



Matrizes

ISSN: 1982-2073

ISSN: 1982-8160

matrizes@usp.br

Universidade de São Paulo

Brasil

Schroeder, Ralph

Big data: moldando o conhecimento, moldando a vida cotidiana 1

Matrizes, vol. 12, núm. 2, 2018, Maio-, pp. 135-163

Universidade de São Paulo

Brasil

DOI: <https://doi.org/10.11606/issn.1982-8160.v12i2p135-163>

Disponível em: <https://www.redalyc.org/articulo.oa?id=143065200016>

- ▶ Como citar este artigo
- ▶ Número completo
- ▶ Mais informações do artigo
- ▶ Site da revista em redalyc.org

UABM 

Sistema de Informação Científica Redalyc

Rede de Revistas Científicas da América Latina e do Caribe, Espanha e Portugal

Sem fins lucrativos acadêmica projeto, desenvolvido no âmbito da iniciativa
acesso aberto

Big data: moldando o conhecimento, moldando a vida cotidiana¹

Big data: shaping knowledge, shaping everyday life

RALPH SCHROEDER ^a

University of Oxford, Oxford Internet Institute. Oxford, Reino Unido

RESUMO

Este artigo discute a *novidade* do *big data*, argumentando que ela pode ser definida em relação ao tipo de conhecimento que é criado. Destaca a importância do *big data* na pesquisa em ciências sociais e, especialmente, na pesquisa em comunicação, apesar de essa investigação ser em parte limitada pelas fontes (comerciais, muitas vezes) de dados. Outro impedimento é que essa pesquisa está avançando em muitas direções, mas sem integrar o conhecimento descoberto em relatos gerais do papel da mídia na mudança social. E a maior parte desse conhecimento não está sendo produzido nas ciências sociais, mas sim no setor privado e em contextos políticos. O *big data* recoloca também, em novas perspectivas, certas questões como a da privacidade e a da vigilância social. **Palavras-chave:** *Big data*, mídia digital, ciência, pesquisa científica, conhecimento aplicado

ABSTRACT

This article discusses the *newness* of big data, arguing that it can be defined in relation to the type of knowledge that is created. It highlights the importance of big data in the research in social sciences, and especially in communication research, despite that research be partly limited by the data sources (often, commercial). Another impediment is that such research is pushing in many directions, but without integrating the new-found knowledge into overall accounts of the role of media in social change. And most of this knowledge is not being produced within the social sciences at all, but rather in the private sector and in policy settings. Big data also raises certain issues in new guises, such as privacy and social vigilance.

Keywords: Big data, social media, science, scientific research, applied knowledge

¹ Este texto corresponde a uma adaptação de capítulo do livro *Social theory after the internet: media, technology and globalization* (UCL Press, 2018). Disponível em: <<https://goo.gl/brs27N>>. Acesso em: 28 jul. 2018. Tradução: Richard Romancini.

^a Concluiu seu doutorado sobre Max Weber na London School of Economics em 1988. Publicou, entre outros livros: *An age of limits: social theory for the 21st Century* (Palgrave Macmillan, 2013) e, com Eric T. Meyer, *Knowledge machines: digital transformations of the sciences and humanities* (MIT Press, 2015). Orcid: <https://orcid.org/0000-0002-4229-1585>. E-mail: ralph.schroeder@oii.ox.ac.uk

O CONHECIMENTO SOBRE A sociedade tomou, recentemente, uma nova direção com os inovadores estudos da mídia digital. Os exemplos incluem análises do Twitter, da Wikipédia, do Facebook, do Google e o uso de *smartphones*. Esses estudos foram possíveis graças à disponibilidade de quantidades inéditas de dados – embora estes sejam muitas vezes limitados pelo fato de terem origem em empresas do setor privado. Outra limitação é que eles não estão bem integrados às análises de outras mídias ou à pesquisa social existente. As novas mídias não se enquadram nas duas principais tradições de estudo da comunicação – a comunicação de massa *versus* a comunicação interpessoal. Por exemplo, se os usuários do Facebook compartilham notícias em um grupo de amigos, isso será diferente da troca de informações entre duas pessoas, mas também diferente da transmissão de notícias. Ou, novamente, se as *hashtags* do Twitter forem criadas para eventos específicos, elas criarão um público em torno do evento em vez de fazerem parte da comunicação de um para um ou da transmissão para muitos. Os mecanismos de busca que adaptam os resultados a usuários ou grupos específicos (o efeito de “filtro bolha”, de Pariser [2011]) são outro exemplo. Mais um exemplo bastante diferente é a Wikipédia, que, apesar de ser um dos sites mais populares, não se enquadra facilmente nas categorias existentes de fontes de informação, como aquelas produzidas por pesquisadores acadêmicos ou por profissionais de mídia. A Wikipédia é uma fonte proeminente de pesquisa em *big data* que, ao contrário de muitas fontes comerciais, está disponível para os pesquisadores de maneira aberta.

Quando examinarmos exemplos da pesquisa em *big data*, portanto, será importante ver se é possível ir além dos modelos de massa *versus* modelos interpessoais. Independentemente de isso ser possível, a pesquisa em *big data* permitiu avanços significativos no conhecimento sobre o papel da mídia na sociedade, remodelando a ciência social em direção a abordagens mais quantitativas. Um argumento comum contra essa visão é que não há nada novo no *big data*; os dados sempre foram usados de várias maneiras e o *big data* é simplesmente uma extensão dessa tendência. Este trabalho parte dessa visão: o *big data* iniciou uma nova direção no tipo de conhecimento disponível – mas com implicações diferentes para a pesquisa e para as aplicações práticas na sociedade de maneira geral. As duas dimensões estão relacionadas, mas têm consequências bem diferentes: o avanço do conhecimento cumulativo em um caso e o maior controle sobre os públicos ou os consumidores – por vezes, os cidadãos – no outro. Este artigo cobrirá dois tópicos: como a pesquisa, e particularmente a pesquisa de mídia, avançou com o *big data*; e como a aplicação do conhecimento de *big data* está moldando os usos da mídia e tem implicações sociais mais amplas. Mas a primeira tarefa, especialmente à luz do argumento do *nada de novo*, é abordar a espinhosa questão de definir *big data*.

DEFININDO *BIG DATA*

As técnicas de quantificação nas ciências sociais e no *marketing* em um sentido amplo (incluindo a logística) têm uma longa história e, nesse sentido, não há realmente nada de novo no *big data*. Beniger (1986), por exemplo, documentou o surgimento de métodos científicos e de técnicas psicológicas para o controle dos mercados de consumo pelo menos desde meados do século XX. Para as ciências sociais, Porter (2008) delineou como a relação entre pesquisa acadêmica e comercial aumentou e diminuiu em termos de proximidade, também quanto à coleta de dados. Assim, o “*big data*” é algo novo? “*Big data*” pode ser definido como uma pesquisa que representa uma mudança na escala e no escopo do conhecimento sobre um determinado fenômeno². Note que essa definição não se baseia em *tamanho per se*, mas no tamanho relativo do objeto dado ou do fenômeno em investigação – em que o objeto pode possuir tantos dados que, anteriormente, sua coleta e análise em uma escala suficiente era difícil, impraticável ou impossível. Portanto a definição diz respeito a como a pesquisa de *big data* vai além da pesquisa anterior sobre tipos específicos de objetos (voltaremos a isso em breve).

² O que se segue é baseado em Schroeder (2014a; 2014b).

Mas o que são *dados*? Em termos de conhecimento científico – ou confiável e objetivo –, os dados têm três características: a primeira é que os dados pertencem (no sentido ontológico, não jurídico) ao objeto ou fenômeno sob investigação; é um material coletado sobre o objeto da pesquisa. A segunda é que os dados existem antes da análise. Como Hacking (1992: 48) coloca, a visão de que “todos os dados são de natureza interpretada” é enganosa: “os dados são elaborados, mas como uma boa primeira aproximação, a feitura e a tomada vêm antes da interpretação”. Ele acrescenta, “é verdade que rejeitamos ou descartamos dados putativos porque eles não se encaixam em uma interpretação, mas isso não prova que todos os dados sejam interpretados” (loc. cit.). Ele também distingue dados de outros aspectos relacionados ao processo científico, como a calibração de instrumentos para a medição de dados. E a terceira, os dados são a unidade de análise útil mais divisível ou atomizada.

