Kate Crawford, danah boyd, Jim Thatcher e Adrian Mackenzie, e o mesmo vale para os trabalhos destes autores.

Tal fato auxilia na compreensão dos sistemas de reputação da BD&S. Como apontam Burt e Doreian (1982, p. 110), “a reputação de uma revista científica é determinada pelo interesse que uma comunidade científica atribui a ela”. Se é nela que se reúnem os ‘pontos de passagem obrigatórios’, os que informam os principais conceitos e debates

⁴⁴ Calculado pelo fator de relação entre autores dentro de um mesmo texto. Ligações mais intensas indicam que dois ou mais autores são citados quase sempre juntos dentro de um mesmo artigo. Metodologicamente, a figura segue o que se convencionou chamar de *collocation graphs*, um tipo de análise textual que investiga o espaço que determinados termos compartilham no texto e, a partir daí, o nível de associação e correlação entre estes termos. Mais detalhes sobre as expressões estatísticas que sustentam os *collocation graphs* podem ser encontradas no texto do criador do conceito, Martin Phillips (1985), e no artigo de Vaclav Bezina (2018).

realizados em um campo, o ingresso em determinado campo científico passa, informalmente, pela leitura e publicação dos pesquisadores nos mesmos canais. De fato, é esperado do pesquisador que deseje ‘entrar’ no mundo da pesquisa em dada área que interaja com os cânones da disciplina e parte sempre da vereda trilhada por estes. Não poucos reconhecem que é este processo que garante a uma publicação científica, apesar do processo de avaliação por pares às cegas (*blind-review*), manter uma função estruturante para os debates da área (VANDERSTRAETEN, 2010; HAYASHI et al, 2016).

Em se tratando dos cânones – um dos elementos centrais na modulação de uma comunidade científica - que informaram os debates na BD&S no período estudado, a Tabela 1 aponta para os dados fornecidos pelo mapeamento acerca dos textos mais citados no escopo dos 131 artigos estudados. A obra mais citada na BD&S no período analisado é o texto “Questões Críticas para o Big Data: provocações para um fenômeno cultural, tecnológico e acadêmico”, um artigo publicado em 2012 de autoria de danah boyd e Kate Crawford. Com quase o mesmo número de aparições está o texto “Big Data, novas epistemologias e mudanças de paradigma”, publicado em 2014 na própria BD&S por Rob Kitchin. Tais textos serão analisados com maior profundidade no capítulo 3, tanto em relação ao seu conteúdo quanto ao modo em que são abordados nos artigos da BD&S.

OBRAS MAIS CITADAS	FREQUÊNCIA EM ARTIGOS
(Boyd & Crawford, 2012)	29
(Kitchin, 2014)	27
(Mayer-Schönberger & Cukier, 2013)	13
(Savage & Burrows, 2007)	10
(Dalton & Thatcher, 2014)	10
(Gitelman, 2013)	10
(Anderson, 2008)	8
(Bowker & Star, 1999)	8
(Edwards, 2010)	7
(Gillespie, 2014)	7
(Amoore, 2011)	6
(Latour, 2005)	6
(Lazer et al., 2014)	6

(Ruppert et al., 2013)	6
(Blei, 2012)	5
(Mackenzie, 2015)	5
(Taylor et al., 2014)	5
(Thatcher, 2014)	5
(Wilson, 2015)	5

Tabela 1: Obras mais citadas nos artigos da revista entre 2014 e 2016⁴⁵.

Entre estas figuras-chave, o período inicial da publicação (2014) é marcado pela presença de Roger Burrows como principal referência nos artigos, como mostra a Figura 4. Seu artigo escrito juntamente com Mike Savage “A crise iminente da sociologia empírica”, um convite para que a sociologia considere os desafios que os métodos de análise de dados trazem para o campo, a ser discutido mais detalhadamente no terceiro capítulo, exerce influência direta sobre a maior parte das discussões iniciais, enquanto o interesse nas discussões posteriores se volta mais para textos de outros nomes como boyd, Crawford, Kitchin e Latour. Interessa no escopo desta pesquisa o processo de sedimentação de certas referências em detrimento de outras.

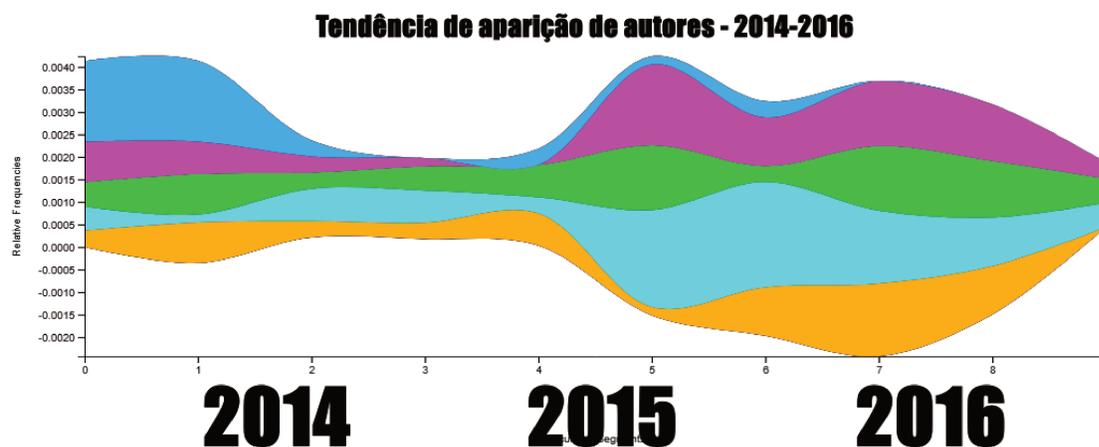


Figura 4: Tendência de aparição dos autores mais referenciados no decorrer do período 2014-2016. De cima para baixo: Roger Burrows (azul); Kate Crawford (rosa); danah boyd (verde); Rob Kitchin (azul claro); Bruno Latour (laranja).

É interessante notar que, ao contrário de outros colégios invisíveis de estudos mais consolidados, articulados por publicações científicas muitas vezes centenárias (e.g.

⁴⁵ Referências completas dos artigos se encontram na lista de referências bibliográficas ao final desta dissertação.

HAYASHI et al, 2016), a organização dos cânones e das lideranças dentro das comunidades nos estudos em Big Data nas Ciências Sociais acompanham o recente surgimento da BD&S, como aponta a Tabela 2. Isso nos permite presumir, e confirmar o que parte da literatura apresentada neste capítulo apresenta (e.g. VANDERSTRAETEN, 2010; D’ANDREA; DELICH, 2005), que *o processo de cristalização de uma comunidade científica e o de organização de uma publicação científica estão intrinsecamente ligados e não podem ser desassociados*. É possível notar que o campo de estudos em Big Data, especialmente nas Ciências Sociais, se encontra em sua infância, já que a maior parte dos materiais mais citados na BD&S é constituída de artigos ou livros inéditos publicados na década de 2010. O número de referências a obras publicadas entre 2010 e 2016 chega quase ao dobro do que o total de todos os outros períodos somados. Muitas referências, inclusive, são de artigos publicados na própria BD&S entre 2014 e 2016.

ANO DAS PUBLICAÇÕES MAIS CITADAS	NÚMERO DE REFERÊNCIAS	ANO DAS PUBLICAÇÕES MAIS CITADAS	NÚMERO DE REFERÊNCIAS
2016	72	1995	23
2015	154	1994	15
2014	365	1993	7
2013	300	1992	15
2012	201	1991	15
2011	162	1990	9
2010	115	1989	5
2009	81	1988	5
2008	70	1987	16
2007	61	1986	14
2006	44	1985	8
2005	52	1984	11
2004	33	1983	7

2003	26	1982	5
2002	28	1981	3
2001	27	1980	3
2000	29	1979	9
1999	24	1978	6
1998	21	1977	4
1997	10	1976	1
1996	7		

Tabela 2: Anos de publicação dos materiais referenciados nos artigos da BD&S e quantidade de referências a materiais publicados naquele ano.

Todo este cenário impacta diretamente os principais interesses da pesquisa publicada na BD&S. Como mostra a Figura 5, alguns dos principais termos utilizados são estruturados em diálogo com alguns dos temas-chaves dos textos canônicos apresentados anteriormente. Termos como “social” (52 vezes), “analysis” (26 vezes), “science” (24 vezes), “media” (23 vezes) e “digital” (20 vezes) aparecem entre os mais utilizados [Para lista completa de palavras-chave, ver Anexo].

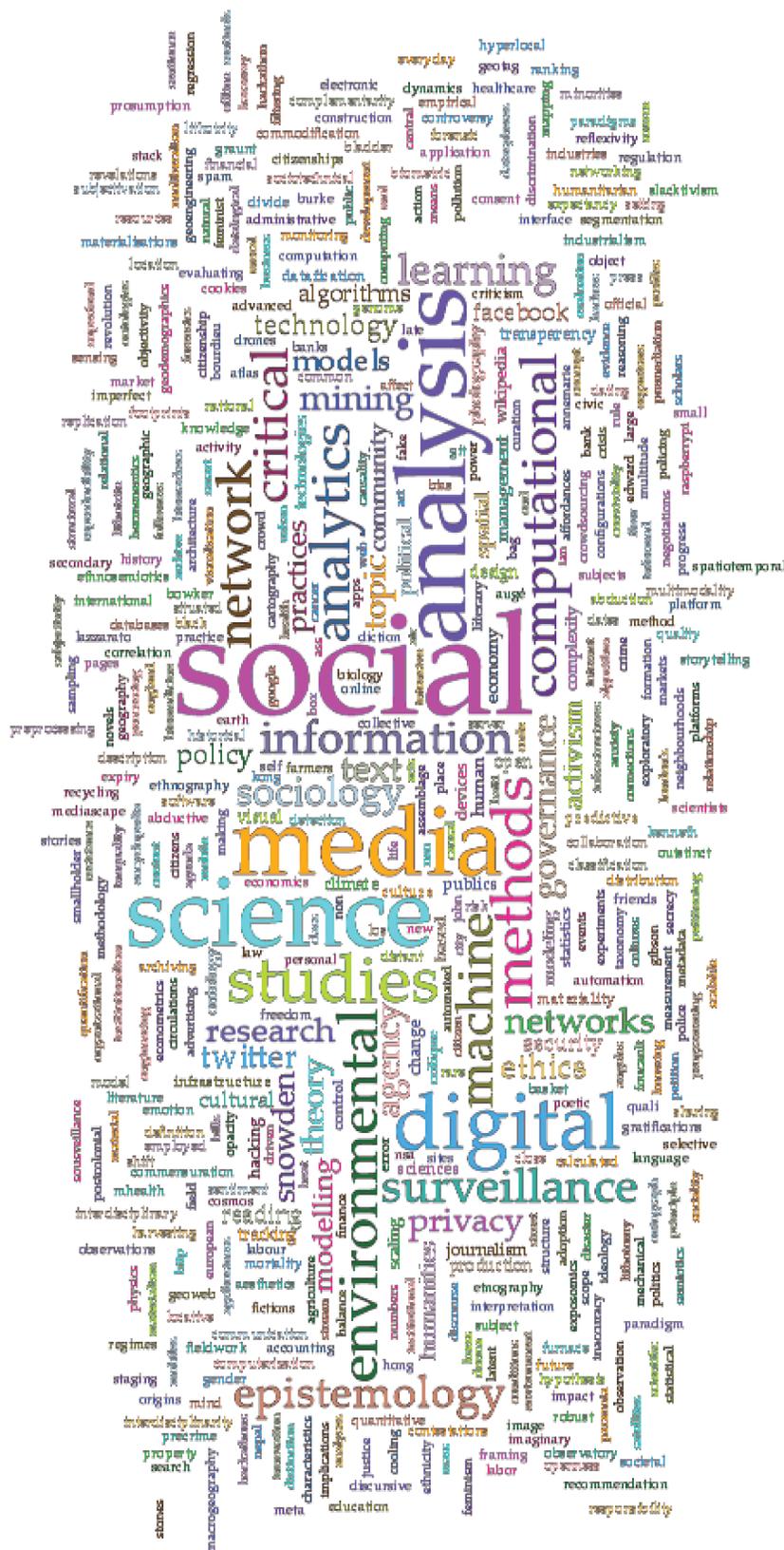


Figura 5: Nuvem de palavras com os termos mais utilizados nas palavras-chave dos artigos publicados entre

2014 e 2016 na BD&S, com a exclusão dos termos “Big” e Data”, devido a frequência muito superior com que aparecem nos artigos, somando um total de 161 vezes.

Apesar do caráter pouco específico de muitas destas palavras - por exemplo, social pode se referir a inúmeras coisas - este levantamento permitiu identificar a presença considerável de temáticas como “mídias digitais”, “redes sociais”, “meio-ambiente” (11 vezes), “métodos computacionais” (10 vezes), “vigilância” (9 vezes) e “epistemologia” (8 vezes) entre os artigos. Tal achado permite inferir a formação de verdadeiros ‘campos’ de pesquisa em Big Data dentro das Ciências Sociais, como, por exemplo, os estudos em vigilância, que englobam termos como “privacidade” (6 vezes), “governança” (6 vezes), “Snowden” (5 vezes), “ativismo” (5 vezes) e “segurança” (4 vezes), entre outros; meio-ambiente, envolvendo questões como “política de modelagem” [5 vezes] e “mudança climática” [3 vezes]; estudos em espaço e geografia (expressões como “Big Data Espacial” [4 vezes] e “rastreamento” [3 vezes]); análises de redes sociais (expressões como “mídias digitais” [20 vezes], “Twitter” [6 vezes] e “Facebook” [4 vezes]); bem como métodos computacionais (expressões como “machine learning” [8 vezes] e “mineração de textos” [6 vezes]).

A variedade maior de temas, todavia, não é vista como proporcional a uma maior variedade de cânones e autores-líderes nas comunidades da BD&S, tendo em vista a disparidade mais profunda existente entre os autores e textos mais citados e o restante do material. Um questionamento que pode ser levantado, então, é se não seria próprio atribuir a tais textos e autores a posição de pontos unificadores - ou *brokers*⁴⁶, na linguagem da Análise Computacional de Redes - das pesquisas em Big Data na BD&S.

Outra questão essencial para a compreensão da formação de colégios invisíveis se encontra nos territórios que arregimenta. Isto é, quais instituições, de que lugares, têm os assentos mais importantes dentro da comunidade científica? Que idiomas esses colégios falam?

⁴⁶ Termo derivado do francês ‘*broceur*’, que significa pequeno comerciante. No contexto atual, representa alguém ou algo que se coloca como intermediário entre duas partes no contexto de um acordo comercial. A Física Social usa desta definição para apontar o broker como figura mediadora obrigatória para o vínculo entre distintos sistemas sociais e mesmo para a coesão geral das redes de conexão social. Nesse caso, quanto maior o isolamento entre dois clusters distintos, maior o poder que um broker terá. Sobre isso, ver o célebre trabalho de Ronald S. Burt (1995).

O predomínio anglófono, no que se refere ao idioma utilizado nas comunidades científicas, e especialmente dos Estados Unidos e do Reino Unido no que tange ao número de publicações científicas abrigadas ali, instituições e acadêmicos que publicam e organização (e verba para isso) de congressos científicos é fato reconhecido amplamente dentro da academia (e.g. BELCHER, 2007; GUIMARÃES; HAYASHI, 2014). Alguns estudos, inclusive, se dedicam a apontar, a partir de relatos etnográficos, as dificuldades de pesquisadores das ‘margens’, geograficamente e linguisticamente falando, em conseguir, de fato, obter um senso de ‘pertencimento’ a tais comunidades⁴⁷. Isso envolve inclusive dificuldades relativas a tentativas de se publicar em inglês por parte de acadêmicos naturais de línguas não-anglófonas – e sem o aporte de instituições de fomento, universidades ou de famílias afortunadas para a realização de períodos de estudo nos EUA ou no Reino Unido – que vão desde críticas recebidas nas avaliações por questões gramaticais ou sintáticas até rejeições por falta de adequação aos termos e estilo argumentativo convencionais da comunidade científica (DUEÑAS, 2012).

A Figura 6 aponta para um mapa geográfico dos países envolvidos nas publicações da BD&S. Houve uma tentativa de classificar também a origem nacional de cada autor, todavia como nem sempre foi possível obter estes dados, preferiu-se levar em consideração apenas os países-sede das instituições de ensino dos autores, informação esta disponibilizada pela revista. A intenção aqui era, a partir deste mapa, considerar a relação da geopolítica dos artigos com as preocupações e questões das Ciências Sociais acerca do Big Data, no contexto da BD&S.

⁴⁷ Para um resumo de alguns destes relatos, ver DUEÑAS, 2012.

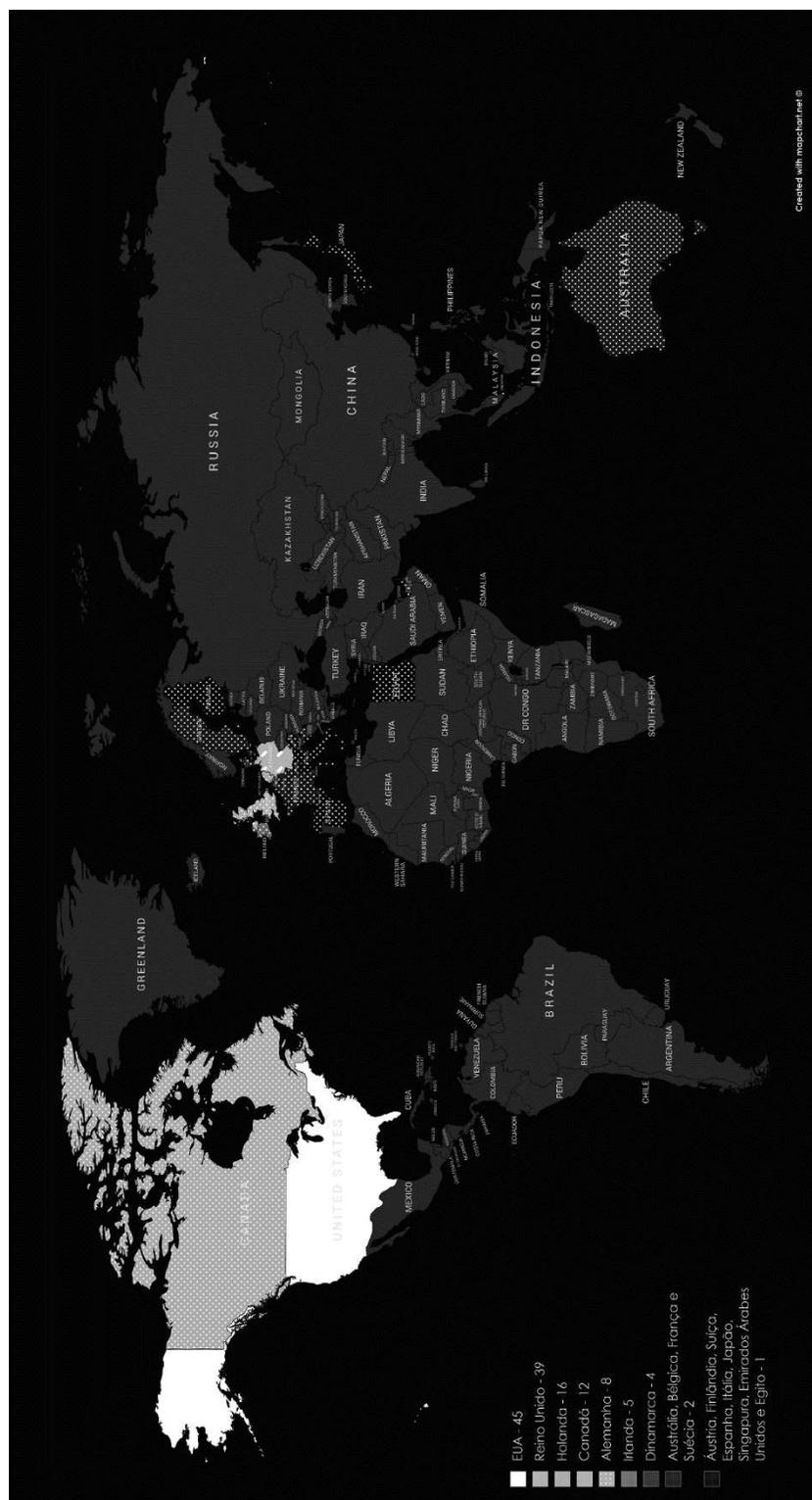


Figura 6: Origem acadêmica dos trabalhos publicados na BD&S entre 2014 e 2016. Mapa em negativo: tonalidades mais claras indicam presença mais expressiva na BD&S. Países marcados em tom cinza sem qualquer variação em sua superfície não tiveram nenhuma representação por parte de suas instituições de ensino na publicação.

O levantamento apontou que instituições estadunidenses, por meio de seus pesquisadores, encabeçam 45 dos 131 artigos publicados pela BD&S no período, o que configura mais de um terço das pesquisas (34,35%). Somadas com as instituições britânicas, que representam um total de 39 artigos (29,77%), a fatia britânico-americana alcança pouco mais de 64% do material produzido na BD&S. Destacam-se também a produção por parte de acadêmicos de instituições neerlandesas (16 artigos no total) e canadenses (12). Apenas quatro trabalhos partem de pesquisadores de instituições que fazem parte de países não-ocidentais: Egito, Singapura, Emirados Árabes Unidos e Japão⁴⁸.

Isso já aponta, desde o início, que a BD&S é uma publicação de circulação de certas comunidades científicas de centralidade euroamericana. Essa é uma informação que aponta para a relação entre a posição que a BD&S se propõe a ocupar, de um canal de interação e circulação de ideias para os pesquisadores de Big Data a partir de perspectivas críticas, e o real local de circulação de seus debates⁴⁹.

A proposta do mapeamento analítico feito em 131 artigos da BD&S publicados entre os anos de 2014 e 2016 foi de fornecer subsídios para uma compreensão da organização da pesquisa na BD&S e quais os principais temas e movimentos que a informam. A via quantitativa permitiu identificar alguns dos elementos básicos do funcionamento dos colégios invisíveis interpelados pela BD&S como principais lideranças, cânones e locais de circulação. Todavia, tratar de uma série de relações entre grupos científicos e o impacto disso na pesquisa em Big Data apresentada na BD&S envolve também uma análise que contemple uma via qualitativa, que alcance os porquês e como que os quês da análise quantitativa não puderam responder.

⁴⁸ Com a ressalva de que a instituição vinculada ao trabalho dos Emirados Árabes Unidos (EAU) é a New York University Abu Dhabi, um campus da New York University (EUA) nos EAU concebido a partir do projeto *NYU's Global Network*. Apesar do reconhecimento de que o conceito de Ocidente é largamente debatido atualmente, inclusive com questionamento por parte de alguns se o Japão ou Singapura já não poderiam ser considerados como parte do Ocidente (MILLER, 2004), a divisão acima segue o entendimento mais comum de civilização ocidental, como ilustrado na obra de Samuel Huntington. Ver HUNTINGTON, S. *O Choque das Civilizações e a Recomposição da Ordem Mundial*. Rio de Janeiro: Editora Objetiva, 1997.

⁴⁹ Discordo aqui da opinião defendida por Caroline S. Wagner (2008) de que a colaboração internacional é o remédio para democratizar a participação dos pesquisadores nas contribuições de sua área. Tal visão é, para mim, inocente ao tomar progresso e simetria como sendo o resultado 'natural' do fomento à rede internacional científica, quando, na verdade, talvez o próprio discurso da internacionalização representativa da ciência seja uma forma de opacizar assimetrias existentes na relação de poder entre o Ocidente e o Não-Ocidente. Trabalho com mais detalhes essa ideia idílica de integração científica no capítulo final.

Afinal, qual a relação entre as filiações daquelas comunidades científicas e os temas que ganham destaque na BD&S? Reconhecendo que toda pesquisa científica e processo de publicação em uma revista científica, inclusive nas próprias Ciências Sociais, tem uma natureza social (D'ANDREA; DECLICH, 2005), como essas interações acabam por modular os debates feitos na BD&S? Se de fato existe um reconhecimento das problemáticas dos dados na consolidação da desigualdade nas suas mais variadas facetas, e especialmente na relação Norte-Sul global, nos estudos e linha editorial da BD&S, como essa própria assimetria se reverbera na própria rede de relações das comunidades situadas ao redor da BD&S? O que leva à aparente discrepância entre o posicionamento crítico da BD&S em relação aos estudos sobre dados e os dados do levantamento realizado nesta pesquisa sobre a ausência epistemológica e de participação do Sul Global? Tais questões são decisivas para o desenvolvimento desta dissertação.

Em um primeiro momento, pretendo aqui tratar sobre meu primeiro contato com a BD&S e a experiência que tive ao participar de um evento promovido pela BD&S e que reunia grande parte de seu grupo editorial. Pretendo reunir aqui também informações que obtive de terceiros, de colegas que conheci na conferência, de outros relatos feitos por participantes do evento bem como outras informações que envolvam os membros destes colégios invisíveis. Reconheço aqui a limitação que se apresenta à narrativa de um evento de dois dias. O relato de experiência fornece um olhar único sobre as expectativas, interações sociais e projetos que cercam membros cuja atuação está diretamente relacionada ao funcionamento da BD&S. Sendo assim, fornece subsídios para se compreender a partir de quais perspectivas, lugares e referências se formam os discursos da BD&S e como o funcionamento desta comunidade científica circunscreve a maneira como avalia Big Data.

2.5 BD&S e sua comunidade científica: um relato de experiência

Uma das formas mais consistentes de encontro, consolidação e transmutação de uma comunidade científica se dá através dos eventos científicos. A importância dos eventos científicos como espaços privilegiados para a entrada de 'iniciantes' no seio da comunidade científica, a consolidação dos principais nomes e (re)configuração das discussões de determinado campo encontra há muito eco na literatura científica (GUIMARÃES; HAYASHI, 2014; ZUCCALA; BESSELAAR, 2009; GARVEY, 1979). Afinal, o caminho

para a publicação de um artigo numa revista científica passa quase que inevitavelmente pelos encontros informais proporcionados pelos eventos.

Na medida em que a comunicação científica cresceu e se tornou mais complicada nos últimos anos, os encontros científicos e conferências passaram a exercer uma função distinta e cada vez mais importante no processo geral de comunicação. [...] Muitas das informações compartilhadas e do valor atribuído a estes encontros vêm das interações [realizadas ali]. De fato, cientistas indicam que tais encontros estão entre as principais razões para participarem destes eventos (GARVEY, 1979, p. 41-42).

Sendo assim, uma das formas de entender os olhares das pesquisas publicadas na BD&S é olhar para os locais de reunião dos grupos de proeminência na publicação. Afinal, como dito anteriormente, tanto o comitê editorial e consultivo da publicação quanto a imensa maioria de seus autores estão vinculados a diferentes instituições euroamericanas, especialmente nos Estados Unidos e Reino Unido. É em conferências da área que estes nomes se encontram e lançam as bases do debate do campo.

Durante o período analisado nesta dissertação, a BD&S promoveu diversos eventos acadêmicos, inclusive publicando artigos apresentados nestas conferências. A primeira delas foi a *International Conference on Social Media & Society (SMSociety)*, organizada pelo *Social Media Lab*, da canadense *Ryerson University*. A BD&S publicou trabalhos apresentados nas edições de 2014 e 2015 do evento, que teve entre seus organizadores dois dos editores da BD&S: Anatolij Gruzd, docente na instituição-anfitriã, e Dhiraj Murthy⁵⁰. Ao todo, sete textos relacionados à conferência foram publicados na BD&S sob o dossiê temático intitulado “*Social Media & Society*”.

Embora não esteja diretamente ligada ao evento em questão, a BD&S também publicou um dossiê especial voltado para as discussões ampliadas realizadas em um dos encontros da Sociedade para a Filosofia da Informação (SPI) intitulado “*Conceptual challenges of data in science and technology*”. Entre os líderes da SPI está Luciano Floridi, membro do Comitê Consultivo da BD&S, além da presença de convidados como Rob Kitchin

⁵⁰ Informações disponíveis no site da conferência. Disponível em <<https://socialmediaandsociety.org/past-conferences/2015-2/2015-conference-program-committee/>> e <<https://socialmediaandsociety.org/past-conferences/2014-2/program-committee-2/>> Acesso em: 20/02/2020.

e Judith Simon, também vinculados à publicação⁵¹. Outros dossiês como o “*Spatial Big Data*”, publicado em 2016, resultado de uma das sessões do encontro anual da *Association of American Geographers* (AAG) de 2015⁵², ganharam espaço nas páginas da BD&S e demonstram a importância destes espaços para a constituição das visões que permeiam os artigos da BD&S.

A fim de aprofundar meu entendimento sobre a BD&S, resolvi converter numa experiência etnográfica minha participação em um dos eventos promovidos pela BD&S e seu corpo editorial e cujos trabalhos farão parte da publicação em 2020. O que se segue nas próximas páginas é um relato da experiência que tive ao participar da terceira edição da Data Power Conference, organizada pelo Zentrum für Medien-, Kommunikations- und Informationsforschung (ZeMKI), da Universität Bremen, na Alemanha, e realizada nos dias 12 e 13 de setembro de 2019. Além de promovido pela BD&S, que publicará em 2020 algumas das contribuições apresentadas no evento, a ocasião oferecia uma chance de encontrar alguns dos nomes constantemente citados na publicação, bem como alguns dos autores que ali publicaram seus trabalhos.