Além de identificar a novidade do *big data* digital, essa definição de dados tem implicações sobre como os avanços nas ciências sociais podem ser mensurados. Pressupõe uma epistemologia realista e pragmatista (Id., 1983), porque a definição exige que exista um objeto *lá fora* (realismo) sobre o qual se tenha obtido conhecimento mais útil ou efetivo (pragmatismo). Hacking (Ibid.: 146) define a ciência como a “aventura da interligação entre representar e intervir”; novamente, essa é uma explicação realista e pragmática da relação entre o conhecimento científico e os mundos físico ou natural. Desenvolvi as ideias de Hacking (Schroeder, 2007: 9) argumentando que a tecnologia é “a aventura da

interligação entre refinar e manipular” o mundo por meio de instrumentos ou ferramentas físicas.

Com essas definições em mente, pode-se reconhecer que ferramentas mais efetivas (por exemplo, o poder computacional) se tornaram disponíveis em relação a fontes de dados em grande escala e prontamente manipuláveis, ligando assim a maior disponibilidade de fontes de dados aos avanços tecnológicos ou às ferramentas que podem manipulá-los. Aqui, como em outros lugares, a quantificação ou a matematização, além da fonte de dados, são um pré-requisito fundamental para o crescimento do conhecimento científico, com as ferramentas (*algoritmos*) proliferando e os instrumentos para utilizá-las (computadores) se tornando mais eficazes. Outro aspecto relacionado à definição de tecnologia é que as novas tecnologias frequentemente saem do laboratório e passam a fazer parte da vida cotidiana na forma de dispositivos de consumo ou de seus usos. Isso, claro, é exatamente o que aconteceu com as mídias sociais, as ferramentas de busca e outras mídias digitais: as novas tecnologias se transferiram do *laboratório* para dispositivos de consumo ou de serviços digitais.

Essas são ideias filosóficas sobre o que o conhecimento científico e as tecnologias fazem, ou como eles fornecem conhecimento e mudam o mundo. A chave aqui é que essas ideias fornecem um esclarecimento sobre as implicações do conhecimento baseado em dados: uma concepção *realista* considera os dados disponibilizados a partir de uma fonte no mundo e em uma escala que é diferente da que estava disponível anteriormente em relação a objetos similares. Aqui podemos pensar, como exemplos concretos, nos dados que temos sobre interações sociais no Twitter, no Facebook ou na Wikipédia, no caso em que todos os dados (ou grandes amostras) dessas mídias estão disponíveis, e como isso se compara com os dados disponíveis sobre registros de telefone fixo, ou sobre assistir à televisão, ou sobre as cartas físicas e seu conteúdo, remetentes e destinatários.

Essa visão da ciência e dos dados tem várias consequências para a natureza do conhecimento das ciências sociais e para os usos planejados dos diferentes tipos de conhecimento. Uma *representação* mais poderosa implica uma compreensão maior do fenômeno, e a *intervenção* ocorre tipicamente em relação à tentativa de fazer mudanças no mundo natural – ou aqui, no mundo social. Para a pesquisa de mídia em *big data*, o *mundo* do fenômeno no qual se intervém inclui infraestruturas digitais, como as mídias sociais e as ferramentas de busca que têm acesso aos dados digitais das pessoas. Mas os pesquisadores acadêmicos normalmente não têm interesse em – nem a possibilidade de – intervir nessas tecnologias digitais (ou com elas), exceto, por exemplo, quando eles podem controlar o ambiente do qual os dados digitais são coletados (como, digamos,

em um experimento). No entanto, esse tipo de intervenção é precisamente no que o Facebook e outras empresas (e, às vezes, os governos) que têm acesso a esses dados e a esses sistemas tecnológicos estão interessados.

O poder da pesquisa em *big data*, ao menos no contexto acadêmico, deriva de sua cientificidade. Mas existe uma diferença entre avançar o conhecimento e mudar ou controlar o ambiente físico ou social. Por exemplo, em um experimento, a *população de laboratório* pode ser submetida a diferentes condições; aplicar as descobertas no contexto de um site de rede social ou ferramenta de busca requer um tipo diferente de esforço, incluindo acessar essas mídias digitais. E pode-se notar, de imediato, que estudos experimentais sobre mídia digital (como o do *contágio emocional* do Facebook, que será discutido em breve) muitas vezes não são científicos em um aspecto crucial: os dados não estão disponíveis, de maneira aberta, para replicação – uma condição crucial para a ciência. A manipulação de fenômenos é, portanto, um exercício mais prático e aplicado, uma maneira mais eficaz de controlar aspectos específicos do mundo físico ou social para determinados propósitos (um exemplo amplamente conhecido é mudar a cor na página inicial do mecanismo de pesquisa Google para ver se isso melhora a *experiência do usuário*).

Novamente, torna-se claro a partir dessas considerações que há possibilidades bastante diferentes na pesquisa acadêmica e na comercial. Os cientistas sociais acadêmicos estão engajados na pesquisa para produzir conhecimento generalizável sobre o comportamento humano, não (em sua maior parte, ou em primeira instância) para intervir nele. Dentro do setor privado ou em outros contextos aplicados, como as políticas públicas, por outro lado, os pesquisadores e aqueles que usam esse conhecimento (como profissionais de marketing e anunciantes) querem intervir no comportamento humano. Assim, os usos de *big data* para aplicações específicas, influenciando o comportamento das pessoas, não são neutros – mesmo que o conhecimento gerado para esses propósitos seja neutro ou científico. O conhecimento que usa dados digitais se aplica a seres humanos tratados como material abstrato regido por certas regularidades estatísticas, enquanto o conhecimento gerado para o uso em sistemas tecnológicos para influenciar comportamentos está muito mais ligado ao contexto de tempos, lugares, populações e propósitos particulares.

Daí também a menor eficácia do conhecimento (no sentido de *representar*) comercial ou aplicado em oposição ao conhecimento científico: os dados são proprietários (enquanto os dados científicos devem ser abertos), as populações não são representativas (o conhecimento científico deve ser generalizável) e o objetivo é o conhecimento que deve aumentar as *vendas* ou algo do tipo (o objetivo do conhecimento científico é representar de maneira mais eficaz o

mundo, e o da tecnologia é refinar o controle sobre o ambiente). De novo, o conhecimento do setor privado é, portanto, tipicamente não cumulativo. Mas, como acontece com todas as tecnologias, e de acordo com a definição dada anteriormente, ele pode aprimorar o *refino e a manipulação*, como acontece com qualquer tecnologia baseada em conhecimento científico. Há, por isso, uma divisão entre os usos de *big data* em análises acadêmicas ou científicas em comparação com os usos de *big data* em ambientes comerciais, governamentais e em outros contextos aplicados. Na pesquisa acadêmica e na ciência, o *big data* é usado para gerar conhecimento abstrato, sem prescrever como usar esse conhecimento para mudar o comportamento. Em ambientes aplicados, o inverso é verdadeiro: o conhecimento das ciências sociais é gerado para que possa ser aproveitado para mudar o comportamento.

Esse aspecto pode ser diretamente relacionado à definição de dados utilizada aqui: em ambientes onde os dados não são obtidos de fontes *brutas* (o mundo físico), eles são tratados *como se* fossem brutos (em relação ao comportamento humano). Considere os dados do Twitter: quando os *tweets* são analisados, isso geralmente é feito pela contagem da frequência de palavras ou mensagens enviadas entre contas *como se* fossem unidades sem contexto. Ou seja, as contas do Twitter são tratadas como se pertencessem a uma pessoa ou organização (embora isso não seja necessariamente o caso – pense nos *bots*), e as interações entre unidades são tratadas como equivalentes (o que, novamente, pode não ser verdade – pense nos usuários com várias contas). Ou, novamente, a frequência das palavras é tratada como indicativo de certo sentimento ou intenção, sem levar em conta o fato de que as palavras podem ser usadas de maneiras diferentes – por exemplo, ironicamente³. Desse modo, os dados do Twitter são tratados como se fossem unidades abstratas e pudessem ser manipulados cientificamente. E se os problemas dessa abordagem (por exemplo, um usuário com várias contas) são vistos como insignificantes, considerar a abordagem baseada em dados como científica é justificável.

³ Para um panorama do uso da mídia social para a análise de sentimento, ver Thelwall, Buckley e Paltoglou (2012), que também apontam para o problema da *ironia*.

PROMOVENDO O CONHECIMENTO ACADÊMICO SOBRE A MÍDIA DIGITAL

Já falamos de pesquisas de *big data*, incluindo as descobertas sobre os usos de ferramentas de busca, pesquisas na Wikipédia e clusters de atenção na web (Schroeder, 2018: 101-125). Há muitos mais (Ekbia et al., 2015), portanto será suficiente dar mais dois exemplos, relacionados aos usos do Facebook e do celular⁴. Então poderemos nos voltar para a questão mais ampla de como esse novo conhecimento se encaixa na ciência social e das suas implicações mais gerais.

⁴ O que se segue é parcialmente baseado em Schroeder (2016).

Um questionamento sobre o Facebook é se os *amigos* que compartilham conteúdo também compartilham perspectivas ou ideologias políticas. Bakshy, Messing e Adamic (2015) investigaram essa questão em mais de dez milhões de usuários estadunidenses do Facebook e descobriram que os amigos do Facebook são ideologicamente bastante diversos, em parte porque seus laços refletem redes off-line, como a família, a escola e o trabalho – em contraste com usuários do Twitter que compartilham interesses ou tópicos, mas não necessariamente laços off-line, e são, portanto, mais ideologicamente polarizados (Conover et al., 2011). Logo, como as variadas mídias sociais são usadas em diferentes contextos, combinar seus usos políticos e usos para a socialização poderia ser uma maneira de pensar sobre como as novas mídias são pessoais, mas, ao mesmo tempo, vão além das mensagens dessa esfera.

O mesmo vale para outro estudo do Facebook (Settle et al., 2013) que analisou o conteúdo das mensagens do Facebook e, em particular, as *atualizações de status*, que a maioria (73%) dos usuários do Facebook faz pelo menos uma vez por semana (Hampton et al., 2011). Também pode se acrescentar aqui que, em um momento anterior, pelo menos 60% dos estadunidenses usavam o Facebook e 66% destes o usavam para atividades cívicas ou políticas (Rainie et al., 2012). Settle e seus colegas (2013) puderam coletar conteúdo de natureza política em relação à eleição presidencial dos Estados Unidos em 2008 e ao debate sobre a reforma da saúde em 2009. Os autores mostraram que as mensagens políticas acompanham de perto os principais eventos (no caso da eleição, estes incluíam as convenções partidárias, a eleição em si e a posse; no caso do debate sobre assistência médica, houve uma mudança no uso do termo *reforma da saúde* para *Obamacare*). Dado o grande número de estadunidenses e de outros que trocam mensagens políticas no Facebook, essa é uma maneira interessante de rastrear mudanças na atenção. Também é intrigante ver o ajuste relacionado ao argumento feito aqui de que há um espaço de atenção limitado para a agenda política. Ao mesmo tempo, o Facebook é apenas um dos vários sites de redes sociais, mesmo que seja o dominante nos Estados Unidos e na maior parte do mundo; portanto esses resultados precisam fazer parte de uma análise mais ampla. E, claro, os dados para esse estudo, assim como os para o de Bakshy, Messing e Adamic (2015) e para o estudo de Ling et al. (2014), são acessíveis apenas para os pesquisadores que trabalham com essas empresas.

A partir daqui, podemos recorrer ao exemplo dos telefones celulares. É claro que houve muitos estudos sobre a frequência com que as pessoas se conectam umas com as outras, e em que variedades de distâncias, via telefones (Fischer, 1992) e telefones celulares (Licoppe, 2004). No entanto, o contato regular mais frequente por meio de telefones celulares, tanto por texto quanto por voz, é feito

por um pequeno número de pessoas (Ling et al., 2014). Analisando registros de chamadas móveis da operadora móvel dominante na Noruega por um período de três meses, Ling e colaboradores (Ibid.) descobriram que a maioria das conexões é feita com um pequeno grupo de pessoas que estão próximas. Desse modo, como Fischer (1992) em relação ao telefone fixo, eles contradizem a ideia frequentemente debatida de *morte da distância* ou de uma *aldeia global* (Ling et al., 2014). Eles também fazem uma distinção entre as populações rurais, em que “a maior proporção de chamadas é para aqueles que estão a menos de um quilômetro de distância”, e as urbanas, em que “a preponderância de chamadas é para pessoas a mais de um e menos de 24 quilômetros de distância” (Ibid.: 288). Essa é uma descoberta contraintuitiva, já que poderia ser esperado que as ligações das pessoas rurais fossem para pessoas mais distantes e vice-versa. No entanto, se pensarmos nas distâncias que os indivíduos urbanos normalmente perfazem, inclusive para chegar ao trabalho, e nas idades das populações urbanas e rurais, as descobertas fazem sentido (e podem ter implicações para os transportes, para as políticas de cobrança das operadoras de telefonia móvel, entre outras coisas).

Como os telefones celulares, assim como os *smartphones*, estão sendo usados para acessar a internet, no entanto, ainda não é bem compreendido, apesar da disponibilidade de grandes quantidades de dados. Talvez a diferença entre eles esteja ficando embaçada, embora Napoli e Obar (2015) argumentem que os usuários de telefones celulares representam uma “subclasse” por causa da funcionalidade muito mais limitada da internet móvel em comparação com o acesso via computadores de mesa ou laptops. Eles desenvolveram esse argumento com base em uma revisão de estudos que mostram que os computadores de mesa ou laptops são mais úteis para a criação de conteúdo e tarefas complexas, enquanto os *smartphones* o são para atividades passivas e limitadas (Ibid.). No entanto, isso vai contra o senso comum, uma vez que os jovens de países de alta renda, em particular, usam *smartphones* para fazer uma ampla variedade de coisas.

Porém, é importante lembrar, como Donner (2015) aponta, que esses ricos e altamente habilidosos usuários de *smartphones* são uma minoria ao redor do mundo. Além disso, os usuários afluentes quase invariavelmente também têm acesso à internet por meio de laptops e de outros dispositivos, como tablets, além de terem conexões de alta largura de banda a um custo (relativamente) baixo. Assim, mesmo que os usos de seus *smartphones* sejam restritos, eles podem combinar isso com tarefas mais complexas em outros dispositivos. Em contraste, os usuários de países de baixa renda no sul da Ásia e na África têm uma *mentalidade controlada*, com escassa largura de banda, que é (relativamente) cara e, portanto, usada de modo mais econômico. Esses usuários tendem a “mergulhar e sorver”, em vez de “surfear e navegar”, como diz Donner (Ibid.),

e provavelmente também têm habilidades muito mais limitadas e usos muito mais restritos de seus *smartphones*.

Essa nova brecha digital pode diminuir com o tempo, mas a diferença entre a grande maioria dos usuários que usam apenas *smartphones* e uma minoria de usuários com múltiplos dispositivos também deve continuar sendo uma divisão profunda por muitas décadas. E os contextos de uso são importantes se pensarmos em como o pagamento on-line e outros serviços são muito mais comuns na China do que nos casos de outros países. Na Índia e na China os *smartphones* são a maneira dominante de acessar a internet. Colocando de outra maneira, a grande maioria dos internautas em escala global acessará a internet majoritariamente através de um dispositivo móvel e pode nunca (ou raramente, exceto pelo trabalho) ter acesso a um laptop ou a um computador de mesa. Esses usos difundidos precisam ser colocados nos contextos de socialização política e cotidiana. Muitos dados serão disponibilizados a partir desses usos, mas isso não evitará a necessidade de um conhecimento desses contextos teoricamente embasado⁵.

⁵ Para um exemplo de pesquisa comparada entre Índia e China que remete a essas dificuldades, ver Haenssger (2016).

OS USOS E OS LIMITES DO *BIG DATA* NAS CIÊNCIAS SOCIAIS

As abordagens de *big data* dirigem a ciência social (e outras áreas do conhecimento) para um rumo mais quantitativo e estatístico – e, portanto, mais científico. Isso tem sido um grande benefício para a ciência social, mas também é importante esclarecer algumas limitações. Mais uma vez, a ciência social quantitativa não é novidade (Porter, 2008) e nem os esforços para introduzir ferramentas e fontes de dados digitais na pesquisa (Meyer; Schroeder, 2015). O que há de novo na pesquisa de *big data* são as fontes de dados, que fornecem acesso a dados prontamente manipuláveis (computáveis). Os dados das ciências sociais, no passado, eram difíceis de ser obtidos, principalmente por meio de entrevistas face a face ou pesquisas por correio e telefone, e os dados digitais costumam ser repletos de dificuldades no caso de dados confidenciais e/ou proprietários. Ainda assim, a disponibilidade de dados é uma pré-condição para o crescimento do conhecimento científico da sociedade: os dados fornecem um meio independente para checar ou verificar (ou falsificar) os resultados; são a matéria-prima que permite que os pesquisadores construam a partir do trabalho uns dos outros. Ter mais desses materiais, sobre um aspecto de nossa vida social, que está crescendo rapidamente, significa que essa área de pesquisa continua a aumentar. Isso ocorre especialmente porque as ferramentas de software para lidar com esses dados também proliferaram recentemente (ver Bright, 2016). As ressalvas são que os dados precisam ser tais que atendam aos critérios da ciência – estando abertos à validação e à replicação.