O relato pretende fornecer uma impressão geral das discussões ali presentes e das expectativas dos participantes em relação àquela comunidade científica e em relação à própria BD&S. Assim, se espera entender como as expectativas, organização e escopo de uma das comunidades que exercem impacto sobre os artigos da BD&S influem nas questões ali presentes e ausentes.

Meu primeiro contato com o evento em questão se deu através de uma lista de *mailing* na qual sou inscrito e que divulga regularmente vagas de emprego, chamadas de trabalhos, livros recém-publicados, chamadas de workshops ou cursos de verão, entre outras coisas, relacionados à área da Comunicação e afins⁵³. O e-mail, enviado em meados de janeiro de 2019, continha uma chamada para submissão de resumos para a Data Power Conference, que teria como tema a relação da concentração de poder sobre os dados com as inseguranças e desigualdades globais.

⁵¹ Informações disponíveis no site da SPI. Disponível em: <<https://socphilinfo.github.io/workshops/wpi7/home.html>> Acesso em: 20/02/2020.

⁵² Informações disponíveis no site da AAG. Disponível em: <http://www.aag.org/cs/annualmeeting/videos/2015_chicago> Acesso em: 20/02/2020.

⁵³ A lista pode ser acessada através do link: <<http://commlist.org/guidelines.html>> Acesso em: 20/02/2020.

O tema “In/seguranças globais” da Data Power Conference de 2019 abrange questões acerca destes fenômenos, perguntando: “Como o poder sobre os dados favorece ou desafia as in/seguranças globais? Como sociedade civil, governo e população se engajam com as in/seguranças individuais e coletivas geradas através dos ou com os dados? Quais são as ontologias apropriadas para se pensar acerca de dados e pessoas? Como podemos contemplar uma sociedade datificada justa? E *como seria uma decolonização das in/seguranças dos dados?* (ênfase acrescida)⁵⁴

Inicialmente, não imaginava que havia um vínculo entre a conferência e a BD&S, apesar de reconhecer vários dos nomes que constavam nas edições anteriores do evento [ocorridas na Universidade de Carleton, no Canadá, em 2015, e na Universidade de Sheffield, no Reino Unido, em 2017] como sendo de autores e figuras do comitê editorial da BD&S. Não possuindo condições financeiras de bancar uma viagem até a Alemanha para participar do evento, estava a ponto de arquivar o e-mail, até que, no final da chamada, reparei que haveria a disponibilidade de bolsas de viagem para participantes do Sul Global e isenção de pagamento de taxa de inscrição para mestrandos ou doutorandos.

Imediatamente comecei a elaborar um resumo de 250 palavras de parte do que foi levantado nesta pesquisa [um artigo contendo um resumo das principais ideias apresentadas na Data Power Conference foi publicado na revista Interações, de Portugal, no final de 2019 (SILVA, 2019)] e completei o processo de submissão. Recebi o aceite do meu texto no dia 03 de abril de 2019 e do aceite do pedido de auxílio viagem e isenção da taxa de inscrição no dia 19 do mês seguinte.

A iniciativa de oferecimento de auxílio para pesquisadores do Sul Global, apesar de não ser incomum na área, é sujeita a diversos fatores, como saúde financeira da instituição promotora, capacidade de atração de financiadores (o que passa pela capacidade de articulação política de lideranças de departamentos e/ou grupos de pesquisa) e alcance do tema de interesse. Mais de um dos participantes, além de um funcionário do RH da instituição, o qual me hospedou em sua casa durante os dias em que ali estive, me informaram que o ZeMKI era um dos departamentos que mais recebem aporte financeiro na universidade. De qualquer forma, tendo em vista as perspectivas sombrias para o fomento à pesquisa (e

⁵⁴ Chamada de trabalhos da Data Power Conference 2019. Disponível em: <<https://www.uni-bremen.de/datapower/call-for-papers-registration/>> Acesso em: 20/02/2020.

consequentemente à interação científica promovida pelos eventos científicos) no Brasil de Bolsonaro - e mesmo no cenário político instável da América Latina - comemorei o auxílio fornecido pelo evento como um título.

A iniciativa, por si só, parecia estar um passo à frente ante outros esforços de ampliar o escopo das discussões sobre dados. Não muito tempo após a realização do Data Power Conference, uma discussão tomou conta de uma lista de mailing relacionado ao grupo “Big Data a partir do Sul”, liderado por Stefania Milan e Emiliano Treré, da *Cardiff University* (Reino Unido), acerca de um evento organizado na universidade galesa abordando a assimetria Norte/Sul Global, porém sem qualquer tipo de auxílio financeiro para participação de pesquisadores de instituições latino-americanas ou africanas. A discussão girava em torno da efetividade do propósito da conferência ante o círculo de debates que ela realmente abrigaria.

Apesar do auxílio oferecido na Data Power Conference, no entanto, não houve muitos pesquisadores vinculados a instituições de ensino do Sul Global no evento. Conferindo a tabela de auxílios da conferência, averigui que o auxílio dado para minha viagem era de longe o maior oferecido (cujo valor alcançou €1325). O auxílio de valor mais próximo a esse fora oferecido a um pesquisador de uma universidade russa, que recebeu uma quantia ao redor de €800. Resta saber se os motivos para tal se devem a um desinteresse de pesquisadores do Sul Global nas temáticas discutidas ali, a uma dificuldade de tais pesquisadores em se adequar ao “roteiro epistemológico” comumente aceito dentro dos grupos ali presentes, ou a falta de conhecimento acerca da conferência. Tudo isso contribuiu para que me sentisse um certo ‘peixe fora d’água’ no evento. Cheguei a perguntar à uma das responsáveis pela equipe de credenciamento se havia algum outro brasileiro entre os participantes para que me sentisse um pouco mais ‘em casa’, porém até aquele momento nenhum outro brasileiro constava na lista.

A pesquisadora turca Güneş Tavmen, que também esteve presente na conferência deste ano assim como na primeira edição do evento, em 2015, também notou o fato de que, ao menos inicialmente, a conferência era “majoritariamente ‘branca’ em termos de palestrantes e assuntos discutidos. “À exceção de dois apresentadores, todos as apresentações de trabalho e palestras vieram de organizações na Europa, Austrália e Estados

Unidos/Canadá”⁵⁵. No seu texto, publicado no blog Dactive pouco após a conferência, Tavmen, que aborda iniciativas *open-source* e *smarts cities* em sua pesquisa, comparou os desenvolvimentos ocorridos entre 2015 e 2019 na Data Power Conference. Para ela, a tendência apresentada foi de maior diversidade geográfica e étnica entre os participantes, porém de maior concentração das temáticas e abordagens. “[Os trabalhos tiveram] um foco esmagador em como tornar os sistemas [de dados] mais éticos às custas da ausência de contestação da dominância destes sistemas”⁵⁶. Apesar de Tavmen traçar um relato mais positivo do evento de 2019 no que concerne à diversidade étnica, o esforço feito pela organização do evento em ampliar o escopo da discussão, incluindo pesquisas feitas fora do eixo América do Norte-Europa-Austrália, pouco surtiu efeito. Desde o momento em que cheguei no evento, não pude fazer mais do que me sentir um ‘peixe fora d’água’, no meio de uma constelação de estrelas que se reuniam ali.

A sensação de deslocamento começou assim que cheguei ao local do evento, que ocupava dois espaços diferentes, porém próximos, dentro da Universidade de Bremen, e percebi tanto no hall de entrada do edifício como na lista de participantes um grande número de pesquisadores que descobrira estarem entre os mais citados na BD&S, bem como uma quantidade considerável de autores e membros do Comitê Editorial da publicação. A impressão inicial naquela chegada é de que a imensa maioria dos presentes tinha de antemão algum tipo de intimidade ou conhecimento do trabalho de cada um. Na fila de credenciamento, por exemplo, entre as cerca de quinze pessoas que se aglomeravam à porta de entrada eu era o único que não estava em alguma rodinha de conversas. Já apressado para realizar o check-in e me dirigir à mesa de abertura, que seria sucedida por minha apresentação em uma das sessões de trabalho, acabei por não me engajar em interações alongadas naquele instante.

Após passar pelo credenciamento, assisti à palestra de abertura, proferida pela pesquisadora de Ciências Políticas Seeta Peña Gangadharan, da *London School of Economics*

⁵⁵ Tavmen publicou um texto no blog da Dactive traçando uma retrospectiva geral, a partir de sua experiência como participante, dos dois eventos da Data Power Conference em que pôde participar. O texto se encontra disponível em: <<https://data-activism.net/2019/11/bigdatasur-widening-the-field-of-critical-data-studies-reflections-on-four-years-data-power/>> Acesso em: 20/02/2020.

⁵⁶ “An overwhelming focus on how to make these systems more ethical and just with a lack of contestation of the domination of these systems”. Texto disponível em: <<https://data-activism.net/2019/11/bigdatasur-widening-the-field-of-critical-data-studies-reflections-on-four-years-data-power/>> Acesso em: 20/02/2020.

intitulada “What do just data governance strategies need in the 21st century?” [Do que estratégias justas de governança de dados precisam no século 21?]. O público escutava atentamente enquanto Gangadharan apresentava os resultados de um trabalho de campo que desenvolvia há alguns anos acerca do impacto de sistemas de coleta de dados e políticas públicas relacionadas aos dados em comunidades socialmente vulneráveis de algumas das grandes cidades estadunidenses.

De início, a apresentação parecia dar o tom de um evento que tinha por objetivo ampliar os escopos de análise tanto no que se refere aos temas discutidos quanto à quantidade de lugares e grupos sociais afetados pelos dados que teriam ali voz. Consumido pela preocupação com os slides da apresentação que faria logo em seguida, ainda mais em um idioma que não era o meu nativo, acabei por não prestar a devida atenção na breve discussão ocorrida após a apresentação. Todavia o engajamento do público foi nítido - tanto que gerou um atraso de quase 30 minutos para o início das sessões de trabalho. Resta saber se isso permaneceria assim para a maratona de palestras e apresentações de trabalho durante a manhã e à tarde, e atividades culturais na parte da noite, durante os dois dias seguintes.



Figura 7: Palestra inicial da Data Power Conference 2019, proferida pela Dra. Seeta Peña Gangadharan, da LSE. Foto borrada para preservar a identidade dos participantes.

Em seguida, me dirigi a um prédio próximo onde apresentaria meu trabalho em uma das salas de aula. A minha sessão tinha por título “*Big Data and humanitarianism vs. corruption and public debt*” [Big Data e humanitarismo vs. corrupção e dívida pública], o que, por si só, já sugere o caráter bastante eclético das contribuições apresentadas, já que o

meu trabalho ficava longe de discutir as questões colocadas pelo título da sessão - os temas variaram entre desenvolvimento sustentável e ético de inteligência artificial, passando por mapeamento de fluxo de dados na construção de novas tecnologias e até mesmo discussões sobre alteridade a partir da política migratória da União Europeia. A debatedora foi Jo Bates, membro do comitê editorial da BD&S, docente na Universidade de Sheffield (Reino Unido) e uma das co-fundadoras da Data Power Conference. A sessão prosseguiu com apresentações de acadêmicos de algumas instituições europeias e com um debate posterior. Para minha surpresa, não houve qualquer comentário da debatedora sobre o meu trabalho. Dos trabalhos apresentados, o meu era o único que, de fato, trazia a assimetria Norte/Sul Global como uma questão a se discutir, o que me impressionou por ser um evento que tinha a temática como um de seus motes.

O evento reuniu várias sessões de trabalho durante os dois dias em que ocorreu. Algumas ocorreram apenas em um determinado recorte de tempo de um dos dias, enquanto outras, como as sessões sobre abordagens feministas em relação aos dados, governança de dados e ativismo de dados/*open data* ocuparam múltiplos horários durante o evento. Devido a afinidade com a temática da minha pesquisa, me interessei em participar das sessões que abordavam tratativas decoloniais e sobre o ativismo de dados/*open data* em contextos não-ocidentais, como entre populações indígenas. Na parte da tarde, no primeiro dia do evento - e sem o peso de apresentar o trabalho nas costas - permaneci quase todo o tempo na sessão “*Data activism, citizen engagement, indigenous data sovereignty and open data*” [Ativismo de dados, engajamento civil, soberania de dados indígenas e dados abertos]. Essa era de longe a sessão mais cheia do evento. Sem conseguir achar assentos livres no ambiente, tive de sentar no chão, onde já havia pelo menos umas dez pessoas.

Chegando atrasado, demorei um pouco a me situar nas apresentações e discussões feitas ali. Não demorou muito para perceber que uma quantidade considerável dos presentes era vinculada aos trabalhos, seja como co-autores ou membros da mesma equipe ou grupo de pesquisa. Isso era perceptível não só pelo número de autores e co-autores presentes na lista⁵⁷, mas nos comentários feitos pelo público e nos agradecimentos proferidos pelos apresentadores. Diferente da sessão em que participei mais cedo, a sala era mais

⁵⁷ que pode ser acessada através do link: <<https://data-power.smart-abstract.com/sessionplanner/#/event/12237>> Acesso em: 20/02/2020.

eticamente diversa, especialmente devido a presença da pesquisadora zimbabuana Primrose Mawire e de membros da *Global Indigenous Data Alliance*. Inclusive, a apresentação do grupo sobre a questão da privacidade em contextos onde não há um sujeito ou indivíduo como o de comunidades *maori* na Austrália foi a grande atração da sessão. Foi na fala deles que ouvi uma pergunta que me fez refletir durante e após o evento, e que deixei registrado em meu bloquinho de anotações: “Onde estão os não-acadêmicos aqui?”. Uma pergunta propícia para a maior sessão de trabalhos do evento e cujo mote era o ativismo de dados e o engajamento civil.

Apesar de reparar uma presença majoritária ‘branca’ na conferência, pude notar pelas apresentações e quantidade de trabalhos que discussões sobre dados reconhecendo o impacto histórico do colonialismo e do neoliberalismo estavam em alta no evento. Numa outra sessão em que participei, ainda na tarde do primeiro dia de evento, intitulada “*Data visualisation at the margins: missing people, invisible people, imaginary people*” [Visualização de dados nas margens: pessoas desaparecidas, pessoas invisíveis, pessoas imaginárias], todas as apresentações se direcionaram à opacização social de grupos minoritários, apontada por alguns relatos como sendo fruto de um longo processo de relações desiguais de poder agravada pelo impulso neoliberal do século anterior. Porém, à exceção de algumas apresentações sobre políticas de dados e inteligência artificial na China e Zimbábue, e vigilância forense na África do Sul, as discussões ainda giravam em torno do contexto europeu, norte-americano e australiano/neozelandês.

As outras duas palestras realizadas no evento abrangeram análises sobre o funcionamento das políticas de dados do governo chinês, feitas pelo acadêmico chinês Dr. Jack Linchuan Qiu, da *Chinese University Hongkong*, além de uma fala sobre os desafios da educação indiana com relação ao avanço da automação na área de tecnologia da informação, feita pela indiana Dra. Nimmi Rangaswamy, do *Indian Institute of Information Technology*. Como não tinha estadia em hotel, e nem dinheiro para pagar uma diária em uma hospedagem próxima da universidade, perdi ambas as apresentações, que ocorreram no último horário do primeiro dia de evento e no primeiro horário do dia seguinte, respectivamente. O deslocamento até o local onde fui abrigado por uma amistosa família local era longo e cansativo. A impressão do primeiro dia, todavia, foi de um clima em geral inclinado a uma pluralização das bases de discussão e dos temas nos estudos sobre Big Data, porém com

pouca produção de fato sobre estes temas e menor participação ainda de acadêmicos de instituições de fora do eixo América do Norte-Europa-Oceania nas proposições ali apresentadas.

No segundo dia de evento, com a cabeça mais livre, tive mais tempo para conhecer os participantes e me engajar em conversas mais longas. Descrevo estes encontros mais abaixo. Como ouvinte, participei de uma sessão intitulada “*Forensic data, human rights and refugees*” [Dados forenses, direitos humanos e refugiados] que reuniu contribuições importantes questionando a instrumentalização do crime nos processos de vigilância - isto é, a ideia de que violência e crime deixam de ser fenômenos de análise sócio-política para alcançarem sua resolução através da ‘transparência’ dos dados e de suas categorias objetivantes como raça, cor, gênero e características étnicas. O sabor eurocêntrico de tais categorias foi apontado com frequência nesta sessão, que reuniu um número considerável de participantes da Universidade de Minho, em Portugal, membros de um projeto de pesquisa sobre dados forenses financiado pelo Conselho Europeu de Pesquisa (ERC). O número de membros na apresentação foi reduzido em comparação às sessões do dia anterior, não somente na sessão onde estava, mas em outras ao redor.

O momento que mais aguardara durante o evento era o painel de encerramento do mesmo, que ocorreria no final da tarde do segundo dia, cujo título era “*Decolonising data: undoing the South*”, e que teria a presença das pesquisadoras Monika Halkort (*Lebanese American University* - Líbano), Stefania Milan (*University of Amsterdam* - Países Baixos), Tracey Lauriault e Merlyna Lim (*Carleton University* - Canadá). E não me decepcionei com o que ouvi, apesar de Lim e Milan não terem contribuído com nenhuma fala. Halkort elaborou uma crítica a um modelo universalista de proposta de decolonização dos dados - como o oferecido, segundo ela, por Couldry e Mejias (2019), por exemplo - e uma mudança de foco da apropriação para a relação construída com e pelos dados. Lauriault, na mesma linha, trouxe um mapeamento pós-colonial sobre as relações históricas, linguísticas e ‘nominalísticas’ entre povos *inuits* do Círculo Polar Ártico e como a infraestrutura de rede pode auxiliar na desconstrução de nomes impostos pela tradição colonial e no resgate da história e dos nomes apagados destes povos antigos. Uma coisa não podia negar naquele momento, o Sul Global como tema nos estudos sobre dados trazia reflexões bastante obtusas em um cenário de início de pesquisa em Big Data.

Devido a quantidade de sessões simultâneas, cujo número chegava a seis, não pude acompanhar todas, porém percebi que algumas das sessões reuniam grupos já bastante familiarizados com as pesquisas apresentadas nestes espaços. Pequenos grupos de pesquisadores vinculados a instituições de pesquisa do mesmo país ou língua se congregavam durante o coffee break, marcados especialmente por pesquisadores de universidades britânicas como a Sheffield, Cardiff e a LSE, em maioria no evento.

Desejando conhecer um pouco mais das expectativas dos participantes em relação ao evento, comecei a me engajar em algumas das rodas de conversa que dominavam os corredores. Duas pesquisadoras portuguesas da Universidade do Minho me contaram estarem em uma maratona de apresentações em congressos organizados no continente europeu sobre o projeto que realizam no Departamento de Sociologia da universidade lusófona, financiado pelo European Research Council, acerca das implicações do uso de dados na investigação criminal. Com muito em jogo, participar de um evento com nomes reconhecidos nos estudos sobre dados, e presença de importantes departamentos de áreas como comunicação e sociologia no continente europeu, oferecia um meio importante de divulgar os resultados da pesquisa. A mesma razão para participação no evento foi apontada por outros pesquisadores, um da Holanda e outro da Alemanha, com os quais conversei em algumas rodas de discussão após as sessões.

Em outras rodas de conversa, conheci pesquisadores que vinham da Inglaterra, Países Baixos, Estados Unidos, Portugal e da própria Alemanha. Alguns estavam na fase final de suas respectivas teses doutorais e viam no evento a oportunidade de consolidar a pesquisa dentro da comunidade científica dos “Estudos Críticos sobre os Dados”⁵⁸ e a maioria dos que abordei era composta de pesquisadores em início de carreira, tendo recém-concluído o doutorado e à procura de projetos de pesquisa ou vagas em alguma universidade. De fato, foram poucas as ‘cabeças-brancas’ que testemunhei no encontro. Poucos pareciam ter mais de 50 anos, o que para mim foi uma surpresa ante a quantidade de citações que alguns

⁵⁸ É importante salientar que, repetidamente, os organizadores do evento sinalizavam ser este um encontro sobre os Estudos Críticos sobre os Dados, termo dado para descrever uma série de propostas que questionem empiricamente ou teoricamente o estatuto puramente empírico e neutro dos dados (cf. ILIADIS; RUSSO, 2016; DALTON; THATCHER, 2014; DALTON; TAYLOR; THATCHER, 2016; KITCHIN; LAURIAULT, 2018; CRAWFORD; MILTNER; GRAY, 2014). Parte dos membros do Comitê Editorial e dos organizadores da Data Power Conference estão entre os principais propositores dos Estudos Críticos sobre os Dados. Em 2016, a BD&S dedicou um espaço dentro da publicação para esta temática. Disponível em: <<https://journals.sagepub.com/page/bds/collections/critical-data-studies>> Acesso em: 20/02/2020.

daqueles nomes recebiam nos estudos sobre dados. Talvez um reflexo do quão recente é o tema.

Ao descobrirem que minha pesquisa girava em torno da BD&S, muitos apontaram que pretendiam submeter um trabalho para a publicação ou mesmo já haviam publicado um artigo na BD&S. Tive a oportunidade de conhecer inclusive um autor cujo texto, publicado na revista em 2015, se encontra analisado no capítulo 3 desta dissertação. A reação de alguns ao descobrirem que era brasileiro e ainda cursando o Mestrado foi cercada de ‘meus pêsames’ e ‘lamento’. Muitos escutaram sobre o desmanche da educação pública promovido pelo governo Bolsonaro. Na época em que fui para o evento, havia um sério risco de que todas as bolsas de estudo oferecidas pela Coordenação de Aperfeiçoamento de Pessoal do Ensino Superior (Capes) para o ano subsequente fossem canceladas. Alguns me incentivaram a que submetesse um artigo para a BD&S ou tentasse ‘cavar’ alguma vaga em uma equipe de pesquisa com bom financiamento e interessada em pesquisadores internacionais - cujos líderes se encontravam na conferência.

Mais de uma vez, inclusive, tive uma das breves conversas informais de corredor interrompida ao informar que era brasileiro. “Ah Brasil? Quero nem falar sobre o Brasil depois que aquele fascista assumiu o poder” foi a reação de uma das participantes de uma mesa de debates ao final do evento após abordá-la com questionamentos sobre sua fala. A maior parte lamentava o estado político em que o Brasil se encontra e a falta de incentivo financeiro para a pesquisa e os pesquisadores nas universidades públicas. Não houve perguntas, todavia, sobre o que se falava por aqui, o que chamava a atenção dos acadêmicos que pesquisam dados ou sobre como enfrentávamos as novas ameaças no contexto digital. Não faz mal lembrar que eu era o único participante de uma instituição sul-americana na conferência.



Figura 8: Um dos momentos para a interação social ocorreu na noite do primeiro dia de evento, com a realização de um jantar para os participantes em um restaurante próximo à universidade que sediou a *Data Power Conference*. Foto borrada para preservar a identidade dos participantes.

Além das sessões de apresentação de trabalho e das plenárias, o evento também organizou algumas atividades sociais e culturais durante sua programação. Cada uma delas tinha um pequeno custo que variava entre 10 e 25 euros. Devido à escassez de recursos, não pude participar da maior parte delas, lamentando em especial a ausência no jantar festivo da conferência, realizado na noite do primeiro dia de evento e reunindo um bom número de participantes da *Data Power Conference*. Todavia, me inscrevi em uma dessas atividades, que envolvia um passeio na zona portuária da cidade (*Überseestadt*) a fim de conhecermos as marcas do colonialismo na infraestrutura daquela região.



Figura 9: Região do Überseestadt onde foi realizado o tour sobre o passado colonial de Bremen.

A atividade foi de longe a programação mais interessante proporcionada pelo evento com relação a sua proposta inicial. A caminhada durou pouco mais de duas horas, para quem aguentou o percurso, ao menos; a maior parte dos participantes ‘abandonou o barco’ antes da metade do *tour*. Durante o percurso fomos apresentados a um legado muitas vezes oculto na progressiva cidade de Bremen de marcas repressivas de um colonialismo que, às custas do enriquecimento da elite da cidade-estado no fim do século XIX, acabou por diminuir consideravelmente as condições de vida dos trabalhadores que vinham ali estabelecer-se e de suas comunidades, além de expor a exploração das colônias alemãs no continente africano.

Todo o percurso foi guiado por um coletivo local chamado ‘*Decolonize Bremen*’⁵⁹, um grupo que busca através de diferentes meios, incluindo cinema experimental, ação política e peças teatrais, além é claro do tour gratuito oferecido, abrir o jogo para a

⁵⁹ O site do coletivo pode ser acessado através do link: <<https://www.decolonizebremen.com/>> Acesso em: 20/02/2020.

população e figuras públicas acerca de como o passado colonialista é visto (ou esquecido) pela sociedade local. O fato de a excursão ter sido liderada pelo coletivo, em um tom distante do ‘academicismo’ que imperava no evento - e de projetos financiados à rodo por instituições com um caixa financeiro elevado, foi um refrigério.

Durante um dos deslocamentos do *tour*, conversei mais de perto com um dos rapazes que atuava como guia. Entre papos sobre a situação política do Brasil, da Alemanha e como viver em uma era ameaçada pela extrema direita, me recordo de ouvi-lo dizer: “Todo o esforço de conter isso aí está na integração de forças, vocês dentro da universidade e nós entre as entidades populares. Ninguém resiste a nada sozinho”. Naquele momento, só pude refletir nos motivos que realmente me levaram ali para Bremen e se ‘resistência’ e ‘integração’ realmente estavam entre eles ou mesmo no que foi aquele evento em que participei. Ao final do *tour*, eu e três outras pesquisadoras que participavam da *Data Power Conference*, e que resistiram até o final do programa social, trocamos contato com os membros do coletivo e nos despedimos. Fim do curto evento que me levou a cruzar o Atlântico.

2.6 Impressões e discussão

Após a experiência de participar de um evento como a *Data Power Conference*, que reuniu um grupo bastante nuclear na área de estudos críticos sobre dados e de influência sobre os rumos das pesquisas publicadas na BD&S, sou obrigado a concordar com Tavmen, também de que certamente houve um reconhecimento da necessidade de se decolonizar a área e de “ampliar o campo em termos de inclusão geográfica e sócio-cultural”⁶⁰. Todavia, enquanto tal esforço foi evidente na temática do evento, no fornecimento de auxílio para participantes do Sul Global e nas sessões ‘inchadas’ sobre Big Data e colonialismo, o evento ainda me pareceu em larga medida um encontro de compadres, ligados por projetos de grande envergadura financeira e institucional realizados no solo europeu e unidos por interesses de pesquisa e estudos de casos mais ou menos comuns.

⁶⁰ “there was a substantial effort to widen the field in terms of geographical and socio-cultural inclusion.” Texto disponível em: <<https://data-activism.net/2019/11/bigdatasur-widening-the-field-of-critical-data-studies-reflections-on-four-years-data-power/>> Acesso em: 20/02/2020.

Embora a presença de pesquisadores ligados à instituições de ensino do Sul Global tenha sido pequena, não é para menos, considerando que, como exposto no relato, a imensa maioria das discussões girava em torno de estudos de casos de dados a partir do eixo Europa-América do Norte-Austrália/Nova Zelândia, com algum enfoque na China e Índia. Além disso, em apenas duas vezes, uma na plenária liderada pelo acadêmico chinês Dr. Jack Linchuan Qiu e pelos membros do movimento *Global Indigenous Data Alliance*, em uma das sessões de trabalho, me recorde de escutar referências a ideias e raciocínios construídos por autores do Sul Global. Muitas das referências evocadas nas apresentações de trabalho, inclusive, se voltavam para algumas das próprias ‘estrelas’ do evento. Somente na minha sessão, por exemplo, Jo Bates, a debatedora e um dos grandes nomes ali presentes, foi citada em duas das cinco apresentações realizadas.

Apesar das intenções pluralizantes e decolonizantes, a *Data Power Conference* me pareceu apresentar um desejo de ampliar os horizontes, mas ainda centralizando a pesquisa naquele nicho. Uma ânsia por ampliar as perspectivas sobre os dados, ‘oxigenando’ os debates feitos ali, mas a partir da estrutura financeira, epistemológica e institucional já tão ‘consagrada’ naquele meio, em especial pelo cânone recente reunido ali. Uma tentativa de congregar o saber e os poderes de todos os quinhões do mundo para pensarem nos problemas e soluções para uma realidade que interessa mormente a órgãos como o Conselho Europeu de Pesquisa. Não à toa, foi pela via cultural, através do tour realizado após as tratativas oficiais da conferência, que o tema do evento alcançou maior potência. Qual a disposição de tal comunidade científica em expandir seus horizontes étnicos e institucionais a fim de responder como viver numa sociedade cada vez mais interpelada por problemas de ordem social e cultural relacionados aos dados?

A constatação acima auxilia na compreensão dos dados apresentados anteriormente sobre a BD&S, que, vale recordar, foi uma das promotoras da conferência em Bremen. Embora se proponha como um espaço por excelência para o livre fluxo de propostas críticas sobre os dados a partir da contribuição das mais diversas subáreas dentro das Ciências Sociais, as comunidades científicas que se articulam ali, especialmente em sua liderança, os temas ali discutidos e as referências que fomentam os debates acabam por limitar o ambiente de circulação desta ao de poucas escolas norte-americanas, europeias e do Pacífico Sul, tendo como escopo o das democracias liberais e das megacorporações do Ocidente. Sendo a BD&S

uma das poucas, senão a única, publicação científica que toma como foco específico o diálogo entre as diversas problemáticas do Big Data a partir das Ciências Sociais, pode-se inferir que essa é uma realidade a ser encontrada nos estudos em Big Data nas Ciências Sociais (GANTER; ORTEGA, 2019; ARORA, 2019; HALKORT, 2019).

Porém, qual a extensão da influência desta realidade no conteúdo da pesquisa em Big Data? De que forma isto informa os artigos da BD&S e seus enfoques? Para isso, a análise se volta agora para quatro dos principais textos referenciados na BD&S e cuja influência, apesar da hodiernidade do tema, faz valer nesta pesquisa o título de ‘canônicos’.

3. Olhares para onde? Uma análise a partir dos cânones da BD&S

3.1 Apresentação

Enquanto no capítulo anterior, o mapeamento analítico e o relato de experiência da minha participação em um dos eventos promovidos pela BD&S permitiram um olhar quanti-qualitativo sobre algumas das principais prioridades, temas, lugares de fala e sobre o ‘organismo social’ que articula a BD&S, seguimos explorando aqui quatro textos que exerceram um perceptível impacto nas pesquisas publicadas na BD&S, dentro do escopo investigado. São estes o artigo ‘*Critical Questions for Big Data: provocations for a cultural, technological, and scholarly phenomenon*’, de danah boyd e Kate Crawford; o livro ‘*The Data Revolution: big data, open data, data infrastructures and their consequences*’ e o artigo ‘*Big Data, new epistemologies and paradigm shifts*’, de Rob Kitchin; o livro ‘*Big Data: a revolution that will transform how we live, work, and think*’, de Viktor Mayer-Schönberger e Kenneth Cukier; o artigo ‘*The Coming Crisis of Empirical Sociology*’, de Mike Savage e Roger Burrows.

Como visto no capítulo introdutório, os trabalhos de Rob Kitchin, Kate Crawford, danah boyd, Bruno Latour, Roger Burrows, Mike Savage, entre outros, ocuparam um lugar central na articulação dos debates da BD&S e mesmo nos temas da Data Power Conference, apresentada no capítulo 2. Do total de artigos analisados, quase metade (62 artigos, 47,3%) contém referências a um dos autores mencionados acima.

Este passo envolve a análise do que fundamenta estes trabalhos e quais suas propostas de pesquisa, além da relação que os artigos da BD&S constroem com estes textos em suas abordagens, reconhecendo, mais uma vez, que a própria existência dos cânones bem como sua influência, ou seu caráter de ser um ‘ponto de passagem obrigatório’, é de uma natureza eminentemente social, podendo assumir variadas formas (BENTIVOGLIO, 2012). Deve-se reconhecer que mesmo que a capacidade fundadora destes textos clame por uma articulação dos artigos que retome seus pontos (VANDERSTRAETEN, 2010),

cânones, que negam a distinção entre conhecimento e opinião e que são instrumentos de sobrevivência construídos para serem à prova do tempo, à prova da razão, são certamente desconstrutíveis. Se as pessoas pensam que tais coisas não são assim, podem muito bem encontrar meios de destruí-los.

A defesa deles [dos cânones] não pode mais ser empreendida por um poder institucional central; eles não podem mais ser compulsórios, apesar de que é difícil imaginar como a operação normal das instituições do saber, incluindo o recrutamento [para suas fileiras], pode subsistir sem eles (KERMODE, 2011, p. 78)

Sendo assim, busca-se aqui uma empreitada de caráter quase genealógico⁶¹ dos artigos da BD&S a fim de entender como os artigos percorrem os caminhos deixados pelos cânones. Reconhecendo a centralidade destes textos nas questões discutidas na BD&S, como os autores interpelam os temas colocados por estes textos em seus próprios trabalhos? O fazem de modo a meramente reafirmar o que está dado ali ou a retrabalhar, ampliar e mesmo examinar sob um caráter crítico estes cânones? Tais questionamentos se mostram essenciais para esta pesquisa pois apontam para o recrudescimento destes pontos de vista canônicos dentro dos colégios invisíveis interpelados pela BD&S, em caso e reafirmação quase unânime destes pontos no material, ou para uma amplitude maior de visões em relação aos cânones.

Estudos com viés relativamente semelhante ao realizado neste trabalho, no contexto da interface Big Data-Ciências Sociais, apontaram para uma prática crescente de “reprodução de estruturas conhecidas e relatos acadêmicos” (GANTER; ORTEGA, 2019, p. 82) centralizados em poucos grupos espalhados por instituições norte-americanas e europeias (YOUTIE et al, 2017). Tal questão, já reconhecida no capítulo anterior no contexto da BD&S, é trazida à tona aqui sob o espectro, desta vez, da relação desta realidade com os conteúdos dos cânones e o modo como estes são apropriados nos artigos da BD&S. O que ali contribui para a consolidação desta realidade? De que forma isto contribui para a ausência de reconhecimento da assimetria existente entre essa realidade ocidental e a não-ocidental?

Para efeitos de facilitação do processo de leitura e classificação dos textos, que aparecem aqui em quantidade considerável, propõe-se um sistema de classificação que elenca os artigos em: 1) referência instrumental – aqui se referindo a reafirmação dos conteúdos apresentados nos cânones, sem maiores elaborações; 2) referência dialógica – aqui se

⁶¹ Nunca é demais ressaltar que uso o termo genealogia não no sentido historiográfico de traçar uma linhagem que aponte todo o conjunto do objeto como pertencendo a uma origem ou ancestral comum, constituindo, portanto, um modelo universal de explicação de determinado objeto. Utilizo genealogia como referência ao próprio ato de lançar mão de um mapeamento de dada situação, neste caso a pesquisa em Big Data, não ao resultado do mapeamento, não a uma explicação causal dos porquês deste campo. Creio que os apelos de Gilles Deleuze e Félix Guattari (1995) por um pensamento antigenealógico são muito úteis para a prevenção da busca por fundamentos últimos ou estruturas-raízes para qualquer tipo de empreitada acadêmica.

referindo à construção de diálogos a partir das questões colocadas nos cânones, ampliando o que fora colocado ali ou explorando suas possibilidades para além dos domínios alcançados pelos textos referidos; 3) referência crítica – aqui se referindo à refutação, por parte dos autores, parcial ou completa, dos pontos de vista apresentados nos cânones, apontando-se limitações destas. Apesar de artificial, e sem nenhuma pretensão de algo como ‘representar’ a realidade destes trabalhos, tal classificação serve de apoio para os argumentos centrais defendidos neste capítulo, além de não ser uma estratégia estranha a levantamentos de natureza bibliométrica e metodologia mista (e.g. ARAÚJO, 2009).

3.2 boyd & Crawford

Dentre tudo o que se foi publicado até o presente momento relacionado ao Big Data, não há dúvida de que o texto “*Critical Questions for Big Data*”, de danah boyd e Kate Crawford, figura entre os trabalhos mais importantes, constituindo-se quase como uma referência básica nos estudos em Big Data e na Ciência Social Computacional (YOUTIE et al, 2017). O artigo, publicado no 22º volume da revista *Information, Communication & Society*, no ano de 2012, uma das principais publicações acadêmicas nos campos de Comunicação e Ciências Sociais do mundo⁶², é, de longe, o mais citado e visto da história da revista, com mais de 123 mil visualizações (quase quatro vezes mais que o segundo colocado) e cerca de mil citações (o dobro em relação ao segundo artigo mais citado)⁶³. Ambas, a norte-americana danah boyd e a australiana Kate Crawford, trabalham como pesquisadoras independentes do *Social Media Collective*, uma iniciativa de pesquisa multidisciplinar vinculada a *Microsoft Research*.

Buscando trazer uma contribuição crítica para o que viam como um oxímoro⁶⁴, usado ‘pobrememente’ apenas no sentido de inovação, solução e neutralidade, as autoras buscaram lançar seis ‘provocações’ para uma análise crítica do fenômeno do Big Data. Para

⁶² Figura em 6º na categoria “Comunicação” e 3º em “Ciências Informacionais” no Journal Citation Reports de 2018, publicado pela SCImago. Disponível em: <<https://www.scimagojr.com/journalrank.php?category=3309>> e <<https://www.scimagojr.com/journalrank.php?category=3315>> Acesso em: 20/02/2020.

⁶³ Até março de 2019. Estas são informações da própria publicação, disponibilizadas em seu site. Disponível em: <<https://www.tandfonline.com>> Acesso em: 20/02/2020.

⁶⁴ Termo consagrado por Geoffrey Bowker em sua obra “*Memory Practices in the Sciences*” (2006, p. 184): “Raw data is both an oxymoron and a bad idea; to the contrary, data should be cooked with care.”

elas, primeiramente, o fator definitivo do Big Data não seria meramente o tamanho “grande” dos dados, mas uma convergência entre aspectos tecnológicos, analíticos e mitológicos:

Tecnológico: maximização do poder computacional e da precisão algorítmica a fim de reunir, analisar, conectar e comparar grandes bases de dados. *Analítico*: dependência de grandes bases de dados para identificação de padrões a fim de se realizar reivindicações econômicas, sociais, técnicas e legais. *Mitológico*: a crença difundida de que grandes bases de dados oferecem uma forma mais elevada de inteligência e conhecimento que pode gerar *insights* previamente impossíveis, com uma aura de **verdade, objetividade e precisão**” (BOYD & CRAWFORD, 2012, p. 663, grifo das autoras)

A partir do que veem como uma necessidade de uma “chamada para reflexão sobre as coisas mesmas com as quais os pesquisadores trabalham - dados”, e em diálogo com autores como Bruno Latour, David Berry, Lev Manovich, David Bollier, entre outros, boyd e Crawford (2012) apresentam seis provocações para os estudos em Big Data. Tais provocações estão sintetizadas nas seguintes afirmações: 1) “o Big Data altera a definição do que é conhecimento”; 2) “reivindicações de objetividade e precisão são enganosas”; 3) “dados maiores não significam dados melhores”; 4) “tirado de contexto, o Big Data perde seu significado”; 5) “o fato de os dados estarem acessíveis não os torna éticos”; 6) “acesso limitado ao Big Data cria novas brechas digitais”⁶⁵.

As provocações, segundo as autoras, não envolvem apenas a reflexão e pesquisa acadêmica, mas também empresas e governos. De fato, elas criticam o que veem como uma constante tentativa de menosprezar o envolvimento da academia na análise do Big Data na base de que a “indústria consegue fazer melhor” (CONOVER apud BOYD & CRAWFORD, 2012, p. 674). O chamado geral de boyd e Crawford para questionar o Big Data a partir de suas provocações pode ser visto pela influência do artigo nas mais diversas áreas de pesquisa, como: administração de empresas (JENNEX, 2019; KWON et al, 2014), saúde pública (FINFGELD-CONNETT, 2014; SHARON, 2017), administração e políticas públicas (EDELLENBOS et al, 2018), ciências da computação (STEHLE & PEUQUET, 2015; LEWIS

⁶⁵ Tradução mais comumente utilizada para “*digital divides*”, termo usado aqui por boyd e Crawford, tanto em português como em espanhol. Parte da literatura das Humanidades Digitais considera o termo “*digital inequalities*” como proporcionando uma compreensão menos dicotômica e impositiva tanto do papel que as TICs operam no avançar da desigualdade social, como do caminho contrário (cf. HARGITTAI; HSIEH, 2013).

et al, 2013), filosofia da ciência (SYMONS & ALVARADO, 2016; SÆTRA, 2018) e ciências sociais (MOATS & SEAVER, 2019; OVADIA, 2013)⁶⁶.

O impacto do trabalho de boyd e Crawford alcançou mesmo documentos de organizações internacionais⁶⁷ e textos jornalísticos⁶⁸. A importância deste artigo para o estudo do Big Data levou Andrew Iliadis e Federica Russo (2016, p. 2) a classificá-lo, apenas quatro anos após sua publicação, como um “texto clássico”, informando um incômodo crescente em diversos campos de estudo acerca do status do Big Data. Não à toa, o artigo se situa exatamente no começo de uma ascensão meteórica nas pesquisas em Big Data, em especial nas Ciências Sociais, como revela o gráfico abaixo:

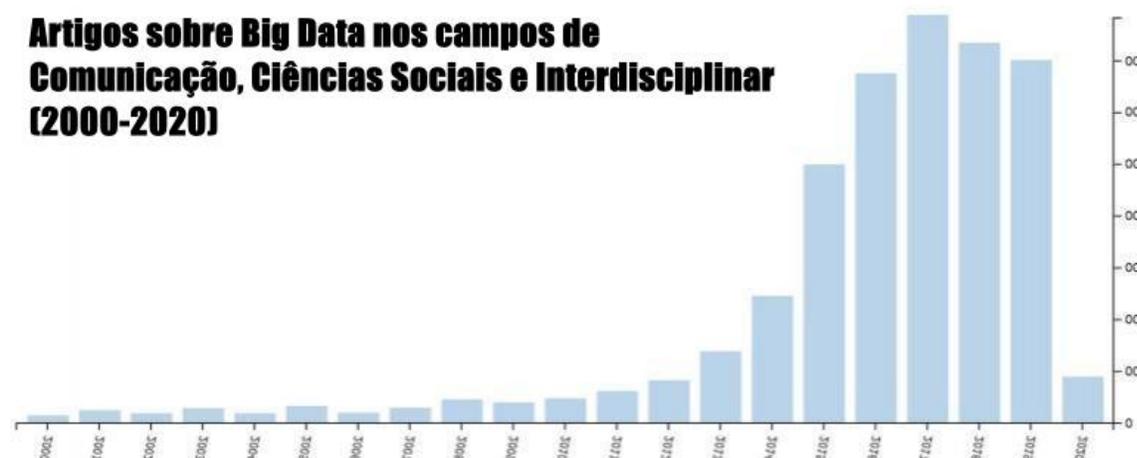


Figura 10: Número de artigos sobre “Big Data”, nos campos de “Comunicação”, “Ciências Sociais” e “Interdisciplinares”, listados na base de dados da Web of Science no período entre 2000 e 2020. Entre 2012, ano da publicação do artigo de boyd e Crawford, e 2017, a taxa de publicações sobre “Big Data” é quintuplicada. Dados atualizados até 16/03/2020.

⁶⁶ Tanto na base de dados da Scopus quanto na do Web of Science, áreas como ciências da computação e administração estão entre as três que mais fizeram uso do artigo de boyd e Crawford em suas pesquisas. Disponível em:

<http://wcs.webofknowledge.com/RA/analyze.do?product=WOS&SID=7EzOCnVxSMmmHDpwjQC&field=TASCA_JCRCategories_JCRCategories_en&yearSort=false>

e <<https://www.scopus.com/term/analyzer.uri?sid=7fbc5502595f3dff2835786c5eb5a9d7&origin=resultslist&sr=c=s&s=REFEID%282-s2.0-84861974217%29&sort=plf-f&sdt=a&sot=cite&sl=26&count=1521&analyzeResults=Analyze+results&imp=t&cite=2-s2.0-84861974217&txGid=c9fbc9d6a118f855eafb8205bf45b180>> Acesso em: 20/02/2020

⁶⁷ Como, por exemplo, o publicado pela Unesco em 2015, intitulado “Keystones to foster inclusive knowledge societies: access to information and knowledge, freedom of expression, privacy, and ethics on a global internet”. Disponível em: <<https://unesdoc.unesco.org/ark:/48223/pf0000232563>> Acesso em: 20/02/2020.

⁶⁸ Ver, por exemplo, “Facebook’s data lockdown is a disaster for academic researchers“. Disponível em: <<http://theconversation.com/facebooks-data-lockdown-is-a-disaster-for-academic-researchers-94533>> Acesso em: 20/02/2020.

Apesar da relevância do artigo e de sua proposta crítica, boyd e Crawford se engajam pouco com relatos que fortaleceriam suas provocações, algo notável para duas autoras cujo trabalho etnográfico constituiu a pesquisa doutoral de ambas, aproveitando pouco do trabalho de campo que realizaram⁶⁹. Isso vale em especial para relatos e narrativas que fogem das questões que preocupam o ambiente euroamericano de pesquisa. Questões como divisões sociais e exploração e uso de dados são enfatizadas como sendo universais, quando, de fato, encontram contornos bastante específicos no Sul Global, por exemplo (e.g MAHRENBACH et al, 2018). Como se verá, isso tem sido uma constante nas pesquisas em Big Data que informaram os materiais analisados na BD&S.

Após a breve introdução ao artigo de danah boyd e Kate Crawford, é importante situar a forma como este influente trabalho informa as pesquisas que constituem o escopo da presente análise. Como visto no capítulo introdutório, o texto aparece como uma das principais referências em 29 artigos da BD&S no período entre 2014 e 2016 (KITCHIN, 2014b; KSHETRI, 2014; SCHROEDER, 2014; HOUSLEY et al, 2014; BLOK & PEDERSEN, 2014; QUAN-HAASE et al, 2015; BAACK, 2015; DALTON & THATCHER, 2015; KENNEDY & MOSS, 2015; KENNEDY et al, 2015; PERNG et al, 2016; METCALF & CRAWFORD, 2016; RIEDER & SIMON, 2016; DALTON et al, 2016; BRONSON & KNEZEVIC, 2016; VAN DER VLIST, 2016; BROOKER et al, 2016; THATCHER, 2016; LOUKISSAS, 2016; LESZCZYNSKI & CRAMPTON, 2016; CHRISTIAENS, 2016; SYMONS & ALVARADO, 2016; FORTUN et al, 2016; CANALI, 2016; LIPPERT, 2016; HONG, 2016; ILIADIS & RUSSO, 2016; THORNHAM & GÓMEZ CRUZ, 2016; VELKOVA, 2016)⁷⁰.

Uma leitura da maneira como cada um dos 29 artigos se apropria do trabalho de boyd e Crawford foi organizada na classificação exposta na introdução deste capítulo: 1) referência instrumental; 2) referência dialógica; e 3) referência crítica. Tais perspectivas foram classificadas e agrupadas abaixo na Tabela 3:

⁶⁹ Por exemplo, dois anos depois do lançamento do artigo na *Journal of Communication*, boyd lança um livro de 281 páginas apresentando o resultado de um trabalho de campo com adolescentes sobre a relação destes com as redes digitais, trabalho este que já realizava na época da publicação de seu artigo com Crawford (BOYD, 2014).

⁷⁰ Vale uma menção ao trabalho de Lagoze (2014), que cita um esboço do mesmo artigo de boyd e Crawford apresentado numa conferência no ano anterior ao da publicação do material na *Information, Communication & Society*.

CATEGORIAS	ARTIGOS
Referência instrumental	(KITCHIN, 2014); (KSHETRI, 2014); (HOUSLEY et al, 2014); (QUAN-HAASE et al, 2015); (DALTON; THATCHER, 2015); (KENNEDY; MOSS, 2015); (KENNEDY et al, 2015); (METCALF; CRAWFORD, 2016); (RIEDER; SIMON, 2016); (DALTON et al, 2016); (BRONSON; KNEZEVIC, 2016); (VAN DER VLIST, 2016); (BROOKER et al, 2016); (THATCHER, 2016); (LOUKISSAS, 2016); (LESZCZYNSKI; CRAMPTON, 2016); (CHRISTIAENS, 2016); (FORTUN et al, 2016); (CANALI, 2016); (ILIADIS; RUSSO, 2016); (THORNHAM; GÓMEZ CRUZ, 2016); (VELKOVA, 2016)
Referência dialógica	(BAACK, 2015); (LIPPERT, 2016); (HONG, 2016)
Referência crítica	(BLOK; PEDERSON, 2014); (SCHROEDER, 2014); (PERNG et al, 2016); (SYMON; ALVARADO, 2016);

Tabela 3: Classificação organizada com base nos modos em que os artigos da análise se apropriam da obra de boyd e Crawford.

A classificação foi feita com base nas referências à obra de boyd e Crawford em relação à proposta de cada artigo, assim como em relação à maneira em que se utilizou a referência durante o texto. Esta classificação não se refere, portanto, à quantidade de vezes em que boyd e Crawford são citadas nos artigos. Estão agrupados na primeira categoria artigos que, de modo geral, referenciam o trabalho entre obras que impactaram os estudos sobre Big Data ou mesmo resumem – e seguem – alguma das provocações elencadas pelas autoras – destas, a sexta foi especialmente citada (e.g. KSHETRI, 2014; KENNEDY; MOSS, 2015; THORNHAM; GÓMEZ CRUZ, 2016).

De modo geral, os artigos elencados na primeira categoria se distinguem por citar a obra das autoras em um contexto geral de pesquisa sobre Big Data, maiormente ressaltando-a como uma importante contribuição crítica (e.g. KITCHIN, 2014; DALTON; THATCHER, 2015), uma reação esperada em relação a um texto fundacional como este. O estudo aqui realizado, no entanto, não constatou nestes textos um engajamento dialógico ou crítico da perspectiva de não problematizar os pontos de vistas das autoras, isto é, este cânon é reproduzido aqui no sentido de filiação autoral às suas propostas, inadvertidamente ou não. Além de contribuir para a proliferação das visões de boyd e Crawford, o modo em que tais

referências se colocam, informando a pesquisa em Big Data na BD&S, tende a circunvenção os problemas de pesquisa da área às provocações elaboradas pelas autoras, o que tende a reduzir o escopo das análises, como discutido anteriormente, a determinados tipos de questões, que, embora importantes, poderiam ser ampliadas.

Na segunda categoria, encontram-se agrupados artigos que dialogam de modo explícito em suas pesquisas com as provocações de boyd e Crawford, apontando possibilidades e limitações destas *em relação à própria pesquisa apresentada ali*, possibilitando leituras atualizadas a partir das feitas pelas autoras. Em uma destas contribuições, Baack (2015) compara o entendimento de ‘dados brutos’ [*raw data*] em discussões acadêmicas como as de boyd e Crawford com a maneira como ativistas da *Open Knowledge Foundation* os compreendem. A partir do que vê como um contraste entre esses entendimentos de dados, e o autor tem em mente especificamente o de boyd e Crawford, Baack apresenta possíveis contribuições epistemológicas de grupos ativistas para os debates sobre Big Data na contemporaneidade. Aqui, as autoras não são apenas citadas ou têm suas ideias resumidas, porém veem sua proposta sendo apresentada *em relação a outras*.

O mesmo ocorre em Lippert (2016), que, mesmo reforçando os argumentos e provocações expostos por boyd e Crawford, traça um diálogo entre estes e seu estudo etnográfico acerca do uso e captação de dados ambientais no contexto de uma empresa da área de meio-ambiente, atualizando as questões das autoras para os estudos em dados ambientais. Já Hong (2016) parte das provocações de boyd e Crawford, em especial a sexta, para realçar que as divisões criadas em torno da exploração e interpretação dos dados são, acima de tudo, divisões de gênero, reforçando um machismo estrutural alimentado pela suposta habilidade técnica do ‘homem’ em interpretar os dados. Neste caso, os achados de Hong ampliam o escopo da provocação de boyd e Crawford para uma questão que não aparecia no texto das autoras.

Já na terceira categoria se encontram trabalhos que oferecem críticas ao trabalho de boyd e Crawford, seja se posicionando em uma posição alheia à das autoras por completo ou, mesmo que seguindo a crítica proposta por elas, apontando limitações das provocações de ambas. O trabalho de Symons e Alvarado (2016), por exemplo, segue a proposta crítica dos dados de boyd e Crawford, citando-as constantemente no decorrer do artigo. Todavia, fornece mais do que um resumo do artigo delas, construindo um diálogo entre a proposta

crítica das autoras e filósofos da ciência. Isso permitiu aos autores notar, inclusive, uma tendência de boyd, Crawford e Kitchin, por exemplo, em apontar os limites dos dados sempre em termos do uso ‘errado’, do funcionamento impróprio das redes, ao invés de serem ligados ao próprio funcionamento dos sistemas computacionais:

[Eles pensam] que o problema tem mais a ver com o viés de amostragem que se origina na implantação de tecnologias, nos métodos de coleção de dados, e na ontologia de dados empregada no processo. Em outras palavras, os problemas acima têm a ver com limitações subjetivas e vieses dos agentes que conduzem o processo (SYMONS; ALVARADO, 2016, p. 5).

A mesma postura é adotada por Perng et al (2016). Os autores argumentam que as provocações de boyd e Crawford são produtivas, porém lançam um desafio a pesquisadores que adotam uma perspectiva crítica sobre os dados, como a de boyd e Crawford, a que apresentem um relato menos generalizado sobre os dados e “pensem cuidadosamente sobre como as mesmas bases de dados são repensadas e retrabalhadas de maneiras diversas e inesperadas, contendo múltiplas vidas paralelas e inter cruzadas e podendo adquirir múltiplas encarnações” (PERNG et al, 2016, p. 9). Para eles, outras perguntas precisam ser feitas além das que até ali estavam a circular nas pesquisas sobre Big Data. O argumento aqui é feito com base em uma pesquisa acerca dos dados de localização e de como expectativas ‘não intencionais’ de experimentos computacionais em relação a eles podem não se conformar a relatos de uma ontologia ‘única’ do Big Data.

Os textos de Schroeder (2014) e Blok & Pederson (2014) vêm a proposta de boyd e Crawford sob um olhar mais pessimista. Blok e Pederson (2014), por exemplo, chegam a reconhecer a importância dos estudos de boyd e Crawford para o campo crítico de pesquisa em Big Data, mas apontam para o que veem como um risco de reprodução, nos escritos delas e de outros autores vinculados aos Estudos Críticos sobre os Dados, de uma “bifurcação problemática entre a evidência quantitativa ‘dura’, em necessidade de interpretação mais atenta, e o dado qualitativo ‘leve’ embutido com ‘o significado’ necessário para reparar esta brecha hermenêutica”, reforçando o que chamam de exclusão mútua entre métodos de ‘big’ e ‘small data’ (BLOK; PEDERSON, 2014, p. 3). Da mesma forma, Schroeder (2014), abordando as possibilidades e perigos da análise de redes sociais e de seu uso de Big Data, lança dúvidas sobre o questionamento das quatro primeiras provocações de

boyd e Crawford a uma certa falta de ‘cientificismo’ dos estudos em Big Data e de suas reivindicações por objetividade. Para Schroeder (2014, p. 4), “é precisamente porque o conhecimento é mais científico, objetivo e poderoso que ele permite mais manipulação”.

Dos 29 artigos que citam este texto canônico, portanto, apenas sete efetivamente adentram nos pontos de vista apresentados por ele, explorando seus potenciais e limites. A maior parte dos olhares aqui partiu do mesmo local onde boyd e Crawford deixaram os seus. Das contribuições que se engajaram de forma mais ampla com o trabalhos das pesquisadoras, as provocações destas foram ampliadas para contemplar as práticas e epistemologias de movimentos ativistas, além de dados ambientais, questões de gênero, enquanto algumas das críticas feitas às autoras giraram em torno de uma certa tendência de crítica ao sujeito e não ao funcionamento computacional na crítica delas ao Big Data, bem como ao modo em que bifurcam Big e Small Data e em como veem a questão da objetividade dos dados.

3.3 Rob Kitchin

Dentre os mais prolíficos pesquisadores sobre o Big Data, o irlandês Rob Kitchin, professor e pesquisador da *National University of Maynooth* [Irlanda], é um dos nomes amplamente referenciados, como demonstram inclusive outros levantamentos sobre a pesquisa em Big Data (e.g YOUTIE et al, 2017). Apesar de sua formação como geógrafo e estudos sobre o uso de dados na geolocalização, dois de seus principais trabalhos publicados estão direcionados a questões epistemológicas do Big Data. O principal deles é o livro “*The Data Revolution: Big Data, Open Data, Data Infrastructures and their consequences*” (KITCHIN, 2014a), lançado pela SAGE em 2014, cuja ‘versão condensada’ foi publicada no mesmo ano, na primeira edição da revista *Big Data & Society*, sob o título de “*Big Data, new epistemologies and paradigm shifts*” (KITCHIN, 2014b).

Em ambos os trabalhos, Kitchin busca dar sequência à proposta crítica de boyd e Crawford acerca de “uma necessidade urgente de uma reflexão crítica mais ampla por parte da academia sobre as implicações epistemológicas dos desdobramentos da revolução dos dados” (KITCHIN, 2014b, p. 1). No artigo publicado na *BD&S*, Kitchin se apropria da teoria dos paradigmas, de Thomas Kuhn, para apontar para uma transição em voga nos modelos epistemológicos das ciências; transição esta que guiada pelo que chama de ‘revolução dos dados’. Para Kitchin, o movimento atual das ciências têm decisivamente se voltado para o

que chama de ‘ciência orientada pelos dados’, e para mais longe de uma ‘ciência orientada pelo conhecimento’, expressão que utiliza com relação aos processos dedutivos característicos do pensamento causal que constituiu o fundamento clássico da epistemologia (cf. DUTRA, 2010).