Ao mesmo tempo, a pesquisa em *big data* está ainda em grande medida numa fase de alta incerteza sobre suas tarefas e a baixa dependência mútua (Whitley, 2000): isto é, os pesquisadores estão explorando muitos novos caminhos, frequentemente sem uma noção de como essa pesquisa pode contribuir para a cumulatividade (Rule, 1997). Em outras palavras, e se tivermos uma visão geral do conhecimento em ciências sociais, o gargalo atual ocorre porque há uma proliferação de estudos empíricos gerados com *big data*, mas esses trabalhos se concentram em certos fenômenos facilmente pesquisáveis, que têm dados em abundância ou em que estes são prontamente analisáveis. Isso implica que há uma necessidade de diálogo entre as várias disciplinas, de modo que as descobertas dessa pesquisa possam ser introduzidas em entendimentos mais sintéticos do mundo social. No entanto, essas percepções baseadas em novas fontes de dados também são limitadas por outras razões: uma, já mencionada, é que muitos estudos não são generalizáveis (ou não podem ser feitos para isso), porque os dados provêm de redes sociais proprietárias ou comerciais. Isso significa que as descobertas não podem ser replicadas, pois os dados não são acessíveis a outros pesquisadores – ou não se sabe como os dados foram gerados, em primeiro lugar. No entanto, a Wikipédia é um contraexemplo, uma vez que os dados são completamente abertos e replicáveis, e há outros⁶. A ciência social acadêmica prosperará com a proliferação dessas e de outras fontes de dados não proprietários (ver Borgman, 2015; Meyer; Schroeder, 2015).

⁶Ver, por exemplo, os conjuntos de dados no Dataverse, disponível em: <<https://goo.gl/qPYS7j>>. Acesso em: 9 ago. 2018.

Uma segunda limitação do *big data* é que as descobertas em contextos aplicados ou comerciais são frequentemente de significância limitada, pois visam objetivos práticos de curto prazo, como quando os dados das mídias sociais são analisados para fins de marketing. Isso significa que é improvável que as descobertas se apliquem além de uma campanha de marketing específica ou a determinada população que seja o alvo de certos produtos vendidos (e a questão de saber se as descobertas são científicas provavelmente não será indagada em nenhum caso). Uma terceira limitação: os estudos podem ser restritos porque a fonte dos dados digitais cobre apenas uma parte da população mundial, mesmo que seja uma grande parte, por razões de linguagem ou censura, ou porque um determinado meio digital é apenas um entre outros populares. E é apenas o contexto da variedade de mídias usadas que revela seu significado: a mídia deve ser colocada no contexto do (sub)sistema do qual ela faz parte. Uma quarta razão, já mencionada, é a de que muitas fontes de dados digitais estão sendo investigadas em diferentes direções, sem uma noção de como as descobertas se encaixam no quadro mais amplo da comunicação ou da pesquisa social; em outras palavras, há uma falta de teorias da mídia digital. As duas primeiras limitações são problemas práticos e a terceira diz respeito ao escopo do estudo.

No entanto, a quarta, que é uma questão de esforço em teorização, síntese e integração de descobertas, pode ser superada.

Os estudos de *big data* também precisam de novas teorias porque os usos da mídia digital estão mudando rapidamente. As mídias digitais estão substituindo as mídias analógicas tradicionais que estão em constante declínio. A mídia digital afasta-se da comunicação interpessoal (uma para um) e da de massa (um para muitos) e provoca uma interação de níveis entre as duas – como quando o conteúdo é compartilhado no Twitter ou Facebook, ou os resultados das ferramentas de busca são adaptados a um grupo particular. Também vemos que as teorias da mídia deixam de lado a busca e o acesso às informações. As novas mídias digitais, portanto, adicionam uma nova camada – a mídia tradicional não pode ser ignorada – à forma como o mundo está sendo mediado, e dão outro passo no processo em curso, pelo qual as tecnologias nos prendem mais intimamente à informação e uns aos outros.

Podemos adicionar outra limitação em termos de como esse conhecimento é aplicado: embora a ciência social esteja fazendo grandes avanços devido à disponibilidade de *big data*, essa pesquisa tem impacto principalmente em certas áreas das políticas. Isso porque, como já mencionado, a ciência social acadêmica é apenas fracamente acoplada à transformação ou à engenharia sociais. Watts (2017) avalia que essa limitação deve ser superada por meio de experimentos frequentes na resolução de problemas sociais, para que o conhecimento possa rapidamente provar seu valor e avançar. Curiosamente, como veremos a seguir, é mais provável que isso aconteça na China, onde há uma tradição desse tipo de experimentação social, e onde as condições sociais – o autoritarismo – podem suportá-la. Quanto às aplicações comerciais, o vasto volume de conhecimento de *big data* no setor de mídia está relacionado à propaganda e ao marketing, assim como o grande volume na política se relaciona ao desenvolvimento de mensagens mais eficazes dirigidas ao eleitorado (ou, na China, ao público em geral). Assim, o conhecimento pode ser visto aqui como parte de uma racionalização em curso – cientificização – de como os consumidores e o público estão sendo medidos e quantificados, para o bem ou para o mal.

Em outras palavras, há um contexto social mais amplo para a pesquisa em *big data*, na qual a pesquisa acadêmica é apenas uma pequena parte do esforço. A pesquisa em *big data* é muito mais difundida no setor comercial e, em menor escala, no governo e em outras organizações, em que é usada para fins práticos – engenharia social, se preferir. O principal efeito, nos Estados Unidos, na Europa e em outros lugares, é que o marketing de consumo está se tornando mais eficaz. Outra área principal de aplicação é a medição da opinião pública. No entanto, novamente, o contexto é importante: na China, por exemplo, esse

⁷ A China possui um ambicioso programa em andamento para criar um sistema de crédito social, usando *big data*, para administrar a sociedade com mais eficiência, embora haja pouca pesquisa (publicada em inglês) sobre esse sistema. Embora as implicações políticas desse sistema sejam, talvez, o que vem primeiro à mente no contexto chinês, pode-se mencionar que, do ponto de vista dos objetivos do governo, esse esforço é apresentado principalmente para melhorar a eficiência econômica (ver Meissner, 2017).

tipo de pesquisa pode ser usado não apenas para obter *feedback* da população, mas também para objetivos de vigilância sistemática (claro, a China não está sozinha nisso, mas as condições prévias para usos mais eficazes desse tipo talvez sejam únicas na China, pelo menos em grande escala). Ao mesmo tempo, é importante não esquecer o fato de que as aplicações comerciais também são centrais na China (Pan, 2017; Stockmann, 2013)⁷.

Esses usos não acadêmicos – engenharia social – da pesquisa em *big data* continuam em expansão, trazendo benefícios (marketing, governança) e perigos (vigilância, manipulação). Além disso, o contexto de mais gestão da mídia – especialmente durante as eleições – pode ser visto como vantajoso para as elites políticas (aqueles que tentam se eleger, ou para governar fora dos períodos das eleições) e uma benção tanto para os eleitores quanto para os cidadãos. Enquanto isso, vale a pena lembrar que, mesmo que os benefícios das novas fontes de dados digitais continuem, devido ao conhecimento científico acadêmico e, inclusive, com o crescente papel das mídias sociais e outras mídias digitais, os resultados serão limitados pela extensão em que dados esclarecem o comportamento do usuário: os dados dizem respeito apenas a determinados comportamentos. E, embora as descobertas sobre mídia digital estejam crescendo, elas também precisam se encaixar no conhecimento mais amplo sobre os usos de mídia e os padrões de interação social das pessoas. A esse respeito, o problema de que as novas mídias digitais muitas vezes não se encaixem nos paradigmas estabelecidos de comunicação de massa *versus* comunicação interpessoal pode ser visto como uma limitação, mas também como uma oportunidade para desenvolver novas teorias a respeito das implicações sociais da informação e da comunicação.

Há uma maneira diferente de destacar essas limitações do conhecimento aplicado com base em *big data*: mesmo se, por exemplo, houver técnicas eficazes de *big data* para comprovar que os relacionamentos do Facebook podem influenciar a maneira como as pessoas votam (Bond et al., 2012), há uma distância entre isso e a capacidade de influenciar como as pessoas realmente votam. Ou, novamente, se pudermos mostrar que os amigos do Facebook compartilham certos gostos culturais (Lewis et al., 2008), isso não é o mesmo que fazer as pessoas efetuarem uma compra com base nesses gostos similares. Em outras palavras, tende a haver um objetivo muito restrito no caso do conhecimento aplicado de *big data*, enquanto na pesquisa acadêmica de *big data* o objetivo é obter o conhecimento mais amplo ou mais generalizável. Por outro lado, o conhecimento aplicado deve ser *acionável* – e isso requer esforço, mas também impõe limitações.

Esses pontos podem ser vistos de uma perspectiva mais abstrata em relação às implicações sociais da ciência. O processo de geração de conhecimento mais avançado invariavelmente produz despersonalização, ou uma abordagem mais

determinista do mundo: na medida em que o mundo é explicado objetivamente, isso não deixa espaço para o conhecimento da individualidade além das leis ou regularidades impessoais. Em relação à legislação, como Mayer-Schoenberger e Cukier (2013) apontaram, o *big data* pode ajudar a minar a ideia de responsabilidade pessoal, e um dos pilares da visão de mundo moderna é a ideia do livre-arbítrio. Mas a questão que eles apontam é muito mais ampla que a legal, já que a pesquisa de *big data* também desafia nossas noções de individualidade e autodeterminação fora do contexto legal: se o objetivo de um estudo do Facebook é ser capaz de prever minha personalidade ou o que farei, isso pode não ser legalmente disruptivo, mas prejudica meu senso de individualidade em um nível pessoal.

Da mesma forma, a própria ideia de determinismo tecnológico ou tecnocientífico – de que meu comportamento pode não apenas ser previsto, mas *manipulado* por uma determinada tecnologia – vai contra autocompreensões e ideias fundamentais sobre como a sociedade opera de acordo com a tomada de decisão individual e coletiva. Além disso, pode-se mencionar que, embora o conhecimento determinístico sobre o comportamento humano possa parecer ameaçador para certos propósitos sociais, esse potente conhecimento também será inevitavelmente necessário – se pensarmos, por exemplo, em mudar os hábitos de consumo de energia das pessoas diante do desafio das mudanças climáticas. Adicionalmente, vale lembrar que não é do interesse das empresas violar a privacidade dos dados das pessoas: as empresas coletam dados pessoais para influenciar nosso comportamento de compra e coisas semelhantes, e os dados são, portanto, um recurso a ser protegido em vez de compartilhado – exceto quando há uma perspectiva de ganho a partir do compartilhamento. Da mesma maneira, os Estados democráticos querem proteger as populações das ameaças e obter conhecimentos mais eficazes para a formulação de políticas e, em alguns casos, *provocar* o comportamento das populações – não necessariamente diminuir suas liberdades (exceto quando a *segurança* está em jogo).

Desse modo, identificar novas fontes de dados também destaca novas oportunidades e perigos decorrentes dessas fontes. Também aponta para os limites das abordagens de *big data*: existem somente tantas fontes quantas são as pessoas que usam os objetos que as fornecem (como mídias sociais ou outros objetos que deixam rastros digitais). Hipoteticamente, uma vez esgotada a utilidade de analisar essas fontes – se todas as possíveis relações sociais cientificamente interessantes no Facebook ou Twitter foram pesquisadas – então haverá retornos decrescentes para o conhecimento científico social – mas não para usos comerciais ou outros usos não acadêmicos de *big data* (embora, similarmente aqui, os usos práticos do conhecimento sejam limitados aos usuários de mídias sociais específicas).

Até agora, discuti principalmente as implicações do *big data* para avanços no conhecimento científico. Quando nos voltarmos para como esse conhecimento é aplicado, bem menos é conhecido, e os efeitos também são muito mais difusos. Aqui também há limites e, sem supervisão regulatória e uso da transparência, estes permanecerão menos claros. Além disso, há benefícios óbvios no uso de *big data* para atingir os consumidores. O Facebook ganha uma vantagem competitiva em conhecer meu círculo social, ou a Google em saber minha localização e as consultas de pesquisa que revelam meus interesses. As empresas usam esse conhecimento e podem querer manipular as pessoas para determinado comportamento – se isso levar a maiores lucros. Mas o Facebook e a Google não têm interesse em ser publicitários ou editores de notícias, por exemplo: como outras empresas, quanto menos regulamentação, melhor – e ser um veículo de notícias pode exigir mais esforço em termos de conformidade regulatória.

As pesquisas baseadas em dados no mundo comercial (e, em menor grau, no governo e no setor sem fins lucrativos) são tipicamente realizadas com objetivos limitados: se certas correlações – por exemplo, em comportamentos de compra – são encontradas, então essas correlações podem ser usadas para incentivar novas compras; ou se determinados pontos críticos do crime são identificados, os recursos policiais podem ser realocados para neutralizá-los (Eagle; Greene, 2014). Mas embora os dados possam ser usados para direcionar indivíduos específicos (para voltar à discussão anterior do *determinismo*), pode não ser possível mudar o comportamento deles, mesmo que seja possível *estimular* as populações. No entanto, em muitos casos, pode ser suficiente que essas correlações funcionem – pelo menos em uma proporção lucrativa ou útil de casos – para que esse conhecimento seja recompensado.

Os aspectos éticos e legais da pesquisa em *big data* têm estado bastante em foco nos debates contemporâneos, pois exigem respostas regulatórias e políticas urgentes. O que é negligenciado com esse enfoque (como acontece com grande parte do avanço tecnocientífico) é a *sutil mudança* em longo prazo em termos dos efeitos de um conhecimento mais eficaz, derivado nesse caso das fontes de *big data*, na sociedade. Um corpo crescente de conhecimento baseado em dados digitais está fadado a ter implicações sociais importantes, mas o faz enquanto conhecimento, em um nível que é quase imperceptível porque é lento e difuso. Para os indivíduos e os formuladores de políticas, é importante responder a questões imediatas e reconhecíveis em relação à proteção de dados, mesmo que as consequências sociais mais amplas do crescimento do conhecimento sejam menos tangíveis e menos claramente identificáveis.

O avanço *insidioso* do conhecimento científico também pode ser visto em termos de como o comportamento do usuário é afetado, e como as pessoas são,

portanto, moldadas por esse conhecimento⁸. Esse tipo de conhecimento avança inexoravelmente em direção a um novo tipo de onisciência – onisciência, no sentido de que tudo que pode ser conhecido via traços digitais será desvendado e levará a uma compreensão abrangente do comportamento humano, das ações individuais às interações em nível global (por exemplo, Eagle; Greene, 2014), sobre uma parte cada vez mais importante da vida cotidiana – a vida on-line. Seria fácil, com base nessas reflexões, tirar conclusões bastante apocalípticas, que são típicas na mídia e em algumas respostas acadêmicas (Grimmelman, 2014, contém exemplos), especialmente em vista da influência bastante ampla que tem sido discutida. No entanto, também é importante colocar esses efeitos em perspectiva: os efeitos desse tipo de pesquisa limitam-se a usos para a vantagem comercial, para a doutrinação política em regimes autoritários e para os avanços nos processos democráticos e da ciência social acadêmica – e os limites a estes que acabamos de discutir. Em vez de exagerar os perigos, é mais correto apontar para a *mudança sutil*, que é difusa, penetrante e, também, um processo amplamente invisível.

O combate à *mudança sutil*, da mesma forma, requer traçar linhas na areia: onde a manipulação do comportamento do usuário deve ser regulada de modo a ser transparente ou sujeita a consentimento explícito (e significativo)? Quando os usuários dependem de um serviço na medida em que é uma parte essencial da infraestrutura social e requer regulamentação? Quando a ciência social acadêmica trabalha em apoio a aplicações comerciais que são eticamente inaceitáveis? Essas perguntas serão (esperançosamente) cada vez mais enfocadas, e fornecem uma base diferente para a potencial regulamentação – ou para a abertura de dados em benefício do bem público – do que as questões mais restritas de ética em pesquisa ou lei e privacidade em casos individuais. E essas perguntas podem ser ilustradas com o estudo do Facebook do *contágio emocional*, ao qual agora podemos nos voltar.