Tal movimento se daria de modo bifurcado: de uma parte, “fora da academia, especialmente nos círculos empresariais” (KITCHIN, 2014b, p. 3), haveria uma forte ênfase em um empirismo quase total, onde os dados estão absolutamente livres de vieses humanos, sendo um reflexo transparente e objetivo da realidade. Seria a visão mais utópica de Anderson (2008), Pentland (2014) e outros. Por outro lado, no contexto do labor acadêmico, Kitchin vê um deslocamento em direção a uma “combinação híbrida de abordagens abduativas, indutivas e dedutivas para o avanço da compreensão de um fenômeno” (KITCHIN, 2014b, p. 5), reconhecendo os dados como apresentando uma certa visão da realidade, de maneira contextual e limitada, porém admitindo sua centralidade no processo de validação científica. Kitchin (2014b, p. 6) resume o pensamento que guia tal deslocamento da seguinte forma:

A ciência orientada pelo conhecimento, lançando mão de uma abordagem diretamente dedutiva, tem utilidade particular na compreensão do mundo dentro de condições como dados escassos e computação fraca. Continuar a utilizar tal abordagem, entretanto, face aos avanços tecnológicos e metodológicos que possibilitam empreendimentos com análises de dados muito mais ricas, [...] faz pouco sentido.

O problema, para Kitchin, é que tal epistemologia ainda reserva pouco espaço para lidar com explicações de padrões de dados. “Um mapa linguístico de tweets [...] pode revelar padrões de concentração geográfica de diferentes comunidades étnicas, mas as questões importantes são o que constitui tais concentrações? Por que elas existem? Quais são os processos de formação e reprodução?” (KITCHIN, 2014b, p. 8). Especialmente para os cientistas sociais, argumenta Kitchin, esta deveria ser a grande preocupação à medida em que o movimento em torno de uma ciência dos dados alcança o status de paradigma. Assim como o artigo de Boyd e Crawford, Kitchin (2014b, p. 10) conclui seu artigo fazendo uma convocação geral para uma “reflexão crítica mais ampla das implicações epistemológicas do Big Data e da análise de dados”, uma tarefa que, segundo ele, “mal começou”.

Em seu livro “Data Revolution” (2014a), Kitchin amplia o alcance da ‘revolução dos dados’, contemplando questões como vigilância, capitalismo, ética e movimentos ativistas. Em cada um destes casos, Kitchin ressalta as mesmas problemáticas de fundo: de quem são os dados? Como eles são analisados? Quem os analisa? Big Data é sempre “Better Data”? Se o suposto benefício universal de uma maior eficácia a que supostamente alcança a todos, todos têm o mesmo acesso aos dados? Para ele, a resposta a estas perguntas deveria ser contemplada a partir dos agenciamentos, conceito que toma emprestado dos Estudos Sociais da Ciência e da Tecnologia (cf. LATOUR, 2005), dos dados. Ao ressaltar o caráter ‘agencial’ dos dados, Kitchin (2014a, p. 24) pretende apontar que “dados não existem independentemente de ideias, técnicas, tecnologias, sistemas, pessoas e contextos” os quais operam continuamente e de maneira complexa e diversa no que o dado é.

Para Kitchin, tais agenciamentos são compostos por diversos elementos os quais resume em sistemas de pensamentos, formas de conhecimento, finanças, economia política, governamentalidade e legalidade, materialidades e infraestrutura, práticas, organizações e instituições, subjetividades e comunidades locais, e mercado financeiro. No livro, ele se volta mais especificamente para a infraestrutura dos dados, questões de ordem técnica e para o movimento open data, ressaltando como as estruturas em torno dos dados têm, em geral, seguido uma agenda neoliberal não-emancipatória para cidadãos e comunidade civil. Porém, como aponta o próprio uso do conceito de agenciamento, “a relação entre dados e seus agenciamentos é uma de constituição mútua; [...] estes agenciamentos não são formações estáveis, mas estão em constante mudança e evolução, deixando e adquirindo diferentes peças ao longo do caminho” (MENICELLI, 2015, p. 319).

O apelo final de seu livro é por um aprofundamento em estudos que tracem os arranjos sociotécnicos das mais diferentes estruturas, sejam elas políticas, econômicas, geográficas, entre outras, de agenciamentos de dados, algo que conclui prevendo que ocorreria nos anos posteriores: “não há dúvida de que nos próximos anos haverá uma verdadeira inundação de estudos semelhantes buscando documentar a natureza e implicações dos agenciamentos emergentes de dados” (KITCHIN, 2014a, p. 232). Apesar do impacto das análises de Kitchin para uma crítica do Big Data, a crítica de Symons e Alvarado ao trabalho de boyd e Crawford também alcança a argumentação de Kitchin: sua crítica não alcança a

estrutura do Big Data, mas os usos que os dados recebem. O funcionamento do Big Data em questão ali é o ‘errado’, o enviesado de acordo com certos interesses.

Assim como o texto de boyd e Crawford, a influência dos trabalhos de Kitchin perpassa campos tão diversos quanto educação (WILLIAMSON, 2016), urbanismo (JOSS et al, 2017) e medicina (CIRILLO; VALENCIA, 2019). Não à toa, como apontado no levantamento do capítulo anterior, é o autor mais citado nas primeiras seis edições da BD&S. No entanto, e novamente, interessa aqui a maneira em que os trabalhos de Kitchin informam os materiais da publicação.

Ao todo, 27 artigos utilizaram os trabalhos de Rob Kitchin como referência, incluindo uma autorreferência da parte de Kitchin (KITCHIN, 2014b; LYON, 2014; BEER, 2015; HUC-HEPHER, 2015; DALTON; THATCHER, 2015; BOLIN; SCHWARZ, 2015; ARADAU; BLANKE, 2015; MOHR et al, 2015; PYBUS et al, 2015; WILLIAMSON, 2015; VAN DIJCK; POELL, 2016; PERNG et al, 2016; RIEDER; SIMON, 2016; DALTON et al, 2016; BEER, 2016; VAN DER VLIST, 2016; KITCHIN; MCARDLE, 2016; LEÓN, 2016; COCKAYNE, 2016; BATES et al, 2016; CHRISTIAENS, 2016; CURRIE et al, 2016; SYMONS; ALVARADO, 2016; CANALI, 2016; ILIADIS; RUSSO, 2016; THORNHAM; CRUZ, 2016). Aqui, a mesma classificação do tópico anterior é feita, sob os mesmos princípios de análise, reunindo três categorias de análise: 1) referência instrumental; 2) referência dialógica; e 3) referência crítica. O resultado é apresentado na Tabela 4, abaixo:

CATEGORIAS	ARTIGOS
Referência instrumental	(KITCHIN, 2014b*); (LYON, 2014); (BEER, 2015); (DALTON; THATCHER, 2015); (BOLIN; SCHWARZ, 2015); (ARADAU; BLANKE, 2015); (WAGNER-PACIFICI et al, 2015); (PYBUS et al, 2015); (WILLIAMSON, 2015); (VAN DIJCK; POELL, 2016); (RIEDER; SIMON, 2016); (DALTON et al, 2016); (KITCHIN; MCARDLE, 2016); (LEÓN, 2016); (CHRISTIAENS, 2016); (CURRIE et al, 2016); (THORNHAM; GÓMEZ CRUZ, 2016); (COCKAYNE, 2016); (CANALI, 2016); (CHRISTIAENS, 2016);
Referência dialógica	(HUC-HEPHER, 2015); (BATES et al, 2016); (ILIADIS; RUSSO, 2016); (VAN DER VLIST, 2016); (BEER, 2016);

Referência crítica	(PERNG et al, 2016); (SYMONS; ALVARADO, 2016)
--------------------	---

Tabela 4: Classificação organizada com base nos modos em que os artigos da análise se apropriam da obra de Kitchin. *Como o artigo de Kitchin, publicado na *Big Data & Society*, apresenta um resumo das implicações epistemológicas das questões levantadas no livro, este uso de referências constitui um caso *sui generis*.

A classificação na primeira categoria adota os mesmos critérios expostos no tópico anterior feita com base nas referências à obra de Kitchin em relação à proposta de cada artigo, assim como em relação à maneira em que se utilizou a referência durante o texto, não tomando necessariamente em questão apenas a quantidade numérica de referências a Kitchin nos artigos. Nesta categoria se encontram obras que reconhecem a contribuição de Kitchin e o mencionam como importante referência nos estudos em Big Data, artigos que citam vários de seus conceitos (e.g. CHRISTIAENS, 2016) ou apontam questões particulares dentro da obra de Kitchin (e.g. CURRIE et al, 2016). Entre estas estão, por exemplo, assuntos como vigilância (LYON, 2014), dados e assimetria de poder (WILLIAMSON, 2015), estudos críticos sobre os dados (DALTON et al, 2016; CHRISTIAENS, 2016), epistemologia dos dados (CANALI, 2016), bem como suas definições de Big Data (COCKAYNE, 2016).

No mais, os pontos de vista apresentados por Kitchin neste texto fundacional são pouco problematizados nos artigos, isto é, não são objeto de análise nestes textos. Na segunda categoria, cinco artigos destacaram-se na análise por incorporarem em seus próprios problemas de pesquisa e nas discussões de seus resultados a proposta de leitura de Kitchin, apresentando possíveis desdobramentos desta leitura em questões que ultrapassam o escopo da análise primária de Kitchin. O artigo de Iliadis e Russo (2016) menciona Kitchin no contexto de várias das reflexões dos Estudos Críticos sobre os Dados, em especial a crítica de uma abordagem positivista dos dados. Em contraste com outras posições críticas sobre os dados, Iliadis e Russo apontam a proposta de Kitchin como oferecendo um melhor fundamento de pesquisa para os estudos críticos sobre os dados, o que deveria ser os “agenciamentos sociotécnicos de dados” (ILIADIS; RUSSO, 2016, p. 3).

No caso de Beer (2016), Kitchin é assumido abertamente como o ponto de partida para a proposta do artigo, uma tentativa de desconstruir a racionalização neoliberal do Big Data. Ele chega mesmo a se referir às reflexões de Kitchin como constituindo uma “visão geral autorizada do Big Data” (BEER, 2016, p. 10). Beer (2016) busca responder o chamado

de Kitchin (2014a, p. 216) para um mapeamento mais detalhado dos enquadramentos [framing] políticos e econômicos sob os quais os regimes discursivos dos dados são construídos. O mesmo acontece com o trabalho de Fernando Van der Vlist (2016), o qual é uma resposta ao chamado de Kitchin por um “engajamento mais crítico e filosófico, bem como empreitadas empíricas detalhadas acerca da formação, funcionamento e sustentação de agenciamentos de dados” (VAN DER VLIST, 2016, p. 11). Sua análise empírica sobre práticas de Big Data no Facebook conclui por oferecer novos olhares sobre as implicações de críticas levantadas por Kitchin em uma rede social como o Facebook.

Bates et al (2016) também se posicionam dentro da linha crítica de Kitchin, porém visando entrar abaixo da superfície a fim de entender “como práticas locais e situadas de Big Data estruturam como os dados trabalham no mundo” (BATES et al, 2016, p. 2). As autoras concluem, a partir dos estudos de casos apresentados no artigo envolvendo a constituição sócio material dos fluxos de dados, que é preciso reconhecer que os esforços de Kitchin necessitam se estender para abordagens empíricas da ‘vida dos dados’ (BATES et al, 2016, p. 1). Da lista acima, o artigo de Huc-Hepher (2015), uma análise etnosemiótica de arquivos da Web, é o que mais cita a obra de Kitchin. Várias de suas definições e discussões sobre metadados, agência, o processo de aglomeração de diferentes cadeias de dados em uma só, por parte de organizações, os repositórios digitais confiáveis [TDR, sigla em inglês], bem como a introdução geral do livro ao fenômeno do Big Data, citadas diretamente várias vezes, fornecem fundamento direto para a discussão da autora sobre processos dinâmicos de arquivamento na web. Seu estudo é direcionado a confirmar previsões já colocadas por Kitchin acerca do papel dos metadados na compreensão da formação de uma base de dados. Os artigos elencados acima apresentam, portanto, tentativas de expandir a proposta de Kitchin, contemplando objetos de análise ou questões não apresentadas na obra do pesquisador irlandês.

Por fim, o levantamento constatou a presença de dois artigos que ofereceram críticas a parte das colocações de Kitchin. O artigo de Perng et al (2016) repete, em relação a Kitchin, a avaliação feita com relação ao trabalho de boyd e Crawford: ambos traçam, na opinião dos autores, um retrato muito generalizado sobre os dados, tratando-os de forma universalizante. A crítica mais elaborada, no entanto, vem de Symons e Alvarado (2016). Os autores traçam um resumo dos principais argumentos do artigo de Kitchin, publicado na

BD&S, o qual colocam como um dos textos fundamentais dos Estudos Críticos sobre os Dados ao lado do artigo de boyd e Crawford. O foco da apropriação de Kitchin no artigo gira em torno dos aspectos epistemológicos da crítica do pesquisador irlandês ao Big Data, tendo em mente em especial as três perspectivas, na visão de Kitchin, que orientam as investigações dos cientistas sobre os dados: paradigmática, empírica e guiada pelos dados. Mais do que sintetizar as ideias de Kitchin, os autores levantam uma crítica “amigável” (SYMONS; ALVARADO, 2016, p. 2) ao modo em que Kitchin e outros teóricos dos estudos críticos sobre os dados utilizam referências da filosofia da ciência.

Segundo Symons e Alvarado, ao construir sua crítica ao Big Data a partir do modelo kuhniano (KUHN, 1962) de história e ao que chamam de uma noção relativista de um certo ramo da filosofia da ciência, Kitchin coloca a natureza da questão epistemológica sobre os dados em limitações subjetivistas por parte dos agentes, e não à própria natureza dos sistemas de dados. Deste modo, Kitchin, boyd e Crawford, entre outros, tenderiam a atribuir, em suas reflexões, os problemas do Big Data não aos próprios dados, mas aos vieses que envolvem sua coleta e manipulação. Contra isso, argumentam que

os sistemas de Big Data também envolvem riscos que não ocorrem apenas por efeitos de design ou escolhas políticas, mas da própria natureza dos sistemas de software. [...] Nós pensamos que os principais problemas epistemológicos com relação aos sistemas largos e complexos de dados ocorrem devido às características próprias de sistemas de software e não de meras limitações contingenciais da parte dos agentes (SYMONS; ALVARADO, 2016, p. 6)

Para os autores, tal fato se deve à uma necessidade de atualização da reflexão epistemológica dos Estudos Críticos sobre os Dados, para que contemple literaturas mais recentes que exponham as características dos métodos computacionais e algoritmos. Eles usam o exemplo de autores da filosofia da ciência da computação e da filosofia da ciência.

No quadro geral, a tendência apresentada aqui é a mesma com relação ao texto de boyd e Crawford e aponta para uma consolidação dos pontos de vista apresentados nestes cânones, que são pouco problematizados até aqui.

3.4. Viktor Mayer-Schönberger e Kenneth Cukier

A terceira obra mais referenciada no recorte investigado da BD&S é a mais popular e menos ‘acadêmica’ dentre as principais. O livro “*Big Data: a revolution that will transform how we live, work, and think*”, do austríaco Viktor Mayer-Schönberger, professor no *Oxford Internet Institute* [Oxford University - Reino Unido] e do estadunidense Kenneth Cukier, jornalista e editor-chefe do jornal *The Economist*, chegou a figurar entre os mais vendidos na lista do jornal *New York Times*. A obra publicada em 2013 é considerada por comentaristas como sendo uma introdução ao Big Data para públicos não-familiarizados(as) com o assunto (e.g. TEMPINI, 2013; ARMSTRONG, 2014).

Mayer-Schönberger e Cukier (2013) pintam, ao longo das mais de 200 páginas do trabalho, um cenário ubíquo de grandes quantidades de dados transformando cada microespaço da sociedade atual. A linguagem para a ‘revolução’ em andamento movimenta constantemente ordens de grandeza: mais dados, mais correlações, mais volume, mais tecnologias. Tais grandezas, para os autores, estão nos ajudando a “fazer sentido do nosso mundo em maneiras que estamos apenas começando a apreciar” (MAYER-SCHÖNBERGER; CUKIER, 2013, p. 10).

Os autores argumentam que o Big Data fomenta três ‘viradas’ decisivas na maneira como se baseia o funcionamento e o próprio conceito de sociedade. Essas viradas se caracterizariam por uma preferência crescente i) pela análise de quantidades grandes de dados ao invés de bases pequenas de dados; ii) por uma quantidade cada vez maior de dados, mesmo que ‘bagunçados’, ao invés de uma procura por informações que sejam “tão acuradas quanto possam ser” (MAYER-SCHÖNBERGER; CUKIER, 2013, p. 24); e iii) por análises correlativas em detrimento da busca por causas ou razões para os dados. Neste último ponto, Mayer-Schönberger e Cukier (2013, p. 35, grifo nosso), após demonstrar a efetividade da correlação para usuários e empresários da Amazon, argumentam que é acima de tudo o pragmatismo que faz do Big Data um projeto indispensável para nosso tempo:

Conhecer o porquê pode ser agradável, mas não é importante para estimular vendas. Saber o quê, entretanto, impulsiona cliques. [...]. O Big Data mostra que existe uma abordagem diferente, em certa maneira mais pragmática. Os inovadores sistemas de recomendação da Amazon apresentaram correlações valiosas sem que se soubesse as causas subjacentes para elas. *Saber o quê, não o porquê, é o suficiente.*

É por isso que, para os autores, é preciso datificar. É necessário converter cada vez mais elementos do mundo - entenda-se realidade aqui - em dados. Para eles, a “datificação representa um enriquecimento essencial na compreensão humana”, uma descoberta de que, no fundo, o universo é compreendido “essencialmente por informação”⁷¹ (MAYER-SCHÖNBERGER; CUKIER, 2013, p. 60). Além do mais, essa é uma atividade que pode ser extremamente lucrativa, de um ponto de vista empresarial. Boa parte dos exemplos dados nos capítulos finais do livro envolvem iniciativas de grandes organizações e pequenas startups que de alguma forma abrangem sistemas de análise e predição baseados em Big Data e que conseguiram prosperar no mercado.

O livro se encerra com algumas preocupações dos autores sobre Big Data e privacidade e o risco de futuros menos abertos graças à presença cada vez mais decisiva da predibilidade a partir de análises de Big Data. Todavia, a nota final é um convite positivo: “devemos usar esta ferramenta com um grau generoso de humildade...e humanidade” (MAYER-SCHÖNBERGER; CUKIER, 2013, p. 117). A obra recebeu, no geral, recomendações positivas com base no modo como os autores introduzem os movimentos da análise de dados e o seu apelo popular (PARISE, 2016). Porém algumas das críticas ao trabalho de Mayer-Schönberger e Cukier não podem ser ignoradas.

Tempini (2013) critica o que vê como uma separação demasiadamente demarcada entre quantitativo e qualitativo nos autores. Para ele, Mayer-Schönberger e Cukier parecem implicar que o momento de início da avaliação qualitativa se dá somente após a apreensão da coisa em um dado quantificado, como se a própria avaliação do que é quantificado não fosse ela mesma qualitativa. Outra de suas apreciações negativas de “Big Data” é ao fato de que o avanço nas analíticas de dados por parte de organizações é sempre tratado como sendo uma vitória dos usuários. Afinal, os sistemas de Big Data jamais são considerados da perspectiva dos atores sociais (TEMPINI, 2013).

Armstrong (2014) também questiona este fato apontando que é preciso pensar o Big Data também a partir do que é excluído pelo sistema. Se pode questionar se o tipo de Big Data descrito, nos termos de Mayer-Schönberger e Cukier, é aquele que é experimentado por usuários brasileiros, bolivianos ou angolanos, ou se o tipo de uso ali descrito se encaixa no próprio contexto empresarial de países latino-americanos, por exemplo. Além das críticas

⁷¹ Aqui, como sinônimo de dado.

elencadas acima, Armstrong (2014, p. 1302) crê que a visão romântica de Big Data de Mayer-Schönberger e Cukier “parece uma exposição antiquada das promessas do Big Data, julga mal o que significa Big Data em um paradigma de conhecimento e carece de um olhar exaustivo sobre o lado negativo” do Big Data.

Apesar de escrito para um público mais amplo e orgulhosamente conter em sua capa a referência ao fato de ser um bestseller do New York Times e do Wall Street Journal⁷², interessa aqui como a obra de Mayer-Schönberger e Cukier informa os artigos contidos na BD&S. Ao todo, no levantamento detalhado no capítulo anterior, 13 artigos fizeram uso extensivo do livro em seus materiais (THRIFT, 2014; LEONELLI, 2014; KITCHIN, 2014b; BURROWS; SAVAGE, 2014; STRUIJS et al, 2014; ZWITTER, 2014; LAGOZE, 2014; GRAY et al, 2015; MADSEN, 2015; HUC-HEPHER, 2015; PYBUS et al, 2015; CANALI, 2016; THORNHAM; CRUZ, 2016). O estudo de cada um destes artigos possibilitou a divisão destes nas seguintes categorias, conforme apontadas na Tabela 5:

CATEGORIAS	ARTIGOS
Referência instrumental	(THRIFT, 2014); (KITCHIN, 2014b); (BURROWS; SAVAGE, 2014); (STRUIJS et al, 2014); (GRAY et al, 2015); (MADSEN, 2015); (THORNHAM; GÓMEZ CRUZ, 2016)
Referência dialógica	(ZWITTER, 2014); (HUC-HEPHER, 2015)
Referência crítica	(LEONELLI, 2014); (LAGOZE, 2014); (PYBUS et al, 2016); (CANALI, 2016)

Tabela 5: Classificação organizada com base nos modos em que os artigos da análise se apropriam da obra de Mayer-Schönberger e Cukier.

Dos 13 artigos que constituem o escopo desta análise, sete foram avaliados como fazendo uma referência instrumental à obra em relação à proposta do artigo. A maior parte destes menciona o livro “Big Data” apenas como uma das obras que deram maior popularidade ao conceito de Big Data (e.g. THRIFT, 2014; BURROWS; SAVAGE, 2014; STRUIJS et al, 2014). Mayer-Schönberger e Cukier também são associados, nos artigos da primeira categoria, a popularização da ideia de uma revolução de dados ou revolução digital

⁷² A obra de Mayer-Schönberger e Cukier foi também traduzida em dezesseis idiomas. Informação disponível em: <https://www.ted.com/speakers/kenneth_cukier> Acesso em: 20/02/2020.

(KITCHIN, 2014b; MADSEN, 2015) e a descrições acerca do volume e da velocidade dos dados na contemporaneidade (GRAY et al, 2015; THORNHAM; GÓMEZ CRUZ, 2016). Nenhum destes trabalhos, todavia, vai além de mencionar o texto canônico e os pontos de vista ali contidos.

Dois dos trabalhos analisados neste tópico apresentaram um diálogo [segunda categoria] entre a pesquisa feita no artigo e as propostas de Mayer-Schönberger e Cukier, ampliando as análises dos autores para questões não contempladas na obra “Big Data”. Zwitter (2014), por exemplo, num trabalho acerca da ética do Big Data, questiona, ao final de sua análise, se as mudanças apontadas por Mayer-Schönberger e Cukier como advindas da ‘revolução dos dados’ seriam refletidas em concepções éticas tradicionais como individualidade, livre-escolha e poder - Zwitter argumenta que sim a partir de sua análise teórica. Já Huc-Hepher (2015), autora de um estudo etnosemiótico de análise de formação de um arquivo na Web, toma por um dos fundamentos de seu estudo a argumentação de Mayer-Schönberger e Cukier acerca do valor da ‘bagunça’ dos dados em uma análise de Big Data *vis-à-vis* modelos que privilegiam “atribuições arbitrárias” (HUC-HEPHER, 2015, p. 10) na classificação e explicação dos dados extraídos. Ela explora o potencial das asserções de Mayer-Schönberger e Cukier sobre o benefício das bagunças de dados e do princípio de correlação em sua análise dos arquivos da Web.

É interessante notar, todavia, que dos 13 artigos, cinco apresentaram perspectivas críticas que refutam Mayer-Schönberger e Cukier, um percentual bastante superior ao recebido pelas obras de Kitchin (2014) e boyd e Crawford (2012), que possuíram mais do que o dobro de participação nos artigos da BD&S. Uma das críticas ao trabalho de Mayer-Schönberger e Cukier (2013) envolve a rejeição de sua visão otimista sobre os potenciais de dados. É nesta linha que Pybus et al (2016) argumentam, por exemplo. Tomando o conceito de datificação como essencial ao argumento do artigo, Pybus et al (2016, p. 3) procuram uma alternativa ao que veem como uma fé cega de Mayer-Schönberger e Cukier na “plenitude dos dados, uma realidade n=tudo onde a análise empírica é transformada em um efeito algorítmico automatizado, já que todas as análises preditivas subsequentes se tornam fatos”. Isso valeria também contra uma visão de dados focada no ganho econômico e que não questionaria quem e para quais fins o processo de datificação é levado a cabo.

Outro questionamento apresentado nos artigos analisados neste tópico é à posição ultra-empirista de Mayer-Schönberger e Cukier, que veem a ciência de dados eliminando, ou, no mínimo, tornando obsoleta, a necessidade de formulação de hipóteses e a importância do conhecimento causal. A proposta de Canali (2016) em seu artigo é, de fato, toda uma crítica à essa noção de Mayer-Schönberger e Cukier. Canali (2016, p. 1) contradiz os autores de “Big Data” em seu estudo de caso sobre o projeto biomédico intitulado EXPOsOMICS demonstrando a centralidade do conhecimento causal “tanto como uma fonte para lidar com a complexidade e como uma saída para se atingir as metas do projeto”.

As críticas de Leonelli (2014) e Lagoze (2014) ao trabalho de Mayer-Schönberger e Cukier, todavia, foram as mais extensas. Lagoze (2014) fornece uma crítica, em especial, ao modelo $n = tudo$ ⁷³, traçado pelos autores. O pesquisador defende que a perspectiva de Mayer-Schönberger e Cukier não é adequada para a reflexão de como o Big Data tem se apresentado e advoga em prol de uma perspectiva sociotécnica que privilegie a investigação das “facetas sociais, culturais, históricas e técnicas [do Big Data] e as interações e tensões entre essas facetas que estabelecem coletivamente o impacto do Big Data na ciência e sua possível transformação” (LAGOZE, 2014, p. 2). Ele termina por argumentar que o falatório dos autores talvez faça mais sentido para a indústria e o mundo dos negócios, não para a ciência e a epistemologia.

Já Leonelli (2014), desde o início de seu artigo, deixa claro que ela toma por ponto de partida para seu escrutínio as características centrais do Big Data conforme sumarizadas por Mayer-Schönberger e Cukier e avalia a plausibilidade de tais ideias “no caso da pesquisa em biologia” (LEONELLI, 2014, p. 2). Leonelli dialoga, em especial, com três aspectos epistemológicos do Big Data conforme apontados por Mayer-Schönberger e Cukier, sendo eles a: i) abrangência dos dados, que poderia permitir aos cientistas ter uma perspectiva o mais próxima do total possível de um fenômeno a partir dos dados; ii) a ‘bagunça’ dos dados, que faria a ciência migrar de uma certa ‘paixão’ ocidental por entender as razões para um fenômeno em direção a um pragmatismo e senso de direção geral providos pelos dados;

⁷³ Mayer-Schönberger e Cukier argumentam, ao tratar da capacidade que maiores quantidades de dados trazem para análises científicas e usos empresariais, que devemos abandonar os modelos baseados em amostras em prol da coleta exaustiva de todos os dados. Para eles, quanto mais dados, melhor. “Usar uma abordagem de $n=tudo$ significa cavar fundo nos dados; as amostras não permitem que isso seja feito tão bem quanto” (MAYER-SCHÖNBERGER; CUKIER, 2013, p. 21).

e iii) as possibilidades de correlação, que permitiriam uma explicação melhor das interações sociais do que os modelos epistemológicos causais de outrora.

Através de um estudo dos modos em que dados biológicos “viajam em diferentes contextos” e bases de dados (LEONELLI, 2014, p. 3), Leonelli questiona estas definições feitas por Mayer-Schönberger e Cukier. A autora enxerga três processos que precisam ser considerados quando avaliamos o impacto dos dados na ciência biológica: i) descontextualização, isto é, um processo anterior à inclusão dos dados nas bases e que envolve a conversão do dado extraído em um contexto particular para seu uso em diferentes contextos e bases de dados, ii) re-contextualização, em outras palavras, o processo de “adoção do dado previamente extraído por um novo contexto de pesquisa, no qual pode ser integrado com outros dados e contribuir para o reconhecimento de novas correlações” (LEONELLI, 2014, p. 4), e iii) reuso, onde teoricamente os dados cumpririam seu propósito contributor para as ciências. A partir de suas análises de como circulam estes dados na biologia, Leonelli demonstra que, em nenhum momento, os dados se apresentam como exaustivos ou todo-abrangentes, como defendem Mayer-Schönberger e Cukier, mas são desde já “fenômenos altamente seletivos”, materiais e contribuições que circulam às custas “da maior parte do trabalho em biologia” (LEONELLI, 2014, p. 7), uma ‘exclusão’ esta que ocorre mesmo à parte da influência de escolhas metodológicas por parte de cientistas.