OS ADMIRÁVEIS MUNDOS NOVOS DO FACEBOOK

Os debates sobre *big data* geralmente concentram-se no Estado e no mundo orwelliano de vigilância por meio do monitoramento de telefones celulares ou da internet. Esses debates foram motivados pelas revelações de Edward Snowden e Julian Assange (WikiLeaks) no início de 2010 sobre vigilância por agências de inteligência, e continuam desde então. Mas em vez do romance *1984*, de Orwell, e da vigilância do Estado, uma visão mais relevante é *Admirável mundo novo*, de Huxley, no qual as empresas e os governos são capazes de moldar a mente das pessoas, fazendo isso de uma maneira que os usuários, consciente ou

⁸ Pode-se mencionar que a visão de *big data* aqui, como em todo este artigo, em termos de seu impacto como conhecimento científico, se afasta de outras abordagens: a nova tecnologia, em vez do *capitalismo*, determina a forma como os dados são usados. O entretenimento e os serviços on-line competem pela atenção e pelos costumes das pessoas, e a principal questão para as ciências sociais é como as mídias digitais modificam os hábitos cotidianos delas. Para os marxistas (Castells, 2009; McChesney, 2013), o capitalismo é explorador, aumenta as desigualdades e as corporações tentam evitar a regulação. Para os liberais e os economistas, o capitalismo e as empresas de mídia geram *valor*, melhoram as vidas e produzem mais prosperidade. Mas há uma explicação mais simples e precisa em relação a como os *big data* estão mudando as indústrias de mídia: as corporações de mídia que operam em mercados competitivos pressionam por condições nas quais tentam aumentar seus lucros (incluindo, quando possível, em relação a isso, a regulamentação estatal e as infraestruturas de tecnologias de informação e comunicação) e otimizam e buscam essas condições independentemente das políticas (incluindo na China) e de como os tipos de mercados criam a desigualdade ou a igualdade digital. Assim, novamente, o conhecimento científico é parte de uma contínua cientização ou racionalização do mundo, para o bem ou para o mal.

⁹ Ver Wilson, Gosling e Graham (2012) para o consentimento e a privacidade na pesquisa do Facebook.

inconscientemente (e pode não ser fácil perceber a diferença), adotam e aceitam isso. O Facebook, ao justificar o experimento de *contágio emocional*, argumentou que seu objetivo principal era melhorar as experiências dos usuários – e quem não gostaria de ter sua experiência no Facebook melhorada? Ou, para usar um exemplo diferente, quem poderia se opor a ser *estimulado* on-line pelo governo para se sentir melhor sobre o pagamento de impostos?

Aqui, podemos tomar o estudo de *contágio emocional* do Facebook (Kramer; Guillory; Hancock, 2014) como um exemplo, pois provocou um amplo debate, embora o argumento se aplique igualmente à Google e a outras mídias digitais que geram muitos dados. Os pesquisadores fizeram um experimento dividindo os usuários do Facebook, em geral, quase 700 mil deles, em dois grupos: um teve mais palavras positivas introduzidas em seus *feed* de notícias, o outro grupo foi exposto a mais palavras negativas. Os pesquisadores então mediram se esses usuários, subsequentemente, à luz dos dois *tratamentos*, publicaram mais palavras positivas ou negativas. Eles descobriram que, de fato, o fizeram, confirmando o *contágio social*. Para nossos propósitos, o ponto principal é que o estudo mostrou experimentalmente que, sem o conhecimento dos usuários do Facebook, seu humor (*contágio emocional*) poderia ser manipulado. Deve-se notar imediatamente que *manipular* é usado aqui em um sentido neutro, equivalente a dizer *mudar o que as pessoas fazem* ou mudar o ambiente físico ou social, que pode ser negativo ou positivo. As questões éticas e legais do estudo podem ser deixadas de lado aqui, exceto ao notar que alguns argumentaram que as críticas ao estudo e à publicidade negativa que ele gerou poderiam levar a que esse e outros experimentos fossem feitos de maneira “subterrânea” (Meyer, 2014); isto é, sem o escrutínio do olhar público da ciência, por causa da má noticiabilidade e, consequentemente, também da visão negativa do Facebook resultante. E, de fato, em consonância com o argumento sobre a pesquisa comercial, uma das implicações sociais desse tipo de pesquisa é que a extensão em que ela possa ir para a clandestinidade provavelmente não será conhecida no todo. Novamente, é por isso que é importante distinguir entre a pesquisa científica aberta e o conhecimento aplicado. Mas as consequências sociais da pesquisa aplicada são que ela se avança sutilmente e afeta cada vez mais a vida cotidiana, mesmo que haja também razões – como a controvérsia provocada pelo estudo no Facebook – para permanecer oculta.

Além desse *encobrimento*, também há pouca pesquisa sobre como os usuários estão respondendo às utilizações de seus dados. Eslami e colaboradores (2015; 2016) perguntaram aos usuários do Facebook sobre como essa rede usa dados para moldar as informações que eles veem. Eles eram, em grande medida, ignorantes ou equivocados sobre isso. E não está claro como

eles poderiam ser mais bem informados, ou se o Facebook poderia ou deveria fazer mais para isso. O mesmo vale para outras mídias digitais – o Facebook é simplesmente apontado aqui como um exemplo proeminente: os usuários de mídia digital não sabem como seus dados estão sendo usados para moldar as notícias, propagandas e mensagens que recebem de outras pessoas. Ainda assim, eles se adaptam a esse estado de coisas, assim como os consumidores em geral em relação a como seus dados estão sendo usados por empresas de mídia (Turow, 2017)¹⁰. Mas vale a pena lembrar que eles não gostam da manipulação sub-reptícia de seus pensamentos e sentimentos, considerando-a não como uma troca desigual para a obtenção de serviços gratuitos (como as empresas e os economistas fazem), mas como uma prática ofensiva. Ao mesmo tempo, como Turow (2017) descobriu em levantamentos com estadunidenses, as pessoas estão aceitando ou se resignando a esse controle sobre as suas vidas.

¹⁰Ver também Rule (2007).

O público e os consumidores estão se adaptando aos usos do *big data* e como isso está moldando a atenção comercial e a opinião política, mesmo que eles continuem gerando controvérsia quando se tornam conhecidos. Novamente, apesar de seu poder crescente, há limites para o *big data*, com base na quantidade de atenção que a mídia digital atrai e na medida em que essa atenção pode ser aproveitada para influenciar consumidores e cidadãos. Ainda assim, como vários estudos (O’Neil, 2016; Pasquale, 2015) apontaram, os usos comerciais (e em certa medida de campanha política) de *big data* são problemáticos quando há pouca ou nenhuma transparência ou responsabilidade. Os exemplos disso são quando os comportamentos são moldados contrariamente aos direitos individuais, ou quando a regulação não acompanha os usos socialmente prejudiciais (como quando o conhecimento sobre indivíduos ou populações é usado para direcioná-los inconscientemente, enganando-os sobre benefícios econômicos).

Concentrei-me aqui em *big data* e mídia digital, mas o *big data* é usado para mais do que moldar a atenção: ele também molda as pontuações de crédito, fornece perfis psicológicos para os empregadores, mede o desempenho educacional, segmenta os compradores e muito mais. Esses estão além do escopo de trabalho sobre a mídia, no qual o foco principal é em notícias e cidadãos, a busca de informações, entretenimento e sociabilidade. Mas as pessoas confiam cada vez mais em transações de mercado de bens e serviços on-line, e, à medida que a atividade econômica on-line penetra cada vez mais na vida cotidiana, os usos da mídia digital também criam mercados para dados. Esses mercados para os dados, por sua vez, em vez de fornecer *conteúdo* como tal, podem ser vistos como mercados de nível superior (ou piramidados), mercados não para atenção, mas para dar forma à atenção. Tais mercados são baseados em comportamentos de busca e socialização de informações e fornecem uma vantagem não apenas

para empresas de mídia, mas também para empresas como Amazon, Uber e LinkedIn, que não são estritamente empresas de mídia, mas empresas que utilizam dados. Aqui, também, novas divisões estão sendo criadas, e as empresas com acesso a dados estão conquistando mercados às custas de outras empresas, e com esses mercados também adquirem maior controle sobre os consumidores.

SEGMENTAÇÃO DE PÚBLICOS E OS USOS E LIMITES DO *BIG DATA* NA VIDA COTIDIANA

A competição pela atenção depende cada vez mais de análises ou métricas em um ambiente de mídia on-line de *alta escolha*. Mas as análises moldam as notícias lidas ou visualizadas? Existe ainda uma limitada pesquisa sobre as consequências dessa mudança, e um problema é que as pessoas não usam exclusivamente fontes on-line de notícias (e nem as notícias são originadas apenas na mídia on-line). Outra complicação é a mudança para o consumo de notícias móveis, sobre o qual, de novo, pouco se sabe. Ainda assim, está claro que as organizações de notícias usam cada vez mais as análises para avaliar o interesse do público, e essa quantificação de interesse molda as notícias que se tornam visíveis (Cherubini; Nielsen, 2016; Petrie, 2015). Uma maneira de abordar como o público é moldado pela análise é comparar os diferentes sistemas de mídia. Os Estados Unidos e a China são os extremos aqui, uma vez que as notícias são predominantemente suscetíveis às forças do mercado e do Estado, respectivamente (Bolsover, 2017).

Onde as forças do mercado são dominantes, as métricas são usadas para medir o que os leitores e espectadores estão interessados. Onde o Estado controla a mídia, como na China, as métricas são usadas para monitorar a opinião pública – mesmo se a lógica de maximizar a atenção por motivos comerciais também se aplicar na China. Mas as métricas não são apenas do setor privado ou do Estado; elas também são usadas pela mídia de serviço público, como na Suécia. E elas também são usadas pela BBC, uma organização de notícias on-line global, para medir seus públicos em todo o mundo, inclusive na Índia. Em todos os casos, as métricas dependem do que as pessoas clicam e do que é compartilhado e apreciado, além de outros dados. Os usos de métricas e *big data* estão se tornando rotineiros e difusos, mas também são moldados pelo contexto de diferentes sistemas de mídia e pelas diferenças nos usos da mídia. E isso também inclui o quanto os *smartphones* são usados. Por fim, as análises de *big data* também são cada vez mais utilizadas para campanhas políticas (Kreiss, 2016). As campanhas políticas estão sujeitas à regulamentação mais rigorosa do que o marketing comercial, mas também se sobrepõem em parte às organizações que fazem análises no setor privado.

Nesse ponto, será útil fazer uma breve digressão para discutir como a análise de *big data* pode ser usada para a mobilização política, que foi recentemente detalhada por Karpf (2015; 2016). A internet tem sido vista com frequência como uma ferramenta para dar voz aos novos atores políticos, mas Karpf (Ibid.) aponta que ela também pode ser uma ferramenta que permite que as campanhas políticas sejam retroalimentadas – avaliando o que os cidadãos querem e aproveitando o seu engajamento. O argumento pode ser levado um passo adiante: Karpf (Ibid.) examina o que ele chama de “ativismo analítico”, principalmente do ponto de vista de organizações ativistas ou de defesa legal que usam as ferramentas analíticas para atingir os apoiadores e envolvê-los. Isso inclui testes A/B ou a realização de experimentos para descobrir o que funciona de maneira mais eficaz. Isso oferece às organizações ativistas novas capacidades, dando-lhes informações mais refinadas sobre o que os públicos (eleitores, apoiadores, membros) estão interessados.

No entanto, de acordo com o argumento apresentado, e da perspectiva daqueles que estão sendo acionados, esse uso do conhecimento poderia ser igualmente empregado para manipulação e, nesse aspecto, é semelhante ao direcionamento comercial, adaptação e manipulação discutidos aqui para consumidores ou públicos. Nesse caso, também, o *big data* permite aproximar ou aproveitar mais os insumos públicos sobre como eles são usados por atores políticos e regimes políticos. Karpf (Ibid.) reconhece essa semelhança e aponta para um *continuum* em que alguns usos da análise (como em campanhas eleitorais ou marketing, que resultam em práticas discriminatórias de seguro, por exemplo) são mais prejudiciais e, portanto, exigem mais regulação e meios para garantir transparência do que outros (defesa legal/ativismo, publicidade ao consumidor) – com algumas gradações entre eles. Mas Karpf (Ibid.) está pensando principalmente em causas ativistas progressistas – e se as análises forem usadas para promover o racismo ou a intolerância, como acontece com os populistas de direita?

Uma abordagem diferente é reconhecer que a análise de *big data* produz um conhecimento mais eficaz e que esse conhecimento deve ser regulado ou restringido onde quer que tenha efeitos prejudiciais (distorcendo o mercado ou propagandísticos não transparentes). Onde não há nenhum dano ou risco de dano, pode ser deixado livre. Dessa forma, os efeitos da análise poderiam ser levados em conta do ponto de vista dos consumidores e cidadãos. As propostas apresentadas por Karpf (Ibid.) a esse respeito – incluindo a transparência e/ou os meios para garantir que as análises não deturpem as populações – são úteis. No entanto, como Karpf (Ibid.) e outros observaram, a transparência nem sempre é possível (porque os dados são proprietários, ou porque a revelação dos algoritmos

permitiria a *aposta* em, digamos, resultados de mecanismos de busca – embora nesse caso a transparência pudesse estar nas mãos de um terceiro confiável: veja Mayer-Schoenberger; Cukier, 2013). E, às vezes, a transparência é indesejável (como em negociações políticas em que o compromisso é necessário e as cartas precisam ser mantidas próximas ao peito ou, novamente, se a transparência permitiria a *aposta* ou a distorção de insumos). Outro problema identificado por Karpf (Ibid.) é que as análises apontam para as preferências do público apenas quando elas são reveladas nos dados – sem fornecer uma compreensão de quaisquer preocupações ou desejos mais profundos (Karpf [Ibid.] os descreve como “preferências reveladas” *versus* “metapreferências”). Em outras palavras, pode ser problemático se organizações ativistas (ou campanhas) forem lideradas por dados discretos em vez de um quadro mais holístico do que os eleitores ou apoiadores querem¹¹.

¹¹Embora novos usos de dados sejam significativos, é importante, novamente, não superestimar sua função. Como Hersh (2015) documentou, pelo menos para as campanhas eleitorais dos Estados Unidos, a maioria dos dados usados pelos partidos para identificar eleitores ou apoiadores são dados de cadastro e de recenseamento eleitoral publicamente disponíveis, em vez de dados de mídia digital, mesmo que os dois sejam combinados às vezes.

¹²Um efeito da segmentação e customização do *big data* é que esse conhecimento consolida uma cultura de consumismo ao permitir que os mercados maximizem seu alcance e profundidade entre os públicos ou os consumidores.

Para retornar aos principais temas deste trabalho: o *big data* é usado principalmente por empresas de mídia que analisam o comportamento do usuário. Em contraste com a “escuta” de Karpf (Ibid.), usei os termos *customização*, *segmentação* e *manipulação*; todos os três podem ser vistos como formas de *moldagem*. As mensagens de *customização* as personalizam; as de *segmentação* dirigem-nas a um alvo; e *manipular* implica melhorar o envolvimento, dirigindo a atenção – às vezes de maneiras inconscientes aos usuários. Isso permite que a atenção seja canalizada para o conteúdo da publicidade e do marketing – e daí a compra. Portanto, a ideia de que as pessoas exercem maior escolha em ambientes on-line do que nos meios de comunicação de massa é enganosa: em ambas elas são tratadas como públicos-alvo, e mecanismos on-line são mais poderosos a esse respeito¹². Outra diferença aqui é que, enquanto o conteúdo de mídia costumava ser transmitido por meio de dispositivos separados – telefone, televisão, rádio e jornais –, a mídia digital pode ser acompanhada em vários dispositivos e, assim, moldar a atenção com mais intensidade. Os dispositivos podem estar proliferando, mas uma infraestrutura de mídia (parcialmente convergente) molda a atenção das pessoas. As ferramentas analíticas também abarcam o comportamento dos usuários em diferentes mídias; por exemplo, a Google é capaz de rastrear a publicidade em todos os sites em que as pessoas clicam se usarem o Gmail ou o Google+ (Miller, 2014). E a focalização depende, em grande parte, da busca de informações que, pelo menos na escala que assumiu desde o uso em massa da web no final da década de 1990, é uma nova prática social. Essa nova prática social permitiu que as ferramentas de busca e o Google, sobretudo, criassem um novo mercado: um mercado não para informações, mas para o conhecimento das informações que os usuários procuram.

A modelagem tecnológica em relação ao *big data* e à mídia digital significa que o conhecimento se torna mais eficaz ou efetivo. Na ciência social, isso implica que as abordagens científicas ou quantitativas estão se tornando mais proeminentes e frequentemente visam à previsão (uma perspectiva determinista). Na economia, o conhecimento de *big data* ajuda (entre outras coisas) a moldar a atenção, porque pode agregar públicos ou consumidores, mesmo quando estes se desagrupam e são reconfigurados pela movimentação on-line. A vantagem das novas infraestruturas on-line é, portanto, uma concentração de conhecimento que é obtida pela criação de mercados de atenção ou consumo, que são agregados a partir de uma variedade de fontes. Assim, para o público, existe um mercado de nível mais alto que agrega fontes de mídia (como nos sites de comparação de preços). Há dois processos em jogo: um é como os consumidores são analisados cientificamente ou por meios computacionais (tecnológicos), como parte de um contínuo processo de racionalização. A outra é como os mercados estão se tornando mais desencaixados à medida que as transações econômicas (ou os interesses de compra, por exemplo) são agregadas quando se tornam on-line. Os dois processos andam de mãos dadas, na medida em que as partes da economia que já estão mais desencaixadas (serviços, finanças) são mais propensas ao on-line, e as transações mais desencaixadas (que acontecem mais on-line) são mais acessíveis à análise computacional.