Do mesmo modo, Leonelli aponta, em seu estudo sobre o movimento dos dados biológicos nas diversas bases de dados do campo, que, *contra* Mayer-Schönberger e Cukier, a tendência do Big Data é por uma ampliação de preconceitos e imprecisões, e não por sua redução. A visão dos autores de que o Big Data tende a se autocorriger e eliminar dados imprecisos é, na visão de Leonelli, motivada pela falha em atentar para os “desequilíbrios inerentes nos tipos e fontes de dados organizados” (LEONELLI, 2014, p. 7), neste caso, nos dados biológicos. Isso também valeria para a ideia de que o advento do Big Data diminui o valor das medidas de eficácia de um estudo. Seu estudo de caso, portanto, não apenas dialoga com o livro de Mayer-Schönberger e Cukier, como também avalia sua plausibilidade ante a sua área de pesquisa. Ela argumenta que pelo menos para o campo da biologia:

[Suas] considerações não tornam os argumentos de Mayer-Schönberger e Cukier sobre o poder do Big Data completamente implausíveis, mas elas certamente diminuem a ideia de que o Big Data está revolucionando a

pesquisa biológica. A disponibilidade de grandes bases de dados está certamente fazendo uma diferença, [...] mas, como argumentei anteriormente, ter muitos dados não é o mesmo de ter todos eles, e cultivar tal ilusão de completude é uma estratégia extremamente arriscada e potencialmente enganosa dentro da biologia (LEONELLI, 2014, p. 7)

Leonelli conclui seu artigo apresentando maneiras distintas em que o Big Data pode contribuir para a biologia e o modo como a especificidade das pesquisas biológicas dá contornos únicos para o modo em que o Big Data é utilizado aí. Com base nesta especificidade, Leonelli (2014, p. 9) finaliza: “Sendo ou não plausível em sua implementação, a epistemologia do Big Data delineada por Mayer-Schönberger e Cukier é, portanto, improvável que se prove atraente para biólogos”.

Percebe-se na análise deste tópico uma disposição maior dos pesquisadores em dialogar criticamente com um texto que, por linguagem e público, se dirige ao cenário das corporações. Sendo assim, suas visões tendem a não receber o mesmo tratamento das de Kitchin ou boyd e Crawford dentro das comunidades científicas envolvidas na BD&S, que se propõe, ao menos, a ser um veículo de comunicação que permita visões antagonistas às do mercado.

3.5. Mike Savage e Roger Burrows

O último texto a ser analisado neste capítulo e o quarto mais referenciado da Big Data & Society no período de 2014-2016 é o artigo “*The Coming Crisis of Empirical Sociology*” [A Crise iminente da Sociologia Empírica], de autoria de Mike Savage, sociólogo e professor do Departamento de Sociologia da *London School of Economics and Political Science*, e Roger Burrows, também sociólogo e professor na Escola de Arquitetura [*School of Architecture, Planning and Landscape*] da *Newcastle University*, ambas no Reino Unido.

O artigo é amplamente considerado um clássico dentro da Sociologia Empírica, tendo inspirado inúmeras respostas e mesmo debates mais amplos no interior deste campo (e.g. MCKIE; RYAN, 2016; DE BOISE, 2012; WEBBER, 2009; CROMPTON, 2008). O texto de Savage e Burrows é o sexto mais citado da história da publicação britânica

Sociology, uma das principais revistas científicas do campo da sociologia⁷⁴ e que publica artigos há mais de cinco décadas. De fato, o artigo se tornou o mais citado da revista no período que abrange o começo do século 21 até 2014 (BURROWS; SAVAGE, 2014), tendo sido publicado no ano de 2007.

É interessante notar que o impacto do artigo de Savage e Burrows é mais acentuado em um período relativamente distante do de sua publicação inicial, como se pode ver na Figura 11 abaixo:

Referências ao artigo de Savage e Burrows desde sua publicação



Figura 11: Número de referências ao artigo “The Coming Crisis of Empirical Sociology” em artigos registrados na base de dados do Web of Science. Dados atualizados até 27/05/2019.

⁷⁴ Na última avaliação do JCR - Journal Citation Reports, a revista científica ficou em décimo oitavo lugar entre todas as publicações do campo da Sociologia. O periódico é publicado pela Associação Sociológica Britânica (British Sociological Association). Informação disponível em: <http://apps.webofknowledge.com/summary.do?product=WOS&parentProduct=WOS&search_mode=GeneralSearch&qid=14&SID=8AXTwcCKshjdJpkqea5&page=1&action=sort&sortBy=LC.D;PY.D;AU.A.en;SO.A.en;VL.D;PG.A&showFirstPage=1&isCRHidden=false> Acesso em: 20/02/2020.

Tal fato, segundo argumentam os próprios autores, se dá exatamente pela proliferação a partir de 2012 do termo e conceito de Big Data, o que seria um exemplo direto do que Savage e Burrows trabalharam no artigo publicado em 2007 (BURROWS; SAVAGE, 2014).

Em linhas gerais, os sociólogos britânicos propuseram no texto clássico da revista *Sociology* que as propostas metodológicas de análise dos fenômenos sociais de uma certa sociologia empírica clássica, se utilizando de técnicas como questionários e entrevistas em profundidade, estavam com os dias contados com o avanço de formas mais eficientes e abrangentes de reunir dados sobre populações a partir do conjunto de dados formados como subprodutos de transações comerciais e atividades em ambiente digital, geralmente obtidos por empresas. A chegada do que Savage e Burrows (2007, p. 886) chamam, citando Nigel Thrift (2005), de ‘capitalismo cognitivo’ ou ‘capitalismo do conhecimento’⁷⁵ retirou da sociologia e de suas metodologias o prisma de “ponto de passagem obrigatório” para a compreensão da sociedade. A provocação de Savage e Burrows, então, se dirige a sociólogos e pensadores das Ciências Sociais para que estes levem em conta “a proliferação de dados coletados por outros, os quais nós [sociólogos] temos amplamente ignorado” (SAVAGE; BURROWS, 2007, p. 895).

Parte da crise iminente da sociologia empírica, para Savage e Burrows, vem da efetividade de análises sociais por parte de instituições não-acadêmicas, vinculadas ao ambiente empresarial. “Em essência, um mundo paralelo e grandemente desconhecido (para pesquisadores da sociologia) de ‘sociologia comercial’ estava a se desdobrar, ao qual certamente não faltava sofisticação” (SAVAGE; BURROWS, 2007, p. 887). Tais análises lançariam sombra especialmente sobre métodos de *survey*, vistos como incômodos para os participantes, reducionistas em seu escopo e excessivamente dependentes de princípios de inferência, e de entrevista em profundidade, entendida como cada vez mais superada por

⁷⁵ Thrift popularizou este termo em sua obra “Knowing Capitalism”, publicada em 2005. Em linhas gerais, com a adoção deste conceito, Thrift preconizou uma mudança na orientação dos processos de comodificação no capitalismo contemporâneo, mudança esta que se direciona para uma apropriação de capital obtido através da coleta de dados sobre comportamentos e preferências de usuários da web. Isto é, um câmbio de um capitalismo do trabalho ou da produção para um capitalismo do conhecimento sobre os usuários/consumidores. Este mesmo movimento é identificado em obras mais antigas, como o “*In the age of the smart machine*”, de Shoshana Zuboff (1988).

sistemas de dados geográficos, por exemplo, que proveriam uma visão mais profunda da particularidade de visões de mundo específicas do que a obtida pelas entrevistas.

Referenciando Bruno Latour (2005) e autores ligados aos Estudos Sociais da Ciência e da Tecnologia, Savage e Burrows (2007, p. 896) convocam a sociologia a abraçar processos de descrição e classificação, em detrimento do interesse na busca de causas e princípios para processos sociais, o que veem como um resquício da sociologia clássica:

Tal chamado em direção a uma sociologia descritiva não envolve uma confiança exclusiva em narrativas, mas procura conectar narrativas, números e imagens em maneiras que se engajam com, e criticam, os tipos de análises transacionais rotineiras que se proliferam atualmente. Ao invés de buscarmos refúgio em nossos debates internos, isso envolveria lançar nossas redes mais longe, se engajando criticamente com as extensas fontes de dados que agora existem e, não menos importante, advogando pelo acesso a tais dados onde eles não estão disponíveis atualmente.

É preciso notar, ao analisarmos a argumentação de Savage e Burrows, que os autores dissertam a partir de e se dirigem de modo mais específico à sociologia britânica (WILLIAMS et al, 2008), que assumiu, para os autores, um caráter por extremo disciplinar na segunda metade do século 20. Sua preocupação reside em convocar sociólogos a salvaguardarem a relevância da pesquisa empírica sociológica em meio a uma certa dificuldade da sociologia, em especial britânica, em ‘surfar’ nos ventos dos novos métodos de análises de dados vindos de grandes corporações.

Vários trabalhos posteriores, influenciados pela provocação de Savage e Burrows, delinearam diferentes maneiras em que dados transacionais e administrativos poderiam ser utilizados na pesquisa sociológica (e.g. WEBBER, 2009; MCKIE; RYAN, 2016). Mesmo os próprios autores do artigo ampliaram o escopo de suas ideias iniciais em reflexões posteriores (SAVAGE; BURROWS, 2009; BURROWS; SAVAGE, 2014), referendando porém a conclusão de seu artigo: “nós mantemos que estávamos corretos no nosso artigo de 2007 ao concluirmos que os sociólogos necessitavam estar preparados para intervir no mundo do Big Data a fim de garantir que tenhamos uma voz neste novo terreno” (BURROWS; SAVAGE, 2014, p. 5).

Entretanto, o tom positivo de Savage e Burrows com relação aos modelos de análise computacional usados por empresas no contexto do ‘capitalismo do conhecimento’ gerou também críticas por parte de comentaristas. Uma delas é a exatamente a de que os

autores britânicos recaem numa leitura histórica tida como inocente por pressupor uma evolução contínua e neutra no fazer científico da sociologia e mesmo nos meandros do capitalismo. Com relação ao uso constante que Savage e Burrows fazem do termo ‘capitalismo cognitivo’, sem contudo se engajar mais profundamente no modo em que este opera, a socióloga Liz Stanley (2008) argumenta que não se pode tomar este termo como um universal ou uma revolução absoluta na forma de se conhecer o real. “Capitalismo não existe em um vácuo e nem ainda provê a totalidade da estrutura social” (STANLEY, 2008, p. 547). Métodos de análise de mercado e o ‘conhecimento’ mais amplo da sociedade, trazidos pelo ‘capitalismo cognitivo, não podem ser medidos meramente pela lógica da eficácia, esta particularmente ligada à consolidação do capitalismo (cf. ELLUL, 1967). Nas palavras de Stanley (2008, p. 546), “o capitalismo sempre conheceu”.

Stanley (2008) critica também uma preocupação de Savage e Burrows que vê como desnecessária, sendo esta acerca da posição da sociologia no ambiente acadêmico. Para ela, a sociologia não estaria simplesmente a ser ‘chutada para escanteio’ por usar métodos ‘desatualizados’. A posição de ‘outro’, estrangeiro, trazendo um olhar distinto do comum sobre a realidade, para ela, sempre fez parte da sociologia. “A sociologia sempre foi o ‘outro’” (STANLEY, 2008, p. 546). Seria exatamente a diversidade de metodologias e respeito à pluralidade de visões que caracterizariam o campo da sociologia no ambiente acadêmico, e não a adoção de metodologias ‘eficazes’ de tamanho único (CROMPTON, 2008).

Considerando o escopo desta dissertação, mais do que uma avaliação crítica do artigo de Savage e Burrows, se busca entender aqui o impacto que este causou na produção sobre Big Data publicada na BD&S. Ao todo, o levantamento detalhado no capítulo anterior encontrou 10 artigos que referenciam mais amplamente o trabalho dos sociólogos britânicos (BURROWS; SAVAGE, 2014; SCHROEDER, 2014; HOUSLEY et al, 2014; BLOK; PEDERSEN, 2014; LYON, 2014; TAYLOR et al, 2014; GRAY et al, 2015; WEBBER et al, 2015; CATON et al, 2015; VAN DER VLIST, 2016).

CATEGORIAS	ARTIGOS
Referência instrumental	(VAN DER VLIST, 2016); (CATON et al, 2015); (GRAY et al, 2015); (TAYLOR et al, 2014); (LYON, 2014); (BLOK; PEDERSEN, 2014)

Referência dialógica	(WEBBER et al, 2015); (HOUSLEY et al, 2014); *(BURROWS; SAVAGE, 2014)
Referência crítica	(SCHROEDER, 2014)

Tabela 6: Classificação organizada com base nos modos em que os artigos da análise se apropriam da obra de Savage e Burrows. *O artigo de Burrows e Savage (2014) é uma releitura do artigo original feita pelos próprios autores.

Dos dez artigos que têm no texto de Savage e Burrows entre as principais referências aos seus trabalhos, seis são classificados aqui como apresentando uma referência instrumental às propostas e ideias dos sociólogos britânicos. Destes, a maior parte menciona o artigo de Savage e Burrows em suas introduções a problemáticas do Big Data, citando-os como importantes referências no estudo do Big Data entre sociólogos e cientistas sociais (e.g. CANTON et al, 2015; LYON, 2014; TAYLOR et al, 2014; GRAY et al, 2015). Algumas das questões específicas do trabalho de Savage e Burrows que foram reverberadas aqui são a proposta por uma ‘política dos métodos’ (VAN DER VLIST, 2016) e o conceito de capitalismo do conhecimento, utilizado por eles a partir de Nigel Thrift (BLOK; PEDERSEN, 2014).

As propostas de Webber et al (2015), Housley et al (2014) e Burrows e Savage (2014) foram classificadas como dialogando com o texto de Savage e Burrows a partir de suas próprias análises, além de apontar desdobramentos e possibilidades de leituras a partir do que propuseram os sociólogos britânicos. No caso de Webber et al (2015), os autores citam fartamente Savage e Burrows, dado que o objeto de análise da pesquisa desenvolvida ali é a própria metodologia utilizada pela comunidade de cientistas sociais e o modo como esta tem recepcionado especialmente abordagens geodemográficas e de taxonomia a partir de amplas bases de dados. Além de proverem um resumo da proposta geral de Savage e Burrows, os autores assumem avançar a proposta destes, no que tange aos motivos para o parco uso de métodos de Big Data por parte de cientistas sociais, especialmente em análises taxonômicas e geodemográficas.

Este artigo desenvolve a tese geral feita por Savage e Burrows de que o Big Data pode fornecer insights sobre o comportamento social que estão além

do alcance dos questionários de pesquisa tradicionais. [...] Embora acreditemos que não é possível fornecer uma explicação teórica para a adoção assimétrica do Big Data como um todo, acreditamos que é possível identificar as razões para a adoção diferencial das taxonomias genéricas que foram desenvolvidas principalmente a partir de fontes de Big Data (WEBBER et al, 2015, p. 14).

Tomando, do mesmo modo, o trabalho de Savage e Burrows como central no artigo deles, Housley et al (2014) têm por proposta da pesquisa e do projeto desenvolvido nela “confrontar os desafios às Ciências Sociais levantados por Burawoy, Savage e Burrows, entre outros” (HOUSLEY et al, 2014, p. 13), apresentando novos meios de coleta e acesso a dados através de iniciativas de código aberto, em especial o trabalho do projeto Cosmos [*Collaborative Online Social Media Observatory*]. Os autores resumem os principais pontos da crise iminente da sociologia empírica, na visão de Savage e Burrows, atentando especialmente para o que estes veem como uma ameaça empresarial que sonda acabar com a relevância do trabalho de cientistas sociais. Housley et al (2014) concordam que a privatização do acesso aos dados pode corroborar para a consolidação de assimetrias entre os diferentes tipos de pesquisa em Big Data, todavia adotam uma posição menos ‘pessimista’ acerca dos desenvolvimentos da pesquisa empírica nas Ciências Sociais. “Quando confrontados com o crescimento de dados disponíveis abertamente a partir de plataformas de mídia social e a dinâmica do movimento de código aberto, essas ameaças à legitimidade da ciência social acadêmica podem não ser tão graves quanto parecem” (HOUSLEY et al, 2014, p. 3). Neste sentido, eles enxergam oportunidades para o florescimento de uma ciência social pública, fruto de uma colaboração direta entre usuários, pesquisadores e cientistas da computação.

Um caso *sui generis* de leitura do artigo de Savage e Burrows é o que parte dos próprios autores. Eles propuseram uma releitura da proposta do artigo publicado na *Sociology*, a partir de comentários recebidos durante os sete anos posteriores à publicação do artigo, bem como ao surgimento de um *boom* no entendimento e nas tratativas sobre Big Data entre 2007 e 2014 [ver capítulo 1]. Apesar de manterem o tom do primeiro artigo, o qual classificam como uma polêmica direcionada aos cientistas sociais, especialmente da Grã-Bretanha e a provocação ao que veem como um ‘rebaixamento’ da sociologia em relação ao que tem sido produzido por cientistas dos dados e outros atores digitais, apontam perspectivas

não contempladas no artigo da *Sociology* a partir de um estudo de caso do maior levantamento de classes sociais da história do Reino Unido, o *Great British Class Survey* (GBCS), publicado em 2013.

Segundo os pesquisadores, este levantamento demonstra que: i) o impacto da pesquisa social não pode ser avaliado apenas em termos de posse de ferramentas de análise de dados, mas também no modo como aquela circula na rede; ii) mais do que uma substituição dos velhos métodos pelos novos, testemunhamos “uma batalha metodológica, cujo resultado ainda é incerto, mas que deve gerar maiores demarcações e diferenciações entre os domínios dos ‘velhos’ e ‘novos’ métodos” (BURROWS; SAVAGE, 2014, p. 4); iii) as pesquisas sociais não apenas representam segmentos da sociedade, como elas mesmas os modulam e ajudam a mobilizá-los; iv) é preciso reconhecer a capacidade de análise de métodos mistos, que se encontram no meio-termo entre Big Data e small data; v) o uso de métodos ortodoxos de análise social, por parte de cientistas sociais, não consegue conceber uma análise tão abrangente quanto a do GBCS. Apesar dos desenvolvimentos, Burrows e Savage mantêm a tese central do artigo anterior: a crise da sociologia empírica está aí, e é maior do que se imaginava.

De modo geral, os três trabalhos apresentados acima se distinguem por tomarem a polêmica de Savage e Burrows como propulsora dos incômodos respondidos nas pesquisas. Como perspectiva crítica, aparece o artigo de Ralph Schroeder (2014), cujo tema aborda a pesquisa sobre redes sociais e Big Data nas Ciências Sociais. Schroeder explicitamente faz menção às propostas de Savage e Burrows, se afastando das premissas destes sobre o suposto poder das entidades privadas em desenvolver pesquisa social de maneiras mais amplas que as feitas pelos cientistas sociais. Em sua opinião, os cientistas sociais têm utilizado amplamente dados providenciados por empresas, com uma taxa considerável de sucesso, além de terem a sua disposição dados que não são proprietários, aqui cita o exemplo do Wikipédia, os quais podem permitir que a pesquisa nas Ciências Sociais floresça. Por fim, ele argumenta que os estudos nas Ciências Sociais tendem a obter avanços cumulativos pois se preocupam menos em aplicar tais dados imediatamente, como feito no contexto empresarial. Apesar de concordar com Savage e Burrows na existência de assimetrias na relação empresa-academia no contexto dos dados, Schroeder (2014) deseja se demorar mais

nas implicações do modo em que os dados transacionais têm sido utilizados, do que em diagnosticar quem está ganhando a ‘quebra de braço’.

3.6 Discussão

Para um texto que se coloca como uma polêmica direcionada aos cientistas sociais, é interessante notar que mais da metade dos artigos analisados aqui não dialoga em suas pesquisas com os pontos de vista levantados por Savage e Burrows. Isso indica, e as análises feitas em tópicos anteriores também, que há uma tendência na BD&S pela manutenção não apenas destes textos canônicos, mas de seus pontos de vista, nas discussões apresentadas ali sobre Big Data e Ciências Sociais. Textos como o de boyd e Crawford, Kitchin, e Savage e Burrows são amplamente reconhecidos como decisivos para qualquer proposta crítica sobre os dados, porém seus pontos são pouco problematizados no material analisado. Somente o trabalho de Mayer-Schönberger e Cukier, reconhecidamente escrito sob uma perspectiva do mundo dos negócios, obteve uma avaliação crítica mais ampla.

Essa constatação é importante, pois identifica uma tendência da pesquisa em Big Data entre cientistas sociais na BD&S em apontar as armas contra o inimigo, isto é, narrativas, ferramentas e leituras do Big Data de cunho claramente neoliberal, mercadológico e instrumentalista, porém negligenciar uma avaliação consistente de seus próprios cânones. Isso se constitui como um solo fértil para a consolidação dos colégios invisíveis, seus funcionamentos e personagens e o apagamento de empreitadas que fogem do discurso canônico nestas comunidades científicas que se reúnem na BD&S (SANTOS, 2019). Tal centralização somente colabora para um euroamericano-centrismo na pesquisa e para que comunidades inteiras que pesquisam Big Data nas Ciências Sociais ignorem a existência de outras realidades não-ocidentais, corroborando ainda mais para estas assimetrias (ARORA, 2016). O modo como o funcionamento da BD&S e de seus colégios invisíveis, analisado no capítulo anterior, e de seus cânones, neste capítulo, acaba por reforçar estas assimetrias, além de ignorar as perspectivas delas será discutido no próximo capítulo.

4. Esgotamento e tensão: o olhar em prol de alternativas ao Sul na BD&S

4.1 Apresentação

Até aqui se apresentou um panorama da pesquisa em Big Data na BD&S, tanto em relação às instituições de origem que fomentam o debate feito ali, principais autores, temas e métodos, quanto acerca do funcionamento das comunidades científicas que ali se articulam e dos textos que informam os esforços de tais grupos. O percurso transitou por um mapeamento analítico realizado em 131 artigos da BD&S, passando por um relato etnográfico em um dos eventos promovidos pela BD&S e uma análise dos textos chamados aqui de canônicos na formação dos grupos que publicam na BD&S. Como visto anteriormente, um dos principais achados deste trabalho se encontra na existência, dentro da BD&S, de suas abordagens e de seus grupos, de uma certa tensão entre, de um lado, um ensejo pelo fomento de perspectivas críticas sobre dados dentro de um contexto de emergência de variados problemas no uso e posse de dados por parte, especialmente, de empresas e governos; e, de outro, na insistência em análises e perspectivas típicas do ambiente liberal norte-americano e europeu, em um diagnóstico feito a partir das próprias epistemologias que se encontram ligadas aos problemas discutidos ali. Uma discussão parcial desta tensão foi tratada no capítulo anterior, na análise de textos como os de Mayer-Schönberger e Cukier (2013), boyd e Crawford (2012) e Rob Kitchin (2014), por exemplo.

Diante disso, este capítulo estende a análise desta tensão existente na literatura e na práxis da BD&S, lançando mão de: i) uma extensão do mapeamento apresentado no capítulo 2, contendo um exame da distribuição étnica e geográfica das vozes e instituições que fomentam a discussão sobre o tema. O mapeamento coletou dados como: composição da Equipe Editorial da publicação, artigos que abordam contextos do Sul Global, referências à pesquisadores latino-americanos e africanos, por exemplo, dentro das pesquisas, bem como à literatura em outros idiomas, colaborações entre atores acadêmicos de diferentes instituições ao redor do globo, entre outros, além de contatos eletrônicos realizados com a equipe editorial da BD&S. ii) revisão bibliográfica de artigos e textos que fornecessem uma vista sobre a questão do Sul Global e de um movimento epistemológico pós-colonial, já que boa parte do olhar alternativo sobre dados buscado na linha editorial da BD&S, em seus

artigos e no evento relatado no capítulo 2 procura um diálogo com o Sul Global na consideração dos problemas relacionados aos dados. O objetivo aqui foi o de entender como assimetrias na produção do conhecimento sobre Big Data se perpetuam e quais as posturas destes esforços teóricos com relação às questões do Sul Global.

Para isto, este capítulo pretende discutir primeiramente o que se entende aqui por Sul Global e em que isso se relaciona com os estudos em Big Data. Neste percurso se pretende explorar as implicações da concentração euroamericana da pesquisa em Big Data, além da reprodução de epistemologias e visões de mundo oriundas de modelos coloniais e capitalistas típicos de contextos do Ocidente, para a própria compreensão do que é o Big Data. Especialmente, porém, objetiva discutir como se pode dar o arraigamento de epistemologias do Norte e de posturas colonialistas mesmo quando existe um esforço para superá-las. Se espera argumentar que a chave da questão não está no mero diagnóstico da necessidade de superação dos modelos liberais euroamericanos, mas nas próprias categorias usadas no diagnóstico. Um movimento que não é de modo algum simples, considerando ainda que a própria emergência do Big Data se deu a partir do contexto neoliberal imperante no Vale do Silício (cf. DIEBOLD, 2012).

O percurso será percorrido com o auxílio de pensadores que há muito alertaram para a contribuição da perpetuação de práticas e pensamentos coloniais para a invisibilidade das peculiaridades das vivências dos povos de margens, como, por exemplo, o sociólogo Boaventura de Sousa Santos, para mencionar um nome dentre os lusófonos. Obviamente, a discussão também não ignora críticas à divisão Norte/Sul global no fazer acadêmico, à sua utilidade conceitual para descrever movimentos de luta contra a hegemonia neoliberal e colonialista nos campos político, social e epistemológico e à presença do ‘Norte’ no Sul e do Sul no ‘Norte’. Tais críticas também serão apresentadas aqui, auxiliando no reforço da escolha epistemológica feita nesta dissertação de utilizar o conceito de Norte/Sul na discussão do levantamento sobre a BD&S (GUTTAL, 2016; KACOWICZ, 2007).

A reflexão aqui realizada se justifica também por se encontrar inserida dentro de um movimento recente iniciado pelo trabalho de Stefania Milan e Emiliano Treré em prol de um “Big Data a partir do Sul”, que busca acender o alerta com relação à necessidade não apenas de maior participação do Sul Global nesta nova área de pesquisa, como também em se traçar teorias do Big Data a partir de contextos do Sul Global. No momento em que esta

agenda passa a encontrar espaço em importantes centros acadêmicos do Norte e do Sul Global, seja em eventos ou em revistas científicas, reunindo contribuições como as de Payal Arora, Sarah Anne Ganter, Félix Ortega, Jan Youtie, entre outros, este capítulo busca contribuir para o fomento de discussões que auxiliem as perspectivas e limites de um “Big Data a partir do Sul”.

Algumas das questões que serão tratadas neste capítulo envolvem: A partir do Sul e de suas múltiplas vivências e contextos, Big Data seria o mesmo dos 3Vs de Laney (2001), como já discutido anteriormente? De que maneira as perspectivas críticas sobre os dados conforme contempladas na BD&S podem servir para a própria reprodução do velho colonialismo que ainda hoje assombra o Sul Global? Há uma indisposição existente nos círculos acadêmicos que publicam na BD&S em dialogar com autores e ideias do Sul Global ao invés de meramente falarem sobre eles? De que formas o achado da tensão Norte/Sul Global existente na BD&S pode contribuir com os desafios para a pesquisa do Big Data, especialmente dentro do contexto das Ciências Sociais?

4.2 O olhar do ‘conquistador’ e as velhas categorias

Em seu clássico “A Crítica à Razão Dualista” (2003[1974]), o sociólogo Francisco de Oliveira, dissertando sobre a especificidade dos meandros da evolução capitalista no Brasil, procurou rebater o que via como um vício do pensamento socioeconômico latino-americano em apontar a realidade brasileira em termos de precariedade social, por um lado, mas de manutenção inquestionada de um estruturalismo do pensamento capitalista tradicional, exemplificada em grandezas duais como ‘progresso’ e ‘subdesenvolvimento’, e ‘modernidade’ e ‘tradicionalismo’.

Enquanto denunciavam as miseráveis condições de vida de grande parte da população latino-americana, seus esquemas teóricos e analíticos prendiam-nos às discussões em torno da relação produto-capital, propensão para poupar ou investir, eficiência marginal do capital, economias de escala, tamanho do mercado, levando-os, sem se darem conta, a construir o estranho mundo da dualidade e a desembocar, a contragosto, na ideologia do círculo vicioso da pobreza (OLIVEIRA, 2003, p. 31).

Tais reproduções de “esquemas aprendidos nas universidades anglo-saxônicas” (OLIVEIRA, 2003, p. 32) afetariam toda uma maneira de se pensar a realidade brasileira, popularizando a concepção do país como sendo uma nação ‘subdesenvolvida’ (sobre críticas semelhantes em outros contextos latino-americanos ver SÁBATO, 1975). Na crítica elucidada acima⁷⁶, o problema não está no diagnóstico da realidade assimétrica envolvendo as instâncias do Norte Global, de um lado, e do Sul Global, de outro - e em especial para ele a América Latina, porém *nas categorias a partir dos quais este diagnóstico é feito*. A partir de registros categóricos euroamericanos, marcados por dualidades como economia-política, natureza-cultura, negro-branco e espírito-corpo⁷⁷, se torna inevitável enxergar a realidade latino-americana sem referência à uma certa incompletude, atraso e vulgaridade do colonizado.

Essa operação da visão do ‘conquistador’ é também encontrada em um relato literário já clássico do escritor Aimé Césaire. Esta é expressa em forma de protesto por Calibã, escravo nativo da ilha quase “esquecido” na peça “A tempestade”, de Shakespeare, e relido por Césaire no poema de mesmo nome publicado três séculos e meio após a peça original, em seu diálogo final com Próspero, seu mestre para quem trabalhou por uma década, um nobre milanês banido para a ilha natal de Calibã em meio ao jogo de poder em sua região:

Próspero, você é um grande ilusionista. Você está familiarizado com o engano e *você mentiu para mim sobre o mundo e sobre mim mesmo, até, por fim, impor sobre mim uma imagem de mim mesmo como sendo um subdesenvolvido, ou em suas palavras, um incapaz. Foi assim que você me forçou a me enxergar!* E eu odeio essa imagem! Ela é falsa! Mas agora eu te conheço, velho câncer, e conheço também a mim mesmo! [...] O velho mundo está caindo aos pedaços! (CÉSAIRE, 1969, p. 88, grifo acrescentado).

Em ambas as leituras, de diferentes tipos e em direção a distintos objetos, o que se coloca em jogo é um modo de ver o mundo sempre a partir dos termos do colonizador europeu, como sendo a única forma de saber possível. Termos que vigoram como modernização, privacidade e capitalismo industrial, para citar alguns, e que podem sequer

⁷⁶ Sintetizada brevemente aqui, sob o risco de incorrer em reducionismo.

⁷⁷ O pensamento ocidental é decididamente marcado pela dualidade, pela redução dos fenômenos à categorias previamente estabelecidas sob os auspícios da “Razão” humana - esta vista como operando de modo autônomo na organização do mundo (cf. SANTOS, 2019).

existir ou não encontrar o mesmo efeito fora da realidade euroamericana (cf. OLIVEIRA, 1984). Essa se torna uma questão séria para a reflexão acerca da pesquisa apresentada na BD&S e de suas propostas mais amplas pois alguns dos objetivos claros do Comitê Editorial da BD&S, bem como dos grupos que o cercam, como mostrou a Data Power Conference, é pensar em alternativas para os modelos liberais euroamericanos no pensamento sobre as implicações dos dados. Um movimento que acompanha uma tendência das Ciências Sociais em torno das chamadas Epistemologias do Sul⁷⁸, que ganham realce especialmente a partir do trabalho do sociólogo português Boaventura de Sousa Santos (SANTOS, 2007). Sob a influência dos estudos decoloniais, que realçaram a *vigência de uma hegemonia de processos de produção de conhecimento eurocêntricos* e, com isso, a opacização de certos agentes e vozes na concepção do que se entende por conhecimento, a proposta de um pensamento do Sul Global é pragmática: “*formular alternativas epistemológicas que possam fortalecer as contendas contra o capitalismo, o colonialismo e o patriarcado*” (SANTOS, 2019, p. 23, grifo acrescentado) e superar modelos que tomem como ponto de partida

prioridade absoluta dada à ciência como conhecimento rigoroso; rigor, entendido como determinação; universalismo, entendido como sendo uma especificidade da modernidade ocidental e referido a qualquer entidade ou condição cuja validade não é dependente de qualquer contexto social, cultural ou político concreto; verdade, entendida como a representação do real; uma distinção entre sujeito e objeto, o que conhece e o que é conhecido; a natureza enquanto *res extensa*; a temporalidade linear; o progresso da ciência por via das disciplinas e da especialização; a neutralidade social e política como condição de objetividade (SANTOS, 2019, p. 24).

Outros aspectos que constituiriam as Epistemologias do Norte, ou o Norte Global seriam a exploração extrativista de valor - seja comercial ou simbólico - e uma concentração hierárquica deste em torno de um centro, deixando os produtores deste valor às margens, tudo sob a bandeira de um progresso e de uma suposta importação neutra de ‘processos tecnológicos eficientes’ do Ocidente em contextos não-Ocidentais (MILAN; TRERÉ, 2019). As consequências de tais epistemologias seriam, para Santos (2019, p. 25) a criação de uma

⁷⁸ É importante lembrar que a discussão sobre Norte/Sul aqui não se refere à famigerada discussão de análises socioeconômicas que trabalham essa assimetria sob o aspecto do desenvolvimento/progresso já desde meados da Guerra Fria (cf. KACOWICZ, 2007). É, inclusive, contra esta perspectiva mais popular de Norte/Sul que parte dos esforços das chamadas Epistemologias do Sul se dirige.

linha abissal, onde “estar do outro lado, do lado colonial, da linha abissal equivale a ser impedido pelo conhecimento dominante de representar o mundo como seu e nos seus próprios termos”.

Diante deste contexto de um movimento em torno de um pensamento ou de epistemologias do Sul Global, parece relativamente fácil entender a importância da expressão do conhecimento próprio de mundo para quem é estranho ao círculo euroamericano, como Calibã ou Francisco de Oliveira. Porém, o que leva uma revista europeia, que circula em comunidades científicas euroamericanas, a tomar como prioridade o incentivo de perspectivas do Sul Global?

Em muitos sentidos, como a própria citação de Boaventura expressa acima, *o olhar para o Sul Global se encontrou*, ao menos na literatura euroamericana, demasiadamente ligado à uma crise da civilização europeia. O grito por um outro mundo, para aludir a Pignarre e Stengers (2011), está diretamente ligado à quebra das aspirações e sonhos que sustentavam o imaginário ocidental. Embora se aponte comumente para os horrores do Holocausto e da Segunda Guerra Mundial como um momento definitivo na derrocada dos sonhos do positivismo e no otimismo da capacidade humana (eg. BAUMAN, 1998), é importante destacar aqui a crise mais específica relacionada à Internet e meios digitais. A virada do milênio marca o auge do otimismo com relação ao futuro das redes digitais e suas potencialidades para a democracia, educação e comunicação à nível global - com talvez o exemplo mais forte desta postura sendo o clássico “Cibercultura”, de Pierre Levy (1999), bem como para os mercados e empreendedorismo, na onda da ‘ideologia californiana’⁷⁹ do Vale do Silício (cf. EVANGELISTA, 2018). Duas ou três décadas depois, no entanto, a realidade se mostrou outra. Os processos opacos de extração de dados comportamentais por parte de empresas a fim de obtenção de valor econômica, escândalos de vazamento de dados pessoais e usos destes em campanhas políticas como o caso da Cambridge Analytica, a disseminação coordenada de notícias mentirosas por meio das redes

⁷⁹ Nome dado a uma postura crescente entre jovens empreendedores das décadas de 70 e 80, situados na região do Vale do Silício, no estado da Califórnia (EUA), que adotava ideais da filosofia objetivista de Ayn Rand e um certo espírito libertariano (CASTELLS, 1996) propondo um rumo tecnológico que agregasse elementos da cultura hacker dos anos 70 a um projeto neoliberal que guiasse o desenvolvimento maquínico, e aqui em especial os computadores, no sentido de fortalecer a competição mercadológica e a tomada privada da rede, diminuir os laços da rede com a esfera pública (governos) e incentivar o uso econômico da rede. Para mais informações, ver o documentário lançado em 2011 sob o título “All Watched Over by Machines of Loving Grace”, de Adam Curtis.

digitais, o crescimento de grupos extremistas na web e o papel das redes sociais na polarização extrema que atinge o Ocidente nos últimos anos colocaram em cheque as utopias de outrora e levantaram perguntas como ‘é possível resgatar a internet?’ (STARR, 2019).

Ante esse contexto, e dentro de um movimento em torno de busca de alternativas, aparece a procura pelas ‘alternativas’ oferecidas pelo Sul Global, historicamente alheio à produção de qualquer dos desenvolvimentos apontados acima⁸⁰. Tal postura de fomento à alternativas para evitar o ‘fim do mundo’ (ocidental) fica evidente nas palavras de Milan e Treré (2017, p. 2) sobre o que seria o Sul Global:

Em primeiro lugar, está o Sul geográfico, isto é, as pessoas, atividades, políticas e tecnologias que emanam literalmente nas margens do mundo [...]. Em segundo lugar, e mais importante, nosso Sul é um lugar de (e um signo que representa) resistência, subversão e criatividade. Podemos encontrar inúmeros ‘Suis’ também no Norte Global, sempre e quando haja gente resistindo à injustiça e lutando por melhores condições de vida contra o iminente ‘capitalismo de dados’.

Nesta postura, a busca pela voz epistêmica apagada pelo legado colonial (REIS; ANDRADE, 2018) se dá *a partir do interesse por alternativas ao esgotamento do Norte global*. O paradigma que orienta a proposta ainda é o modelo europeu, as patologias e os termos do jogo são ocidentais, e permanecem inquestionadas. Afinal, nas palavras do filósofo camaronês Achille Mbembe (2018, p. 29),

só o Ocidente foi capaz de inventar ‘os direitos das gentes’. Só ele conseguiu edificar uma sociedade civil das nações compreendida como um espaço público de reciprocidade do direito. Só ele deu origem a uma ideia de ser humano dotada de direitos civis e políticos [...] e, enquanto tal, interessado em tudo o que é humano. Só ele codificou uma gama de costumes aceitos por diferentes povos. [...] O Resto [...] constituía a manifestação por excelência da existência objetificada.

⁸⁰ Aí se dá a visão de Sul Global como laboratório, um caso clínico que fornece ao mundo [europeu e norte-americano, claro] uma mirada privilegiada dos males e benesses de determinadas iniciativas de cunho sanitário, sociopolítico ou econômico. Sobre um exemplo de tal ‘funcionamento laboratorial’ do Sul, especificamente em relação à população negra a partir do século 19, ver a entrevista do filósofo camaronês Achille Mbembe, cedida à revista *Esprit* em 2006 sob o título “Qu'est-ce que la pensée postcoloniale?” [Que é o pensamento pós-colonial?]. Uma versão em inglês da entrevista se encontra disponível em: <https://www.cairn-int.info/load_pdf.php?ID_ARTICLE=E_ESPRI_0612_0117> Acesso em: 16/03/2020.

Tais apontamentos se revelam importantes no contexto crescente de um movimento em torno de um “Big Data a partir do Sul”⁸¹, liderado pelo coletivo de pesquisa *Dataactive*, vinculado ao Departamento de Estudos em Mídia da Universidade de Amsterdam, nos Países Baixos. Situando-se na esteira de propostas dos Estudos Críticos sobre os Dados⁸² (ILIADIS; RUSSO, 2016), dos Estudos Culturais de Martín-Barbero (1997) e de trabalhos anteriores que identificaram a ausência de interpretações contextuais e de ‘baixo para cima’ do Big Data (COULDRY; POWELL, 2014; ARORA, 2016), a agenda tomou por objetivo “propor uma mudança dos meios às mediações, da datificação para o ativismo de dados” e lançar questionamentos como:

Como seria a datificação de cabeça para baixo? Que questões perguntaríamos? Que conceitos, teorias, métodos adotariamos ou temos que desenvolver? O que perdemos ao ater-nos à(s) perspectiva(s) convencional(is) do Ocidente? (MILAN; TRERÉ, 2017, p. 1)

A grande questão que se coloca aqui é do que se toma por dados ou Big Data *a priori*. Como visto no capítulo introdutório, Big Data, como conceito e prática, tem sua origem nos contextos empresariais do Vale do Silício, com preocupações entrelaçadas às do mercado. Na curta história do Big Data, esta tem sido a perspectiva principal nas abordagens sobre o tema - a partir do mote da ‘solução de problemas’ (cf. KITCHIN, 2014). O perigo que se apresenta é o de pensar em propostas mais justas, éticas ou igualitárias em relação ao Big Data para o Sul ou a partir deste, em detrimento de se *partir de categorias distintas às*

⁸¹ O primeiro grande marco de uma iniciativa transcultural e multidisciplinar em prol de um Big Data a partir do Sul se deu em julho de 2017 na cidade de Cartagena, na Colômbia, com a organização de uma pré-conferência, precedendo a conferência anual da *International Association for Media and Communication Research [IAMCR]*, intitulada “De la dataficación al activismo de datos. Congreso “Big Data from the South”. A conferência de um dia teve a apresentação de dezessete trabalhos, divididos em quatro mesas cujas temáticas abrangeram estudos de caso e experiências, perspectivas críticas, trabalhos conceituais e discussões sobre métodos e epistemologias. Desde então, a iniciativa organizou um workshop e uma mesa na conferência anual de 2018 da *Latin American Studies Association (LASA)*, bem como publicou, em 2019, uma edição especial na revista científica *Television & New Media*, além de lançar a chamada de trabalhos para uma edição especial na revista *Palabra Clave*, a ser publicada em meados de 2021.

⁸² Proposta de análise do Big Data que argumenta que “ao invés de tratarmos o Big Data como um fenômeno unicamente empírico e, portanto, amplamente neutro, [...] Big Data deve ser visto como sendo sempre-já constituído dentro de agenciamentos sociotécnicos mais amplos” (ILIADIS; RUSSO, 2016, p. 1). A revista *BD&S* dedicou uma edição especial para os Estudos Críticos sobre os Dados no segundo semestre de 2016. Disponível em: <<https://journals.sagepub.com/page/bds/collections/critical-data-studies>> Acesso em: 20/02/2020.

do próprio *Big Data* e de seu contexto liberal, numa espécie de subversão epistêmica⁸³. Afinal, se “as definições pertencem àqueles que as definem, e não aos definidos” (MORRISON, 2004, p. 190), a procura por alternativas ao esgotamento dos modelos de dados e vigilância euroamericanos não deveria buscar outras formas de entender o mundo, mesmo que estas subvertam por completo as bases da tradição ocidental? Neste aspecto, mais do que resistência e criatividade (MILAN; TRERÉ, 2019), as perspectivas do Sul Global pedem “um pensamento alternativo de alternativas” (SANTOS, 2019, p. 24).

Mesmo que o chamado em direção ao Sul Global e o uso deste conceito, ajude a identificar assimetrias - o “primeiro passo no sentido de as ultrapassar, quer ao nível epistemológico, quer ao nível político” (SANTOS, 2019, p. 28); articular formas de superá-las e ressaltar geograficamente os locais cujos conhecimentos foram historicamente invisibilizados pelo colonialismo, como mostram as Figuras 12 e 13, é preciso que este se direcione tanto para as especificidades das relações de poder no Sul Global (cf. FIRMINO et al, 2019; MANN; DALY, 2019), quanto para as próprias formulações de mundo que se desenrolam fora da tradição ocidental.

⁸³ Talvez um dos melhores exemplos a serem apontados aqui do que poderia ser um caminho para uma subversão epistêmica está no trabalho de Viveiros de Castro (2015) a partir do que chama de ‘perspectivismo’ no pensamento de sociedades ameríndias. Criticando uma certa tendência, na tradição ocidental de pensamento, de reificar o sujeito e sua percepção como sempre estando ‘ali’ de antemão, Castro vê a necessidade de formulação de novas nomenclaturas para a compreensão das perspectivas ameríndias. O etnógrafo, então, procura partir do modo como os ameríndios constituem o mundo para a formulação de novas maneiras de se pensar a realidade. Para os ameríndios, por exemplo, não havia algo como uma natureza e diversas culturas, porém uma multinaturalidade, onde é a cultura, o espírito, que se mantém o mesmo, enquanto as naturezas e ontologias eram cambiáveis. Nesta perspectiva, o ‘sujeito’ é constituído a partir do modo reflexivo de cada agente, isto é, do seu ponto de vista. “Todos os seres vêem (“representam”) o mundo da mesma maneira — o que muda é o mundo que eles vêem. Os animais impõem as mesmas categorias e valores que os humanos sobre o real: seus mundos, como o nosso, giram em torno da caça e da pesca, da cozinha e das bebidas fermentadas, das primas cruzadas e da guerra, dos ritos de iniciação, dos xamãs, chefes, espíritos... Se a Lua, as cobras e as onças vêem os humanos como tapires ou pecaris, é porque, como nós, elas comem tapires e pecaris, comida própria de gente. Só poderia ser assim, pois, sendo gente em seu próprio departamento, os não-humanos vêem as coisas como “a gente” vê. Mas as coisas que eles vêem são outras. [...] O relativismo (multi)cultural supõe uma diversidade de representações subjetivas e parciais, incidentes sobre uma natureza externa, una e total, indiferente à representação; os ameríndios propõem o oposto: uma unidade representativa ou fenomenológica puramente pronominal, aplicada indiferentemente sobre uma radical diversidade objetiva. Uma só “cultura”, múltiplas “naturezas”” (VIVEIROS DE CASTRO, 1996, p. 127-128)

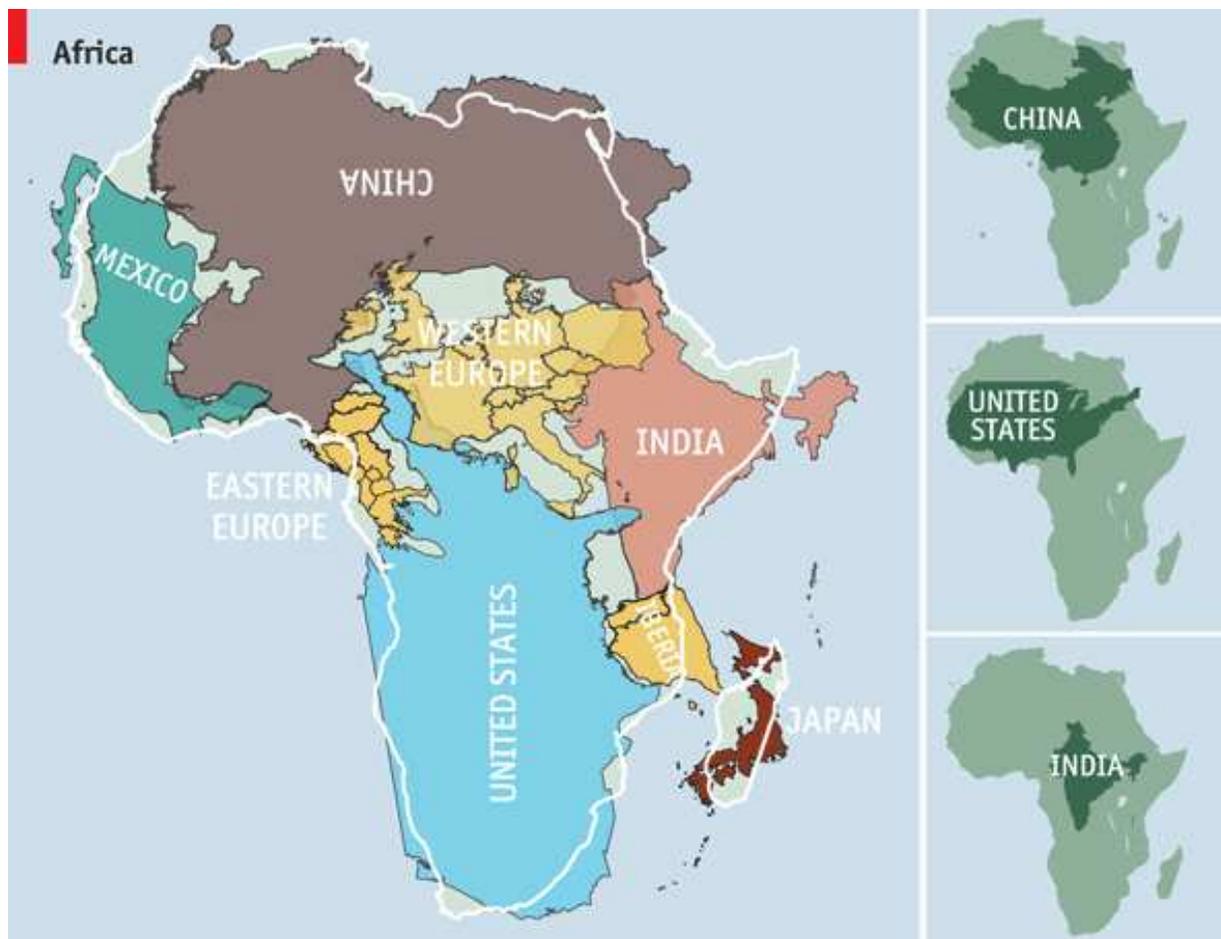


Figura 12: Mapa que critica o exagero dado em mapas influentes como o de Gerardus Mercator, criado no contexto das explorações marítimas de colonizadores europeus. Aqui, por exemplo, a real proporção do continente africano em relação a outras áreas do globo. Disponível em: <<http://bit.ly/tricolorbaiano>> Acesso em: 20/02/2020.



Figura 13: Mapa que critica o exagero dado em mapas influentes como o de Gerardus Mercator, criado no contexto das explorações marítimas de colonizadores europeus, para a posição europeia no globo. Aqui um mapa mundi revertido, jogando com as noções de centramento do Norte, geralmente associado a qualidades de grandeza (cf. NELSON; SIMMONS, 2009). Disponível em: <<http://bit.ly/borabaheaminhap>> Acesso em: 20/02/2020.

No contexto dos estudos sobre dados, o *Indigenous Data Sovereignty* [IDS], já apresentado no capítulo 2 e que surge do contexto de contra-hegemonia de grupos indígenas em relação a práticas digitais do governo australiano em relação a comunidades indígenas como as do povo Yawuru, fornece outro exemplo de procura por novas categorias para a compreensão do impacto do capitalismo de vigilância, para usar um termo popular atualmente (cf. ZUBOFF, 2019), nas comunidades indígenas. Inicialmente partindo de experiências e perspectivas particulares de povos que vivenciaram o tipo de vigilância estatal na Austrália e Nova Zelândia, o IDS se tornou uma coalizão global reunindo comunidades indígenas de todo o globo não somente em prol da luta pelo direito, por parte de povos indígenas, de determinar os meios envolvidos na coleta, interpretação e uso de dados relacionados à e produzidos pelos povos aborígenes (KUKUTAI; TAYLOR, 2016), como também na produção de saberes onde categorias ocidentais como ‘sujeito’ não existem (MANN; DALY, 2019).

Tendo essa discussão em mente, é preciso ainda avaliar em um escopo mais amplo as implicações de como a BD&S se situa em tal movimento em direção ao Sul Global. De que forma a tensão entre o diagnóstico da falência de modelos ocidentais somada à busca por alternativas alheias às do Ocidente e o vínculo persistente à categorias e pensamentos provenientes do contexto euroamericano se manifesta nos temas dos artigos da BD&S e na composição de seu Comitê Editorial, por exemplo?

4.3 BD&S e o olhar para o Sul

Como parte do levantamento realizado na BD&S, buscou-se investigar dados relativos ao tratamento que atores, iniciativas e pesquisas do Sul Global recebem na publicação. O mapeamento identificou desde já uma presença parca de pesquisadores afiliados a instituições fora do eixo da América do Norte e Europa no comitê editorial da publicação, como apontado na Figura 14. Um fato que é visto também em outras áreas de interesse das Ciências Sociais (e.g GANTER; ORTEGA, 2019).



Figura 14: Distribuição geográfica de membros do Comitê Editorial por filiação institucional, em um total de 72 pessoas, entre equipe editorial, supervisores e membros conselheiros. Do total, 17 são afiliados à instituições estadunidenses e 27 à instituições britânicas. Informação atualizada até fevereiro de 2020. Informações disponíveis em: <<https://us.sagepub.com/en-us/sam/journal/big-data-society>> Acesso em: 20/02/2020

A mesma disparidade se vê em relação aos países das instituições de origem dos autores que publicaram na BD&S no período, conforme ilustra a Tabela 7. Instituições estadunidenses, por meio de seus pesquisadores, encabeçam 45 dos 131 artigos publicados pela BD&S no período, o que configura mais de um terço das pesquisas (34,35%). Somadas com as instituições britânicas, que representam um total de 39 artigos (29,77%), a fatia britânico-americana alcança pouco mais de 64% do material produzido na BD&S. Enquanto apenas quatro trabalhos partem de pesquisadores de instituições fora do eixo América do Norte-Europa e Austrália⁸⁴.

PAÍSES	NÚMERO DE PESQUISAS
EUA	45

⁸⁴ A situação ímpar da Austrália e da Nova Zelândia nas discussões pós-coloniais, e sua posição histórica ao lado de nações tradicionalmente capitalistas do Norte, pode ser vista em Mann & Daly (2019).

Reino Unido	39
Países Baixos	16
Canadá	12
Alemanha	8
Irlanda	5
Dinamarca	4
Austrália	2
Bélgica	2
França	2
Suécia	2
Áustria	1
Croácia	1
Espanha	1
Finlândia	1
Itália	1
Suíça	1
Emirados Árabes Unidos	1
Singapura	1

Egito	1
Japão	1

Tabela 7: Distribuição geográfica dos autores dos artigos das seis primeiras edições da BD&S por filiação institucional.

A análise também contemplou a quantidade de artigos que tomavam por seu objeto de estudo contextos vividos na América Latina e na África. Os resultados ressaltam ainda mais a tensão existente entre a busca por perspectivas alheias ao Ocidente, como visto no evento promovido pela BD&S em Bremen, e as discussões efetivas que tomam parte no meio da publicação. Em conformidade com os resultados obtidos anteriormente, como mostra a Figura 15, dos 131 artigos publicados no período apenas três (2,2%) tomam como objeto de análise movimentos, situações e/ou contextos de grupos de ambos os continentes⁸⁵: são eles o artigo de Mulder, Ferguson, Groenewegen, Boersma e Wolbers (2016), que aborda a atuação de grupos humanitários digitais, utilizando financiamento coletivo e plataformas de código aberto, em momentos de crises humanitárias no Nepal e no Haiti; a pesquisa de Rieder, Abdulla, Poell, Woltering e Zack (2015) sobre a página de Facebook “*We are All Khalid*”, fundada por um executivo egípcio da Google no Egito, e que foi uma das propulsoras do movimento político que levou à renúncia do então presidente do país, Hosni Mubarak, em 2011; e o texto do pesquisador Nir Kshetri (2014), sobre os benefícios de aplicações de Big Data em contextos de produção nos países ‘em desenvolvimento’ como Quênia e Brasil, e em regiões como América Latina e África.

⁸⁵ É preciso fazer uma ressalva aqui ao trabalho de Cardullo (2015), que trata do bloqueio do Twitter feito pelo governo turco ante às manifestações contra o primeiro-ministro Erdogan e a maneira como usuários reagiram a isso. O autor construiu sua análise a partir de entrevistas com usuários turcos do Twitter, apresentando características específicas da construção das manifestações a partir de grupos marginais em contextos de censura no Twitter.

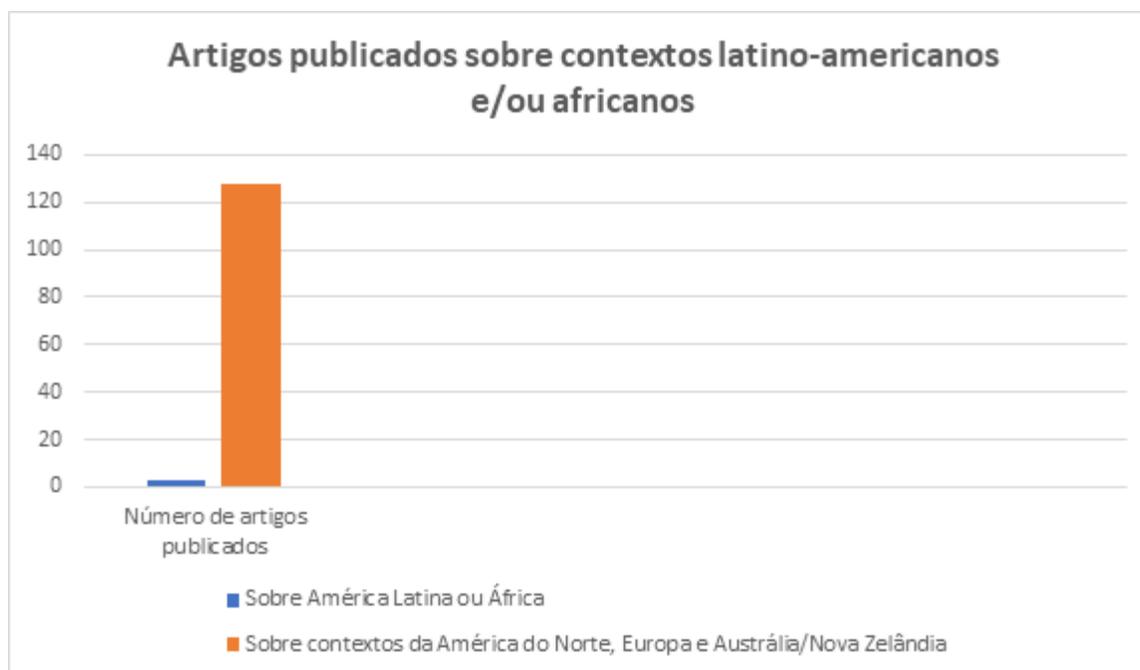


Figura 15: Distribuição de artigos por escolha do objeto de estudo ou contexto de análise em relação a contextos geopolíticos.

Dos três artigos, todavia, apenas dois citam autores locais (Kshetri, 2014; Rieder et al, 2015). O trabalho de Mulder et al (2016) menciona relatórios da Unicef, mapas feitos por ONGs e uma reportagem da revista Wired, porém não cita reportagens, iniciativas, entrevistas, artigos ou textos em geral de populações locais. Rieder et al (2015) contém entre os co-autores uma pesquisadora egípcia, que é referenciada no material (Abdullah, 2014) juntamente com outro autor egípcio e pesquisadores que realizam um longo trabalho de campo na Síria. Kshetri (2014) é, dentre os três, o que mais faz referência a autores do Sul Global, mencionando pesquisas de autores da Indonésia, Índia, Zimbábue e Zâmbia sobre os contextos locais. Seu trabalho argumenta em prol de um uso mais extensivo de práticas de Big Data, bem como um maior acesso a estas, que ‘deram certo’ no contexto de multinacionais e empresas de países da Europa e Estados Unidos em países ‘em desenvolvimento’.

O levantamento, portanto, identificou uma quase total invisibilidade, não apenas na presença de autores do Sul Global, como também de discussões sobre e a partir do Sul nos três primeiros anos de discussão sobre Big Data na BD&S. Outro achado é o referente ao baixo número de referências a literaturas e autores locais quando se abordam contextos do Sul Global, fato visto mesmo dentre os três artigos que abordam tais questões. É a isso que

Ganter e Ortega (2019, p. 79) se referem quando apontam para uma “*tendência de se falar sobre ao invés de com*” o Sul.

Os editores da publicação foram contatados acerca da quantidade de trabalhos submetidos por autores de instituições de ensino latino-americanas ou africanas na revista e localização geográfica de acessos/downloads de artigos, a fim de se avaliar se a ausência de autores do Sul Global nas pesquisas em Big Data pode ser creditada também a um desinteresse sobre o tema ou sobre o material da publicação em si. Os editores responderam que não tinham informações precisas sobre tais questões e que isso demandaria uma “*análise puramente manual*”⁸⁶. Quase 50 dias depois⁸⁷, recebi tal análise em e-mail. Os dados foram organizados na Tabela 8 a seguir e se referem aos dados geográficos de submissão de trabalhos por parte de autoridades com filiação acadêmica a instituições de ensino do Sul Global. Os dados enviados pela BD&S se referem a todo o período de publicação da revista até julho de 2019. Neste contexto, 16% das submissões à revista no período foram feitas por acadêmicos de instituições do Sul Global, enquanto 8% dos artigos da BD&S foram publicados por estes.

País/Região	Aceitos	Rejeitados	Total
Afeganistão	2	0	2
África do Sul	0	1	1
Alemanha	15	19	34
Arábia Saudita	0	1	1
Argentina	0	1	1
Austrália	16	10	26
Áustria	3	2	5
Bélgica	3	5	8
Canadá	14	14	28
Cazaquistão	0	2	2
Chéquia	0	1	1
China	0	7	7

⁸⁶ Mensagem recebida por e-mail no dia 02/06/2019.

⁸⁷ Dados recebidos por e-mail no dia 27/07/2019.

Coréia do Sul	1	1	2
Croácia	1	0	1
Dinamarca	14	8	22
Eslováquia	0	1	1
Espanha	0	2	2
Estados Unidos	81	67	148
Estônia	0	1	1
Filipinas	0	1	1
Finlândia	3	2	5
França	5	9	14
Gana	0	1	1
Hong Kong	0	1	1
Hungria	0	2	2
Índia	1	17	18
Irã	0	5	5
Irlanda	5	2	7
Israel	0	1	1
Itália	3	6	9
Japão	1	0	1
Líbano	0	1	1
Malásia	0	3	3
Malta	0	1	1
Nepal	0	2	2
Nigéria	0	5	5
Noruega	1	2	3
Nova Zelândia	1	2	3
Omã	0	1	1
Países Baixos	21	12	33
Polônia	0	2	2
Portugal	0	2	2

Quênia	0	1	1
Reino Unido	56	41	97
Rússia	0	1	1
Singapura	0	1	1
Suécia	7	5	12
Suíça	2	9	11
Taiwan	1	0	1
Tunísia	0	1	1
Turquia	0	1	1

Tabela 8: Distribuição geográfica das submissões feitas à BD&S desde o início da publicação até julho de 2019. Números separados por país de filiação institucional dos autores.

Os editores foram questionados sobre possíveis razões para esta discrepância, e ofereceram a seguinte resposta: “Um grande número de artigos que recebemos de acadêmicos de instituições do Sul Global (em relação aos de instituições europeias ou norte-americanas) são prontamente rejeitados por não se encaixarem no escopo da publicação”⁸⁸. Além disso, argumentaram que tais artigos evitavam discutir aspectos sociais dos dados. Infelizmente, mais detalhes sobre o que seriam os ‘aspectos sociais dos dados’ ou para outros motivos de rejeição dos artigos não foram dados pelos editores.

Com isso em mente, é preciso notar que diversos fatores necessitam ser considerados para a compreensão da baixa participação de autores, instituições e movimentos do Sul Global nas publicações da BD&S, como, por exemplo, a barreira linguística. Apesar da necessidade de pesquisas mais amplas sobre este assunto, como Ganter e Ortega (2019) também o identificaram em estudo similar na área de Estudos em Mídia e Comunicação, é possível concluir que a [quase] invisibilidade do Sul Global na BD&S e o baixo número de trabalhos sobre contextos e a partir de ideias do Sul Global estão interligados.

4.4 Discussão

A ausência de autores do Sul Global ou de publicações em direção a epistemologias não-ocidentais na BD&S serve de reflexo de uma certa ‘crise de identidade’

⁸⁸ Mensagem recebida por e-mail no dia 27/07/2019.

persistente na BD&S e, como sugerido anteriormente, em diversas outras comunidades científicas no contexto dos estudos sobre dados e mesmo nas Ciências Sociais no geral. Os consecutivos reveses no sonho de uma internet livre, justa e universal, capaz de unir um mundo esfacelado por décadas de Guerra Fria, além do crescente reconhecimento da extensão das práticas de rastreamento de dados de usuários feitas por empresas como o Google e as implicações de seu uso comercial para o enfraquecimento mesmo dos processos democráticos (cf. LYON, 2015), afastam grupos como os vinculados à BD&S das visões mais comuns e empresariais de Big Data e internet - à procura de entendimentos mais críticos sobre as tecnologias de informação e comunicação (TICs).

Por outro lado, como discutido neste e em capítulos anteriores, a procura por alternativas se encontra na chave categórica da tradição ocidental - aqui, especialmente, da própria tradição neoliberal do termo Big Data. A busca por saídas se dá ainda por autores, conceitos e projetos ancorados no contexto euroamericano, ainda que haja um desejo explícito por epistemologias pós-coloniais, que resgatem pensamentos apagados pelo legado colonial. O engajamento quase nulo com perspectivas não-ocidentais e mesmo atores sociais latino-americanos ou africanos, por exemplo, isso dentro de um universo de 131 artigos publicados em um intervalo de três anos em uma publicação científica voltada para olhares críticos sobre os dados, é um sintoma dos tensionamentos que envolvem o projeto científico das comunidades atreladas à BD&S. Embora esta pesquisa não conclua com uma postura pragmática, apresentando soluções para a resolução desta tensão, aponta, todavia, alguns questionamentos que podem contribuir para uma real subversão epistêmica nos estudos sobre dados. Questões como: Onde estão outros pensamentos sobre o mundo? Quem os está produzindo? De que forma dados e sistemas de TICs construídos no contexto euroamericano se relacionam com estes outros mundos? Que efeitos produzem e reproduzem? E no contexto de quem habita o Sul Global, saindo da referência ao europeu, tomando empréstimo, talvez indevidamente, de Achille Mbembe (2018, p. 22-23),

não estará convencido de ser habitado por um duplo, uma entidade estrangeira que o impede de se conhecer a si mesmo? Não vivenciará seu mundo como um definido pela perda e pela cisão e não nutrirá o sonho do regresso a uma identidade consigo mesmo, que regride ao modo da essencialidade pura e, por isso mesmo, muitas vezes, do que lhe é dessemelhante? A partir de quando o projeto de sublevação radical e de

autonomia em nome da diferença se tornará mera inversão mimética daquilo que passamos nosso tempo a cobrir de maldições?

5. CONSIDERAÇÕES FINAIS

O percurso desta pesquisa começou com o objetivo proposto de fornecer um olhar sobre a pesquisa em Big Data nas Ciências Sociais. Em meio a tantos caminhos possíveis e um escopo tão largo, o estudo se voltou para a revista *Big Data & Society*, voltada especificamente para contribuições de perspectivas relacionadas às Ciências Sociais na análise das problemáticas do Big Data na atualidade. A proposta era entender não apenas os principais temas, autores, ênfases e instituições que fomentam o debate naquele espaço, mas como esses autores e comunidades científicas organizam ali suas prioridades, alimentando a própria organização de tais grupos. No caminho, percorremos textos aqui chamados de ‘canônicos’ no contexto da publicação, o relato de um evento científico apoiado pela BD&S e reflexões afins. Um elemento, porém, perpassou cada etapa desta dissertação: uma crescente tensão existente entre a procura por perspectivas críticas dos dados, alternativas às pesquisas de Big Data relacionadas ao mercado, e a manutenção de categorias e linguagem do próprio modelo que buscam repensar. Um choque entre o esforço por pensar novas formas de habitar o digital para além dos modelos sociais, econômicos e políticos do Ocidente e a realidade de não saber se referir ao esgotamento sem esgotar-se ainda mais no funcionamento de mundo típico ao Ocidente ou Norte Global.

Se há um lugar onde esse trabalho deve ser pensado é nesta tensão, de resolução não tão fácil e pronta. Considerando a hodiernidade da pesquisa em Big Data, especialmente em se tratando de perspectivas de caráter mais crítico provindas das Ciências Sociais, essa é uma problemática que produz um momento de autorreflexão em um contexto de crescentes contribuições científicas. O reconhecimento da existência e das implicações desta tensão exigem um pensamento distinto dentro de uma conjuntura onde o volume da produção e a quantidade de congressos, operando quase que num ritmo comercial, imperam como ‘selo do progresso’, ou de utilidade (cf. ORDINE, 2016), na ciência. O que se coloca em xeque aqui é o próprio fazer científico, a própria noção de produção, a própria ideia de Big Data e de resolução de seus problemas. Se há algo que os eventos da modernidade nos ensinaram é acerca do perigo de buscarmos novamente, em meio à falência do modelo ocidental, a resolução dos problemas, as respostas úteis para reerguer as ruínas da ‘democracia ocidental’ – até porque a própria ideia de ‘resolução de problemas’ não parece ocidental demais?

Talvez a melhor resposta para a procura por alternativas para o fim do mundo é esquecer que ele está de fato acabando. Da utilidade de salvação de um projeto em migalhas para a inutilidade do contentamento [já dizia Heidegger (2001, p. 159): “o mais útil é o inútil. Mas experienciar o inútil é a tarefa mais difícil para o homem moderno”] e da inocência da curiosidade por outros mundos.

Sim, é verdade que a ruína do projeto europeu também afeta o Sul Global – e como! São séculos e séculos de esvaziamento de modos de viver não-ocidentais – isto é, da incapacidade, como aponta Mbembe (2018), de poder chamar o mundo de seu ou mesmo de ter um mundo. Porém, ainda assim, não se trata de reunir todo o mundo sob a bandeira da salvação, por meio da resistência e da criatividade, para restaurar um projeto falido, mas de explorar outras epistemologias, outras definições, e de contentar-se nestes espaços. No contexto desta dissertação, vale mesmo questionar o fetiche por um termo tão intrinsecamente atrelado ao Vale do Silício como Big Data. Uma tarefa árdua dentro de uma publicação onde dos mais de 130 artigos, apenas três se preocupam em reconhecer outros espaços fora da América do Norte, Europa e Oceania.

REFERÊNCIAS

AKOKA, Jacky; COMYN-WATTIAU, Isabelle; LAOUFI, Nabil. Research on Big Data: A systematic mapping study. *Computer Standards & Interfaces*, v.54, n.2, p. 105-115, 2017.

AMOORE, Louise. Data derivatives: On the emergence of a security risk calculus for our times. *Theory, Culture & Society*, v.28, n.6, p. 24-43, 2011.

ANDERSON, Chris. *The End of Theory: the data deluge makes the scientific method obsolete*, 2008. Disponível em: <<https://www.wired.com/2008/06/pb-theory/>> Acesso em: 18/11/2018.

ARADAU, Claudia; BLANKE, Tobias. The (Big) Data-security assemblage: Knowledge and critique. *Big Data & Society*, v.2, n.2, p. 1-12, 2015.

ARAÚJO, Ronaldo Ferreira de. *Apropriações de Bruno Latour pela ciência da informação no Brasil: Descrição, explicação e interpretação*. Dissertação (Mestrado em Ciência da Informação). Universidade Federal de Minas Gerais, 2009.

ARORA, Payal. Bottom of the data pyramid: Big Data and the Global South. *International Journal of Communication*, v.10, p. 1681-1699, 2016.

ARMSTRONG, Katharine. Big Data: A revolution that will transform how we live, work, and think (Book Review). *Information, Communication & Society*, v.17, n.10, p. 1300-1302, 2014.

BAACK, Stefan. Datafication and empowerment: How the open data movement re-articulates notions of democracy, participation, and journalism. *Big Data & Society*, v.2, n.2, p. 1-11, 2015.

BANKS, Alan J. Beyond building block theory? A critique of Thomas Kuhn. *Sociological Spectrum*, v.3, n.2, p. 131-141, 1983.

BAUMAN, Zygmunt. *Modernidade e holocausto*. Rio de Janeiro, RJ: Jorge Zahar, 1998.

BATES, Jo; LIN, Yu-Wei; GOODALE, Paula. Data journeys: Capturing the socio-material constitution of data objects and flows. *Big Data & Society*, v.3, n.2, p. 1-12, 2016.

BEER, David. Productive measures: Culture and measurement in the context of everyday neoliberalism. *Big Data & Society*, v.2, n.1, p. 1-12, 2015.

_____. How should we do the history of Big Data? *Big Data & Society*, v.3, n.1, p. 1-10, 2016.

BELCHER, Diane D. Seeking acceptance in an English-only research world. *Journal of Second Language Writing*, v.16, n.1, p. 1-22, 2007.

BENTIVOGLIO, Julio. Entre a história e o cânone: A ciência histórica oitocentista e seus textos fundadores. *História da historiografia*, v.5, n.8, p. 175-186, 2012.

BERRY, Wendell. *Standing by words: Essays*. Berkeley, CA: Shoemaker Hoard, 2005.

BEZINA, Vaclav. Collocation graphs and networks: Selected applications. In: CANTOS-GÓMEZ, P.; ALMELA-SÁNCHEZ, M. (eds). *Lexical collocation analysis*. Quantitative Methods in the Humanities and Social Sciences. Springer, Cham, CH: Springer, 2018. pp. 59–83.

BLEI, David M. Topic modeling and digital humanities. *Journal of Digital Humanities*, v.2, n.1, p. 8–11, 2011.

BLOK, Anders; PEDERSEN, Morten Axel. Complementary social science? Quali-quantitative experiments in a Big Data world. *Big Data & Society*, v.1, n.2, p. 1-6, 2014.

BOLIN, Göran; SCHWARZ, Jonas Andersson. Heuristics of the algorithm: Big Data, user interpretation and institutional translation. *Big Data & Society*, v.2, n.2, p. 1-12, 2015.

BOURDIEU, Pierre. The specificity of the scientific field and the social conditions of the progress of reason. *Social Science Information*, v.14, n.6, p. 19-47, 1975.

_____. *Os usos sociais da ciência: Por uma sociologia clínica do campo científico*. São Paulo, SP: Editora Unesp, 2004.

BOWKER, Geoffrey C. *Memory practices in the sciences*. Cambridge, MA: The MIT Press, 2006.

_____; STARR, Susan Leigh. *Sorting things out: Classification and its consequences*. Cambridge, MA: The MIT Press, 1999.

BOYD, danah. *It's complicated: The social lives of networked teens*. New Haven, CT: Yale University Press, 2014.

_____; CRAWFORD, Kate. Critical questions for Big Data: Provocations for a cultural, technological, and scholarly phenomenon. *Information, Communication & Society*, v.15, n.5, p. 662-679, 2012.

BRONSON, Kelly; KNEZEVIC, Irena. Big Data in food and agriculture. *Big Data & Society*, v.3, n.1, p. 1-5, 2016.

BROOKER, Phillip; BARNETT, Julie; CRIBBIN, Timothy. Doing social media analytics. *Big Data & Society*, v.3, n.2, p. 1-12, 2016.

BURRIS, Val. The academic caste system: Prestige hierarchies in PhD exchange networks. *American Sociological Review*, v.69, n.2, p. 239-264, 2004.

BURROWS, Roger; SAVAGE, Mike. After the crisis? Big Data and the methodological challenges of empirical sociology. *Big Data & Society*, v.1, n.1, p. 1-6, 2014.

BURT, Ronald S. *Structural holes: The social structure of competition*. Cambridge, MA: Harvard University Press, 1995.

_____; DOREIAN, Patrick. Testing a structural model of perception: Conformity and deviance with respect to journal norms in elite sociological methodology. *Quality and Quantity*, v.16, n.2, p. 109-150, 1982.

CANALI, Stefano. Big Data, epistemology and causality: Knowledge in and knowledge out in EXPOsOMICS. *Big Data & Society*, v.3, n.2, p. 1-11, 2016.

CAPPELL, Charles L.; GUTERBOCK, Thomas M. Visible colleges: The social and conceptual structure of sociology specialties. *American Sociology Review*, v.57, n.2, p. 266-273, 1992.

CAPURRO, Rafael. Toward an ontological foundation of information ethics. *Ethics and Information Technology*, v.8, n.4, p. 175-186, 2006.

CARREIRA, Krishma. Imortalidade digital: A era dos grandes dados. *Brazilian Journal of Technology, Communication, and Cognitive Science*, v. 4, p. 1-13, 2016.

CASTELLS, Manuel. *The rise of the network society, the information age: Economy, society and culture vol. I*. Cambridge, MA: Blackwell, 1996.

CATON, Simon; HALL, Margeret; WEINHARDT, Christof. How do politicians use Facebook? An applied Social Observatory. *Big Data & Society*, v.2, n.2, p. 1-18, 2015.

CÉSAIRE, Aimé. *Une tempête: d'après La tempête de Shakespeare*. Paris, FR: Editions du Seuil, 1969.

CHEN, Shu-Heng (Ed.). *Big Data in Computational Social Science and Humanities*. Cham, CH: Springer, 2018.

_____; YU, Tina. *Big Data in Computational Social Science and Humanities: An introduction*. In: CHEN, Shu-Heng (Ed.). *Big Data in Computational Social Science and Humanities*. Cham, CH: Springer, 2018.

CHRISTIAENS, Tim. Digital subjectivation and financial markets: Criticizing Social Studies of Finance with Lazzarato. *Big Data & Society*, v.3, n.2, p. 1-15, 2016.

CIRILLO, Davide; VALENCIA, Alfonso. Big data analytics for personalized medicine. *Current Opinion in Biotechnology*, v.58, p. 161-167, 2019.

COCKAYNE, Daniel. Affect and value in critical examinations of the production and 'prosumption' of Big Data. *Big Data & Society*, v.3, n.2, p. 1-11, 2016.

CONFORTO, Edivandro Carlos; AMARAL, Daniel Capaldo; SILVA, Sérgio Luis da. Roteiro para Revisão Bibliográfica Sistemática: Aplicação no desenvolvimento de produtos

e gerenciamento de projetos. In: Congresso Brasileiro de Gestão de Desenvolvimento de Produto - CBGDP, 8., 2011, Porto Alegre-RS. *Anais... IGDP*, 2011.

COULDRY, Nick; MEJIAS, Ulises A. *The costs of connection: How data is colonizing human life and appropriating it for capitalism*. Palo Alto, CA: Stanford University Press, 2019.

COX, Michael; ELLSWORTH, David. *Managing Big Data for scientific visualization* (01/05/1997). Disponível em: <https://www.researchgate.net/profile/David_Ellsworth2/publication/238704525_Managing_big_data_for_scientific_visualization/links/54ad79d20cf2213c5fe4081a/Managing-big-data-for-scientific-visualization.pdf> Acesso em: 15/01/2020.

CRAWFORD, Kate; MILTNER, Kate; GRAY, Mary L. Critiquing Big Data: Politics, ethics, epistemology. *International Journal of Communication*, v. 8, p. 1663–1672, 2014.

CROLL, Alistair. *Big data is our generation's civil rights issue, and we don't know it*, 2012. Disponível em: <<http://radar.oreilly.com/2012/08/big-data-is-our-generations-civil-rights-issue-and-we-dont-know-it.html>> Acesso em: 18/11/2018.

CROMPTON, Rosemary. Forty years of Sociology: Some comments. *Sociology*, v.42, n.6, p. 1218–1227, 2008.

CRUZ, Leonardo Ribeiro da. Google Suite for Education e o avanço do capitalismo de vigilância sobre as tecnologias educacionais. In: Simpósio Internacional LAVITS “Vigilância, democracia e privacidade na América Latina: vulnerabilidades e resistências”, 5., 2017, Santiago-CH. *Anais... LAVITS*, 2017.

CURRIE, Morgan; PARIS, Britt S.; PASQUETTO, Irene; PIERRE, Jennifer. The conundrum of police officer-involved homicides: Counter-data in Los Angeles County. *Big Data & Society*, v.3, n.2, p. 1-14, 2016.

DALTON, Craig; THATCHER, Jim. *What does a Critical Data Studies look like and why do we care?* - Society and Space, 2014. Disponível em: <<http://societyandspace.org/2014/05/12/what-does-a-critical-data-studies-look-like-and-why-do-we-care-craig-dalton-and-jim-thatcher/>> Acesso em: 16/12/2018.

_____; _____. Inflated granularity: Spatial “Big Data” and geodemographics. *Big Data & Society*, v.2, n.2, p. 1-15, 2015.

_____; TAYLOR, Linnet; THATCHER, Jim. Critical Data Studies: A dialog on data and space. *Big Data & Society*, v.3, n.1, p. 1-9, 2016.

D’ANDREA, Luciano; DECLICH, Andrea. The sociological nature of science communication. *Journal of Science Communication*, v.4, n.2, p. 1-9, 2005.

DE BOISE, Sam. The coming crisis? Some questions for the future of empirical sociology in the UK. *Graduate Journal of Social Science*, v.9, n.2, p. 40-64, 2012.

DELEUZE, Gilles; GUATTARI, Félix. *Mil platôs - capitalismo e esquizofrenia*, vol. 1. Rio de Janeiro: Ed. 34, 1995.

DIEBOLD, Francis D. *On the origin(s) and development of “Big Data”: The phenomenon, the term, and the discipline* (13/02/2013). Disponível em: <https://www.sas.upenn.edu/~fdiebold/papers/paper112/Diebold_Big_Data.pdf> Acesso em: 15/01/2020.

DUBOIS, Didier; HÁJEK, Petr; PRADE, Henri. Knowledge-Driven versus Data-Driven Logics. *Journal of Logic, Language and Information*, v.9, n.1, p. 65-89, 2000.

DUEÑAS, Pilar Mur. Getting research published internationally in English: An ethnographic account of a team of Finance Spanish scholar’s struggles. *Ibérica*, v.24, p. 139-156, 2012.

DUTRA, Luiz Henrique de A. *Introdução à Epistemologia*. São Paulo, SP: Editora UNESP, 2010.

EDELENBOS, Jurian; HIRZALLA, Fadi; VAN ZONEN, Liesbet; VAN DALEN, Jan; BOUMA, Geiske; SLOB, Adriaan; WOESTENBURG, Alexander. Governing the Complexity of Smart Data Cities: Setting a Research Agenda. In: BOLÍVAR, Manuel Pedro Rodríguez. *Smart technologies for smart governments*. New York, NY: Springer, 2018. pp. 35-54.

EDGE, David. Quantitative measures of communication in science: A critical review. *History of Science*, v.17, n.2, p. 102-134, 1979.

EDWARDS, Paul N. *A Vast Machine: computer models, climate data, and the politics of global warming*. Cambridge, MA: The MIT Press, 2010.

ELLUL, Jacques. *The Technological Society*. New York, NY: Vintage Books, 1964.

EVANGELISTA, Rafael. *Para além das máquinas de adorável graça: Cultura hacker, cibernética e democracia*. São Paulo, SP: Edições SESC, 2018.

FAZEKAS, Mihály. *The Use of ‘Big Data’ for Social Sciences Research: an application to corruption research* - SAGE Research Methods Cases, 2014. Disponível em: <<http://methods.sagepub.com/case/big-data-for-social-sciences-research-an-application-to-corruption-research>> Acesso em: 16/12/2018.

FINFGELD-CONNETT, Deborah. Twitter and health science research. *Western Journal of Nursing Research*, v.37, n.10, p. 1269–1283, 2014.

FIRMINO, Rodrigo José; CARDOSO, Bruno de Vasconcelos; EVANGELISTA, Rafael. Hyperconnectivity and (Im)mobility: Uber and Surveillance Capitalism by the Global South. *Surveillance & Society*, v.17, n.1/2, p. 205-212, 2019.

FORTUN, Kim; POIRIER, Lindsay; MORGAN, Alli; COSTELLOE-KUEHN, Brandon; FORTUN, Mike. Pushback: Critical data designers and pollution politics. *Big Data & Society*, v.3, n.2, p. 1-14, 2016.

FOSTER, Ian; GHANI, Rayid; JARMIN, Ron S.; KREUTER, Frauke; LANE, Julia (Eds.). *Big Data and Social Science: A practical guide to methods and tools*. Boca Raton, FL: CRC Press, 2017.

FOUCAULT, Michel. *Microfísica do poder*. 13^a ed. Rio de Janeiro: Edições Graal, 1998.

GANTER, Sarah Anne; ORTEGA, Félix. The Invisibility of Latin American scholarship in European Media and Communication Studies: Challenges and opportunities of de-westernization and academic cosmopolitanism. *International Journal of Communication*, v.13, p. 68-91, 2019.

GARVEY, William D. *Communication: The essence of science*. Oxford, UK: Pergamon Press, 1979.

GILLESPIE, Tarleton. The Relevance of Algorithms. In: GILLESPIE, Tarleton; BOCZKOWSKI, Pablo J.; FOOT, Kirsten A. (Eds.). *Media Technologies: Essays on Communication, Materiality and Society*. Cambridge, MA: MIT Press, 2014. pp. 167-194.

GITELMAN, Lisa (Ed.). *"Raw Data" is an oxymoron*. Cambridge, MA: The MIT Press, 2013.

GRAY, Emily; JENNINGS, Will; FARRALL, Stephen; HAY, Colin. Small Big Data: Using multiple data-sets to explore unfolding social and economic change. *Big Data & Society*, v.2, n.1, p. 1-6, 2015.

GUBER, Rosana. *La etnografía: método, campo y reflexividad*. Bogotá: Grupo Editorial Norma, 2001.

GUIMARÃES, Vera Aparecida Lui; HAYASHI, Maria Cristina Piumbato Innocentini. Os eventos científicos: Espaços privilegiados para a comunicação da ciência. *Comunicologia: Revista de Comunicação da Universidade Católica de Brasília*, v.7, n.2, p. 204-229, 2014.

GUTTAL, Shalmali. Interrogating the relevance of the Global North-South divide, 3 fev. 2016. Disponível em: <https://www.cetri.be/IMG/pdf/shalmali_guttal_23000_eng_3-2.pdf> Acesso em: 15/11/2019.

HARGITTAI, Eszter; HSIEH, Yuli Patrick. Digital inequality. In: DUTTON, William H. (Ed.), *Oxford handbook of Internet studies*. Oxford, UK: Oxford University Press, 2013. pp. 129-150.

HARAWAY, Donna. Saberes situados: A questão da ciência para o feminismo e o privilégio da perspectiva parcial. *Cadernos Pagu*, v.5, p. 7-41, 1995.

HAYASHI, Carlos Roberto Massao; RIGOLIN, Camila Carneiro Dias; GUIMARÃES, Vera Aparecida Lui; HAYASHI, Maria Cristina Piumbato Innocentini. Colégio invisível, revisão pelos pares e normalização documentária na revista do Instituto Histórico e Geográfico de São Paulo. *Informação & Informação*, v.21, n.3, p. 5-33, 2016.

HEIDEGGER, Martin. *Introduction to Metaphysics*. New Haven, CT: Yale Nota Bene, 2000[1953].

_____. *Zollikon seminars: Protocols, conversations, letters*. Evanston, IL: Northwestern University Press, 2001.

_____. The Provenance of Art and the Destination of Thought. *Journal of the British Society for Phenomenology*, v. 44, n. 2, p. 119-128, 2013.

HONG, Renyi. Soft skills and hard numbers: Gender discourse in human resources. *Big Data & Society*, v.3, n.2, p. 1-13, 2016.

HOUSLEY, William; PROCTER, Rob; EDWARDS, Adam; BURNAP, Peter; WILLIAMS, Matthew; SLOAN, Luke; RANA, Omer; MORGAN, Jeffrey; VOSS, Alex; GREENHILL, Anita. Big and broad social data and the sociological imagination: A collaborative response. *Big Data & Society*, v.1, n.2, p. 1-15, 2014.

HUC-HEPHER, Saskia. Big Web data, small focus: An ethnosemiotic approach to culturally themed selective Web archiving. *Big Data & Society*, v.2, n.2, p. 1-15, 2015.

ILIADIS, Andrew; RUSSO, Federica. Critical Data Studies: an introduction. *Big Data & Society*, v.3, n.2, p. 1-7, 2016.

JENNEX, Murray E. Big Data, the Internet of Things, and the Revised Knowledge Pyramid. *ACM SIGMIS Database: the DATABASE for Advances in Information Systems*, v.48, n.4, p. 69-79, 2017.

JEREZ-ROIG, Javier; GUEDES, Marcello Barbosa Otoni Gonçalves; SILVA, João Maria Dias; LIMA, Kenio Costa de. Análise da produção científica da Revista Brasileira de Geriatria e Gerontologia: uma revisão bibliométrica. *Rev. Bras. Geriatr. Gerontol.*, v.17, n.3, p. 659-671, 2014.

JOSS, Simon; COOK, Matthew; DAYOT, Youri. Smart cities: Towards a new citizenship regime? A discourse analysis of the British smart city standard. *Journal of Urban Technology*, v.24, n.4, p. 29-49, 2017.

KACOWICZ, Ariel M. Globalization, poverty and the North-South divide. *International Studies Review*, v. 9, n. 4, p. 565-580, 2007.

KAPLAN, Frédéric. A map for big data research in digital humanities. *Frontiers in Digital Humanities*, v.2, n.1, p. 1-7, 2015.

KENNEDY, Helen; MOSS, Giles. Known or knowing publics? Social media data mining and the question of public agency. *Big Data & Society*, v.2, n.2, p. 1-11, 2015.

_____; POELL, Thomas; VAN DIJCK, José. Data and agency. *Big Data & Society*, v.2, n.2, p. 1-7, 2015.

KERMODE, Frank. *Forms of attention: Botticelli and Hamlet*. 2^a ed. Chicago, IL: The University of Chicago Press, 2011.

KIM, Gang-Hoon; TRIMI, Silvana; CHUNG, Ji-Hyong. Big-Data Applications in the Government Sector. *Communications of the ACM*, v.57, n.3, p. 78-85, 2014.

KITCHIN, Rob. *The Data Revolution: Big Data, Open Data, data infrastructures and their consequences*. London, UK: Sage, 2014a.

_____. Big Data, New Epistemologies and Paradigm Shifts. *Big Data & Society*, v.1, n.1, p. 1-12, 2014b.

_____; MCARDLE, Gavin. What makes Big Data, Big Data? Exploring the ontological characteristics of 26 datasets. *Big Data & Society*, v.3, n.1, p. 1-10, 2016.

_____; LAURIAULT, Tracey P. Toward Critical Data Studies: Charting and unpacking data assemblages and their work. In: THATCHER, Jim; SHEARS, Andrew; ECKERT, Josef (Eds.). *Thinking Big Data in Geography: New regimes, new research*, p. 3-20. Lincoln, NE: University of Nebraska Press, 2018.

KRIPPENDORF, Klaus. Content analysis. In: BARNOUW, Erik; GERBNER, George; SCHRAMM, Wilbur; WORTH, Tobia L.; GROSS, Larry (Eds.). *International Encyclopedia of Communication*. Vol. 1, p. 403-407. New York, NY: Oxford University Press, 1989.

KSHETRI, Nir. The emerging role of Big Data in key development issues: Opportunities, challenges, and concerns. *Big Data & Society*, v.1, n.2, p. 1-20, 2014.

KUHN, Thomas. *The Structure of Scientific Revolutions*. Chicago, IL: The University of Chicago Press, 1962.

KUKUTAI, Tahu; TAYLOR, John (Eds.). *Indigenous Data Sovereignty: Toward an agenda*. Canberra, AU: ANU Press, 2016.

KWON, Ohbyung; LEE, Namyoon; SHIN, Bongsik. Data quality management, data usage experience and acquisition intention of Big Data analytics. *International Journal of Information Management*, v.34, n.3, p. 387-394, 2014.

LABOVITZ, Sanford. Statistical usage in Sociology: sacred cows and rituals. *Sociological Methods & Research*, v.1, n.1, p. 13-17, 1972.

LAGOZE, Carl. Big Data, data integrity, and the fracturing of the control zone. *Big Data & Society*, v.1, n.2, p. 1-11, 2014.

LALUMIA, Joseph. *Kuhn and his critics on normal and revolutionary science*. Diogenes, v.39, n.154, p. 39-45, 1991.

LANEY, Doug. *3D data management: Controlling data volume, velocity and variety*, 2001. Disponível em: <<http://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf>> Acesso em: 18/11/2018.

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

LAZER, David; PENTLAND, Alex; ADAMIC, Lada; ARAL, Sinan; BARABÁSI, Albert-László; BREWER, Devon; CHRISTAKIS, Nicholas; CONTRACTOR, Noshir; FOWLER, James; GUTMANN, Myron; JEBARA, Tony; KING, Gary; MACY, Michael; ROY, Deb; VAN ALSTYNE, Marshall. Computational Social Science. *Science*, v.323, n.5915, p. 721-723, 2009.

LEÓN, Luis F. Alvarez. Property regimes and the commodification of geographic information: An examination of Google Street View. *Big Data & Society*, v.3, n.2, p. 1-13, 2016.

LEONELLI, Sabina. What difference does quantity make? On the epistemology of Big Data in biology. *Big Data & Society*, v.1, n.1, p. 1-11, 2016.

LESZCZYNSKI, Agnieszka; CRAMPTON, Jeremy. Introduction: Spatial Big Data and everyday life. *Big Data & Society*, v.3, n.2, p. 1-6, 2016.

LEVY, Pierre. *Cibercultura*. São Paulo, SP: Editora 34, 1999.

LEWIS, Seth C.; ZAMITH, Rodrigo; HERMIDA, Alfred. Content analysis in an era of Big Data: A hybrid approach to computational and manual methods. *Journal of Broadcasting & Electronic Media*, v.57, p. 34-52, 2013.

LIEVROUW, Leah A. Reconciling structure and process in the study of scholarly communication. In: BORGMAN, Christine L. (Ed.). *Scholarly communication and bibliometrics* (pp. 59-69). Newbury Park: SAGE, 1990.

LIPPERT, Ingmar. Failing the market, failing deliberative democracy: How scaling up corporate carbon reporting proliferates information asymmetries. *Big Data & Society*, v.3, n.2, p. 1-13, 2016.

LIZ, Stanley. It has always known, and we have always been 'other': Knowing capitalism and the 'coming crisis' of sociology confront the concentration system and Mass-Observation. *The Sociological Review*, v.56, n.4, p. 535-551, 2008.

LOHR, Steve. *The origins of 'Big Data': An etymological detective story* (01/02/2013). Disponível em: <<https://bits.blogs.nytimes.com/2013/02/01/the-origins-of-big-data-an-etymological-detective-story/>> Acesso em: 15/01/2020.

LOUKISSAS, Yanni Alexander. A place for Big Data: Close and distant readings of accessions data from the Arnold Arboretum. *Big Data & Society*, v.3, n.2, p. 1-20, 2016.

LUCHETTA, Sara. Literary Mapping: At the Intersection of Complexity and Reduction. *Literary Geographies*, v.4, n.1, p. 6-9, 2018.

LUPTON, Deborah. *Digital Sociology*. New York, NY: Routledge, 2015.

LYON, David. Surveillance, Snowden, and Big Data: Capacities, consequences, critique. *Big Data & Society*, v.1, n.2, p. 1-13, 2014.

_____. *Surveillance after Snowden*. Cambridge, UK: Polite Press, 2015.

MACKENZIE, Adrian. The production of prediction: What does machine learning want? *European Journal of Cultural Studies*, v.18, n.4-5, p. 429-445, 2015.

MADSEN, Anders Koed. Between technical features and analytic capabilities: Charting a relational affordance space for digital social analytics. *Big Data & Society*, v.2, n.1, p. 1-15, 2015.

_____; FLYVERBOM, Mikkel; HILBERT, Martin; RUPPERT, Evelyn. Big Data: Issues for an international political sociology of data practices. *International Political Sociology*, v. 10, p. 275-296, 2016.

MAHRENBACH, Laura C.; MAYER, Katja; PFEFFER, Jürgen. Policy visions of Big Data: Views from the Global South. *Third World Quarterly*, v.39, n.10, p. 1861-1882, 2018.

MANN, Monique; DALY, Angela. (Big) data in the North-in-South: Australia's informational imperialism and digital colonialism. *Television & New Media*, v.20, n.4, p. 379-395, 2019.

MARTIN, Olivier. Da estatística política à sociologia: desenvolvimento e transformações da análise estatística da sociedade (séculos XVII – XIX). *Revista Brasileira de História*, v.21, n.41, p. 13-34, 2001.

MAYER-SCHÖNBERGER, Viktor; CUKIER, Kenneth. *Big Data: A revolution that will transform how we live, work and think*. London, UK: John Murray, 2013.

MBEMBE, Achille. *Crítica da razão negra*. São Paulo, SP: n-1 edições, 2018.

MCFARLAND, Daniel A.; LEWIS, Kevin; GOLDBERG, Amir. Sociology in the era of Big Data: The ascent of forensic Social Sciences. *American Sociologist*, v.47, n.1, p. 12-35, 2017.

MCKIE, Linda; RYAN, Louise (Eds.). *An end to the crisis of empirical sociology? Trends and challenges in social research*. London, UK: Routledge, 2016.

MENICELLI, Francesca. Review of Kitchin's 'The Data Revolution'. *Surveillance & Society*, v. 13, n.2, p. 319-321, 2015.

METCALF, Jacob; CRAWFORD, Kate. Where are human subjects in Big Data research? The emerging ethics divide. *Big Data & Society*, v.3, n.1, p. 1-14, 2016.

MILAN, Stefania; TRERÉ, Emiliano. *Big Data from the South: The Beginning of a Conversation We Must Have*. 2017. Disponível em: <<https://ssrn.com/abstract=3056958>>. Acesso em: 16/12/2018.

_____; _____. Big Data from the South(s): Beyond data universalism. *Television & New Media*, v.20, n.4, p. 319-335, 2019.

MILLER, John H. *The outlier: Japan between Asia and the West*, 2004. Disponível em: <<https://apcss.org/Publications/Ocasional%20Papers/OPTheOutlierJapanBetweenAsiaWest.pdf>> Acesso em: 14/12/2018.

MOATS, David; SEAVER, Nick. "You social scientists love mind dames": Experimenting in the "divide" between data science and critical algorithm studies. *Big Data & Society*, v.6, n.1, p. 1-11, 2019.

MOHR, John W.; WAGNER-PACIFICI, Robin; BREIGER, Ronald L. Toward a computational hermeneutics. *Big Data & Society*, v.2, n.2, p. 1-8, 2015.

MORRISON, Toni. *Beloved*. New York, NY: Vintage International, 2004.

NELSON, Leif D., SIMMONS, Joseph P. On southbound ease and northbound fees: Literal consequences of the metaphoric link between vertical position and cardinal direction. *Journal of Marketing Research*, v.46, n.6, p. 715-724, 2009.

OLIVEIRA, Francisco de. *A economia da dependência imperfeita*. 4ª ed. Rio de Janeiro, RJ: Edições Graal, 1984.

_____. *Crítica à razão dualista - O ornitorrinco*. São Paulo, SP: Boitempo Editorial, 2003.

ORDINE, Nuccio. *A utilidade do inútil: um manifesto*. Rio de Janeiro, RJ: Jorge Zahar, 2016.

OVADIA, Steven. The role of big data in the Social Sciences. *Behavioral & Social Sciences Librarian*, v.32, n.2, p. 130-134, 2013.

QUAN-HAASE, Anabel; MARTIN, Kim; MCCAY-PEET, Lori. Networks of digital humanities scholars: The informational and social uses and gratifications of Twitter. *Big Data & Society*, v.2, n.1, p. 1-12, 2015.

PAISLEY, William. The role of invisible colleges in scientific information transfer. *Educational Researcher*, v.1, n.4, p. 5-19, 1972.

PARISE, Salvatore. Big Data: A revolution that will transform how we live, work, and think, by Viktor Mayer-Schonberger and Kenneth Cukier. *Journal of Information Technology Case and Application Research*, v.18, n.3, p. 186–190, 2016.

PENTLAND, Alex. *Social Physics: how social networks can make us smarter*. New York, NY: Penguin Books, 2014.

PERNG, Sung-Yueh; KITCHIN, Rob; EVANS, Leighton. Locative media and data-driven computing experiments. *Big Data & Society*, v.3, n.1, p. 1-12, 2016.

PHILLIPS, Martin. *Aspects of text structure: An investigation of the lexical organisation of text*. Amsterdam: North-Holland, 1985.

PIGNARRE, Philippe; STENGERS, Isabelle. *Capitalist sorcery: Breaking the spell*. Basingstoke & New York, NY: Palgrave Macmillan, 2011.

PIRYANI, Rajesh; MADHAVI, Devaraj; SINGH, Vivek Kumar. Analytical mapping of opinion mining and sentiment analysis research during 2000–2015. *Information, Processing and Management*, v.53, n.1, p. 122-150, 2017.

POIRIER, John C. An ontological definition of “canon”? *Bulletin for Biblical Research*, v. 24, n. 4, p. 457-466, 2014.

PRICE, Derek J. de Solla. *O desenvolvimento da ciência: análise histórica, filosófica, sociológica e econômica*. Rio de Janeiro: Livros Técnicos, 1976.

PUSCHMANN, Cornelius; BURGESS, Jean. Metaphors of Big Data. *International Journal of Communication*, v.8, n.1, p. 1690-1709, 2014.

PYBUS, Jennifer; COTÉ, Mark; BLANKE, Tobias. Hacking the social life of Big Data. *Big Data & Society*, v.2, n.2, p. 1-10, 2015.

RAFTERY, Adrian E. Statistics in Sociology, 1950-2000: A selective review. *Sociological Methodology*, v.31, p. 1-15, 2001.

RIEDER, Gernot; SIMON, Judith. Datatrust: Or, the political quest for numerical evidence and the epistemologies of Big Data. *Big Data & Society*, v.3, n.1, p. 1-6, 2016.

RODRÍGUEZ, Pablo. *Historia de la información*. Buenos Aires: Capital Intelectual, 2012.

RUPPERT, Evelyn; LAW, John, SAVAGE, Mike. Reassembling social science methods: The challenge of digital devices. *Theory, Culture & Society*, v.30, n.4, p. 22–46, 2013.

SÆTRA, Henrik Skaug. Science as a Vocation in the Era of Big Data: the Philosophy of Science behind Big Data and humanity's Continued Part in Science. *Integrative Psychological and Behavioral Science*, v.52, n.4, p. 508–522, 2018.

SANTOS, Boaventura de Sousa. *Um discurso sobre as ciências*. 15ª ed. Porto, PT: Afrontamento, 2007.

_____. *O fim do império cognitivo: a afirmação das epistemologias do sul*. Belo Horizonte, MG: Autêntica, 2019.

SAVAGE, Mike; BURROWS, Roger. The coming crisis of empirical sociology. *Sociology*, v.41, n.5, p. 885-899, 2007.

_____; _____. Some Further Reflections on the Coming Crisis of Empirical Sociology. *Sociology*, v.43, n.4, p. 762-772, 2009.

SCHROEDER, Ralph. Big Data and the brave new world of social media research. *Big Data & Society*, v.1, n.2, p. 1-11, 2014.

SHARON, Tamar. Self-Tracking for health and the quantified self: Re-articulating autonomy, solidarity, and authenticity in an age of personalized healthcare. *Philosophy & Technology*, v.30, n.1, p. 93–121, 2017.

SILVA, Guilherme Cavalcante. *Elementos para uma crítica da noção cibernética da comunicação: um estudo sobre as premissas cibernéticas da comunicação à luz do avanço da inteligência artificial*. Monografia (Bacharelado em Comunicação Social – Jornalismo). Centro Universitário Adventista de São Paulo, 2016.

_____. Para repensar Comunicação e Big Data: Situando caminhos a partir da edição especial do Journal of Communication. *Revista do Edicc*, v.5, p. 107-118, 2018.

_____. North Perspectives for a Better South? Big Data and the Global South in Big Data & Society. *Interações: Sociedade e as Novas Modernidades*, v. 37, p. 84-107, 2019.

SILVA, Rafael Alves da. *A exaustão de Sísifo: gestão produtiva, trabalhador contemporâneo e novas formas de controle*. Dissertação (Mestrado em Sociologia). Universidade Estadual de Campinas, 2008.

SILVEIRA, Sérgio Amadeu; AVELINO, Rodolfo; SOUZA, Joyce. A privacidade e o mercado de dados pessoais. *Liinc em Revista*, v.12, n.2, p. 217-230, 2016.

STARR, Paul. *Ainda é possível salvar a internet?* 2019 (08/10/2019). Disponível em: <<https://outraspalavras.net/mercadosdemocracia/ainda-e-possivel-resgatar-a-internet/>> Acesso em: 18/02/2020.

STEHLE, Samuel; PEUQUET, Donna J. Analyzing spatio-temporal patterns and their evolution via sequence alignment. *Spatial Cognition & Computation*, v.15, n.2, p. 68-85, 2015.

STRATHERN, Marilyn. *O efeito etnográfico e outros ensaios*. São Paulo, SP: Cosac Naify, 2014.

STRUJIS, Peter; BRAAKSMA, Barteld; DAAS, JH. Official statistics and Big Data. *Big Data & Society*, v.1, n.1, p. 1-6, 2014.

SWAN, Melanie. Philosophy of Big Data: expanding the human-data relation with Big Data Science Services. *Proceedings of the 2015 IEEE First International Conference on Big Data Computing Service and Applications* (p. 468-477). Washington, DC: IEEE Computer Society, 2015.

SYMONS, John; ALVARADO, Ramón. Can we trust Big Data? Applying philosophy of science to software. *Big Data & Society*, v.3, n.2, p. 1-17, 2016.

TAYLOR, Linnet; SCHROEDER, Ralph; MEYER, Eric. Emerging practices and perspectives on Big Data analysis in economics: Bigger and better or more of the same? *Big Data & Society*, v.1, n.2, p. 1–10, 2014.

TEDDLIE, Charles; TASHAKKORI, Abbas. *Foundations of mixed methods research: Integrating quantitative and qualitative approaches in the social and behavioral sciences*. Thousand Oaks, CA: SAGE Publications, 2009.

TEMPINI, Niccolò. *Book review: Big Data: A Revolution that will transform how we live, work, and think by Kenneth Cukier and Viktor Mayer-Schonberger*. 2013. Disponível em: <<https://blogs.lse.ac.uk/lsereviewofbooks/2013/05/02/book-review-big-data-a-revolution-that-will-transform-how-we-live-work-and-think/>> Acesso em: 21/05/2019.

TIMANS, Rob; WOUTERS, Paul; HEILBRON, Johan. Mixed methods research: what it is and what it could be. *Theory and Society*, v. 48, n. 2, p. 193-216, 2019.

THATCHER, Jim. Living on Fumes: Digital footprints, data fumes, and the limitations of spatial big data. *International Journal of Communication*, v.8, p. 1765–2783, 2014.
_____. The object of mobile spatial data, the subject in mobile spatial research. *Big Data & Society*, v.3, n.2, p. 1-7, 2016.

THE ANALYTICAL Greek Lexicon: consisting of an alphabetical arrangement of every occurring inflexion of every word contained in the Greek New Testament Scriptures. New York, NY: Harper & Brothers Publishers, 1852.

THORNHAM, Helen; GÓMEZ CRUZ, Edgar. Hackathons, data and discourse: Convolutions of the data (logical). *Big Data & Society*, v.3, n.2, p. 1-11, 2016.

THRIFT, Nigel. *Knowing capitalism*. London, UK: Sage, 2005.

_____. The ‘sentient’ city and what it may portend. *Big Data & Society*, v.1, n.1, p. 1-21, 2014.

TOMAZ, Tales; SILVA, Guilherme Cavalcante. Repensando big data, algoritmos e comunicação: Para uma crítica da neutralidade instrumental. *Revista Parágrafo*, v.6, n.1, p. 31-42, 2018.

UPRICHARD, Emma. *Focus: Big Data, little questions?* 2013. Disponível em: <<https://discoversociety.org/2013/10/01/focus-big-data-little-questions/>> Acesso em: 16/12/2018.

VANDERSTRAETEN, Raf. Scientific communication: Sociology journals and publication practices. *Sociology*, v.44, n.3, p. 559-576, 2010.

VAN DER VLIST, Fernando N. Accounting for the social: Investigating commensuration and Big Data practices at Facebook. *Big Data & Society*, v.3, n.1, p. 1-16, 2016.

VAN DIJCK, José; POELL, Thomas. Understanding the promises and premises of online health platforms. *Big Data & Society*, v.3, n.1, p. 1-11, 2016.

VELKOVA, Julie. Data that warms: Waste heat, infrastructural convergence and the computation traffic commodity. *Big Data & Society*, v.3, n.2, p. 1-10, 2016.

VIVEIROS DE CASTRO, Eduardo. Os pronomes cosmológicos e o perspectivismo ameríndio. *Mana*, v.2, n.2, p. 115-144, 1996.

_____. *Metafísicas canibais: elementos para uma antropologia pós-estrutural*. São Paulo, SP: Cosac Naify/n-1 edições, 2015.

WAGNER, Caroline S. *The new invisible college: science for development*. Washington, D.C: Brookings Institution Press, 2008.

WEBBER, Richard. Response to 'The coming crisis of empirical sociology': An outline of the research potential of administrative and transactional data. *Sociology*, v.43, n.1, p. 169-178, 2009.

_____; BUTLER, Tim; PHILLIPS, Trevor. Adoption of geodemographic and ethno-cultural taxonomies for analysing Big Data. *Big Data & Society*, v.2, n.1, p. 1-16, 2015.

WIENHOFEN, Leendert W.M.; MATHISEN, Bjørn Magnus; ROMAN, Dumitru. *Empirical Big Data research: A systematic literature mapping*, 2015. Disponível em: <<https://arxiv.org/pdf/1509.03045.pdf>> Acesso em: 07/07/2019.

WILLIAMS, Malcolm; PAYNE, Geoff; HODGKINSON, Liz; POADE, Donna. Does British sociology count? Sociology students' attitudes toward quantitative methods. *Sociology*, v.42, n.5, p. 1003-1021, 2008.

WILLIAMSON, Ben. Educating the smart city: Schooling smart citizens through computational urbanism. *Big Data & Society*, v.2, n.2, p. 1-13, 2015.

_____. Digital methodologies of education governance: Pearson plc and the remediation of methods. *European Educational Research Journal*, v.15, n.1, p. 34-53, 2016.

WILSON, Matthew W. Morgan Freeman is dead and other big data stories. *cultural geographies*, v.22, n.2, p. 345-349, 2015.

YLIJOKI, Ossi; PORRAS, Jari. Conceptualizing Big Data: Analysis of case studies. *Intelligent Systems in Accounting, Finance and Management*, v.23, n.4, p. 295-310, 2016.

YOUTIE, Jan; PORTER, Alan L.; HUANG, Ying. Early social science research about Big Data. *Science and Public Policy*, v.44, n.1, p. 65-74, 2017.

ZANATTA, Rafael A.F. *Proteção de dados pessoais como regulação de risco: Uma nova moldura teórica?* In: I Encontro da Rede de Pesquisa em Governança da Internet, novembro de 2017. Disponível em: <http://www.redegovernanca.net.br/public/conferences/1/anais/ZANATTA,%20Rafael_2017.pdf>. Acesso em: 05/07/2019.

ZUBOFF, Shoshana. *In the age of the smart machine: The future of work and power*. New York, NY: Basic Books, 1988.

_____. *The age of surveillance capitalism: The fight for a human future at the new frontier of power*. New York, NY: Public Affairs, 2019.

ZUCALLA, Alesia Ann. *Revisiting the invisible college: A case study of the intellectual structure and social process of singularity theory research in mathematics*. Tese (PhD – Faculty of Information Studies). University of Toronto, 2004.

_____; BESSELAAR, Peter van den. Mapping review networks: Exploring research community roles and contributions. *Scientometrics*, v.81, n.1, p. 111-122, 2009.

ZWITTER, Andrej. Big Data ethics. *Big Data & Society*, v.1, n.2, p. 1-6, 2014.

ANEXO - Tabela com termos utilizados nas palavras-chave da BD&S entre 2014 e 2016

TERMO	NÚMERO DE APARIÇÕES NAS PALAVRAS-CHAVE
data	161
big	73
social	52
analysis	26
science	24
media	23
digital	20
studies	14
analytics	11
critical	11
environmental	11
methods	11
computational	10
information	9
machine	9
network	9
surveillance	9
epistemology	8
learning	8
sociology	7
text	7
agency	6
ethics	6

governance	6
mining	6
networks	6
privacy	6
research	6
theory	6
topic	6
twitter	6
activism	5
modelling	5
models	5
policy	5
practices	5
snowden	5
technology	5
algorithms	4
community	4
cultural	4
facebook	4
humanities	4
political	4
reading	4
security	4
spatial	4
based	3
change	3

climate	3
complexity	3
control	3
critique	3
culture	3
datafication	3
design	3
devices	3
economy	3
error	3
hacking	3
human	3
infrastructure	3
journalism	3
literary	3
management	3
materiality	3
modeling	3
ontology	3
open	3
photography	3
predictive	3
production	3
publics	3
scaling	3
sciences	3

statistics	3
technologies	3
tracking	3
transparency	3
visual	3
wikipedia	3
affordances	2
agriculture	2
apps	2
assemblage	2
automated	2
biology	2
cartography	2
causality	2
citizen	2
city	2
class	2
classification	2
close	2
code	2
collaboration	2
computation	2
computing	2
crowdsourcing	2
curation	2
detection	2

development	2
discourse	2
distant	2
economics	2
ethnography	2
etnography	2
events	2
experiments	2
farmers	2
finance	2
freedom	2
git	2
google	2
heat	2
historical	2
infrastructures	2
institutional	2
intensive	2
internet	2
interpretation	2
knowledge	2
life	2
making	2
mobile	2
monitoring	2
mortality	2

national	2
numbers	2
online	2
opacity	2
personal	2
place	2
pollution	2
power	2
practice	2
public	2
quantitative	2
risk	2
sentiment	2
situated	2
sociotechnical	2
software	2
street	2
structure	2
subject	2
subjects	2
taxonomy	2
urban	2
visualisation	2
web	2
abduction	1
abductive	1

accounting	1
action	1
activity	1
administrative	1
adoption	1
advanced	1
advertising	1
aesthetics	1
affect	1
agenda	1
agribusiness	1
air	1
algorithm	1
analyses	1
angeles	1
annemarie	1
anxiety	1
application	1
architecture	1
archive	1
archiving	1
art	1
arts	1
ass	1
atlas	1
augé	1

automation	1
bag	1
balance	1
bank	1
banks	1
basket	1
bias	1
bills	1
biomarkers	1
biomedicine	1
biometric	1
black	1
bladder	1
bourdieu	1
bowker	1
box	1
burke	1
business	1
calculated	1
cancer	1
causal	1
central	1
characteristics	1
circulations	1
citizens	1
citizenship	1

citizenships	1
civic	1
collective	1
commensuration	1
commodification	1
common	1
communication	1
complementarity	1
computerization	1
concept	1
conditions	1
configurations	1
connections	1
consent	1
construction	1
content	1
contestations	1
controversy	1
conviviality	1
cookies	1
cooling	1
correlation	1
cosmos	1
crime	1
crisis	1
criticism	1

crowd	1
cultures	1
databases	1
datalogical	1
dataspheres	1
dates	1
dating	1
definition	1
description	1
diction	1
disaster	1
discrimination	1
discursive	1
distinction	1
distribution	1
divide	1
donna	1
driven	1
drones	1
dynamics	1
earth	1
econometrics	1
education	1
edward	1
electronic	1
emotion	1

empirical	1
employed	1
encyclopedia	1
end	1
engineering	1
england	1
environment	1
ethnicity	1
ethnosemiotics	1
european	1
evaluating	1
evaluation	1
everyday	1
evidence	1
expectancy	1
expiry	1
exploratory	1
exposomics	1
fake	1
feminism	1
feminist	1
fictions	1
field	1
fieldwork	1
filtering	1
financial	1

flow	1
followers	1
footprints	1
forensic	1
forensics	1
formation	1
foucault	1
framing	1
friends	1
furnace	1
future	1
gender	1
genome	1
geodemographics	1
geoengineering	1
geographic	1
geography	1