O controle de um conhecimento mais eficiente, portanto, leva a uma nova assimetria, na qual as empresas que coletam muitos dados digitais sobre os usuários têm vantagem. Mas a concorrência no mercado depende de os usuários poderem comparar os preços de maneira justa e igualitária. Com preços específicos e adaptados, no entanto, é frequentemente difícil ou impossível fazê-lo, uma vez que os preços são oferecidos de forma não transparente (Turow, 2017). Se os mercados on-line não operarem de forma transparente, eles precisam ser regulamentados para tanto. E embora algumas empresas de mídia digital dominem a atenção do público e, assim, exercitem um controle mais centralizado, a mudança on-line também permite novas formas de produção cultural que se desviem das mídias tradicionais, como as estrelas com muitos seguidores no YouTube.

Essas implicações do conhecimento também são, novamente, moldadas pelos sistemas de mídia. Como em outros aspectos, a China é única em termos do controle que tem sobre a sua mídia. É verdade que as empresas chinesas estão interessadas, principalmente, em aumentar a quantidade de atividade de seus próprios serviços, a fim de maximizar os lucros. Stockmann e Luo (2017) entrevistaram gerentes e desenvolvedores do Sina Weibo; do WeChat, da Tencent, e do Baidu Tieba, as três mídias sociais mais utilizadas na discussão política

¹³Qin, Stroemberg e Wu (2017: 137) argumentam que essa é uma razão pela qual o governo chinês não censura as mídias sociais tão fortemente quanto poderia: enquanto as mídias sociais não representarem uma grande ameaça, remover o seu conteúdo “prejudicaria a capacidade do regime de aprender com informações de base e abordar os problemas sociais antes que se tornem ameaçadores”. Eles baseiam essa descoberta em um estudo de 13,2 bilhões de postagens no Sina Weibo entre 2009 e 2013, do qual eles extraíram posts sobre temas sociais e políticos. Eles também argumentam que a vigilância da mídia social muda o equilíbrio de poder para a liderança do governo central e para os usuários de mídia social em prejuízo dos líderes locais, já que os primeiros compartilham um interesse, por exemplo, em combater a corrupção local ou abusos de poder (2017: 138).

on-line. Essas mídias sociais são influenciadas pelo governo, mas as empresas estão interessadas principalmente em aumentar a atividade política para maximizar o envolvimento dos usuários. O que é diferente aqui é que, como outros meios de comunicação chineses, eles também veem como seu papel *guiar* a opinião pública. A China tem a vantagem de ter um mercado dominado por sua própria mídia digital, de modo que o conhecimento sobre sua população pode ser usado para monitorar e influenciar a opinião pública, também em conluio com o governo¹³. A desvantagem é que esse uso do conhecimento de *big data* não vai além de suas fronteiras. A Google, o Twitter e o Facebook, ao contrário, têm mercados globais e a capacidade de atingir públicos globais, com a desvantagem (para os formuladores de políticas) de influenciar apenas públicos e consumidores, em vez de cidadãos ou comunidades.

Isso nos leva de volta à dicotomia feita antes: Orwell pensava que as pessoas são cada vez mais enganadas pela propaganda e se submetem a ela. Huxley, por outro lado, defendia a visão de que estamos cada vez mais sendo distraídos pelos prazeres e nos acostumando com isso. O exemplo do Facebook, entre outros, sugere que Huxley tinha razão. Como Thomson (2016: 264-265) coloca:

Todos nós estamos sendo perfilados o tempo todo como parte do sistema tecnocomercial pelo qual nossas estruturas corporativas (privadas e governamentais) tentam manter-se encarregadas de nós e nos manter ajustados como elementos previsíveis no mercado. [...] Isso pode parecer alarmante, mas seguimos o sistema com alegria suficiente. [...] Entendemos que nossas informações e nossos caprichos na internet estão sendo usados para nos posicionar e nos servir “o que queremos”. E não deveríamos nos contentar se tivermos o que queremos – mesmo que esses desejos sejam cada vez mais codificados?

Essas implicações se aplicam globalmente, mas é um pouco diferente com o *big data* para elaboração de políticas.

BIG DATA E POLÍTICAS EM DIFERENTES SISTEMAS DE MÍDIA

A formulação de políticas relacionadas à internet e à web deve ser vista nos contextos nacionais dos sistemas de mídia, de modo que podemos examinar o *big data* brevemente em relação a isso. Na China, o governo está desenvolvendo ativamente uma estratégia de *big data*, agora voltada mais para a *gestão social* do que para o *governo eletrônico*. Zeng (2016) argumenta que há um perigo de que esses esforços de *big data* possam sair pela culatra, permitindo que as facções de elite ataquem umas às outras, usando dados para acusar os inimigos

de corrupção e coisas similares (o faccionalismo da elite tem sido visto como uma causa-chave do colapso do regime autoritário). Mas é mais provável que os principais efeitos do *big data* sejam triplos: primeiro, permitir que o governo efetue *testes beta* das políticas em certas áreas para ver se elas devem ser adotadas de forma mais ampla, realizando efetivamente experimentos de política social; o segundo é através de arquivos individuais digitais ou registros (*dang-an*), que são mantidos dos cidadãos, monitorando seu bom e mau comportamento, e desenvolvendo um sistema de *crédito social* que os recompense a partir disso; o terceiro é avaliar a opinião pública nas mídias sociais para poder responder a ela. As condições para todos os três efeitos são obviamente únicas na China, particularmente uma vez que as mídias sociais chinesas trabalham com o governo, e as mídias digitais estão dentro de limites geográficos e politicamente aceitáveis.

O sistema de identificação pessoal da Índia, o *Aadhaar*, por outro lado, ainda está em seus estágios iniciais, embora seja o maior sistema biométrico do tipo no mundo¹⁴. No entanto, em um país onde a infraestrutura é fraca, um único sistema que conecta os registros individuais e vincula uma grande parte da população ao Estado, aos bancos e aos meios de pagamentos pela primeira vez tem um grande impacto potencial. Dito isso, a viva sociedade civil da Índia manifestou-se contra o sistema *Aadhaar*, apontando possíveis preocupações com a privacidade e abusos comerciais. Os proponentes, por outro lado, disseram que o sistema permitirá a eliminação da corrupção – por exemplo, na entrega de benefícios sociais que, de outra forma, são muitas vezes *desviados* pelos intermediários. Essas afirmações devem ser colocadas em contexto, sendo os *smartphones*, de longe, a forma mais comum pela qual as pessoas usam a mídia digital, principalmente para entretenimento e socialização, como em outros lugares. Doron e Jeffrey (2013: 224) dizem que “o telefone celular atraiu as pessoas da Índia para as relações com o estado capitalista de manutenção de registros de forma mais abrangente do que qualquer mecanismo ou tecnologia anterior”. E o pagamento on-line por meio de *smartphones*, juntamente com os esforços do governo para introduzir uma maneira mais racional e eficaz de coletar o imposto sobre vendas, são, portanto, os efeitos previsíveis mais prováveis do *big data*.

Nos Estados Unidos, o governo Obama publicou relatórios destinados a promover o uso de *big data* e também a proteção contra riscos¹⁵. Os Estados Unidos também têm uma exigência legal segundo a qual os dados financiados pelo contribuinte devem ser abertamente utilizáveis. Ao mesmo tempo, há uma forte tradição de desconfiança do governo nos Estados Unidos, exacerbada pelos recentes escândalos em torno da vigilância governamental. Talvez mais importante, as empresas estadunidenses de mídia digital, apesar de protegerem

¹⁴Rao e Greenleaf (2013) concluem que o *Aadhaar* está em um estado de limbo legal e que sua implementação é irregular e incerta. Mas esse mesmo estado de incerteza, argumentam, também significa que suas utilizações e potenciais usos podem se expandir sem controle. Embora seja útil ampliar as implicações dessa forma de *big data* para o caso indiano, simplesmente não há pesquisa suficiente (tanto quanto sei) sobre os resultados iniciais desse sistema.

¹⁵White House (2014) e Smith, Patil e Muñoz (2016).

os dados de seus usuários diante do Estado, utilizam os dados do usuário para atingir os consumidores, o que (como mencionado) os consumidores não gostam, mas sentem-se impotentes para reagir (Turow, 2017). No entanto, essas empresas também dominam a mídia on-line em todo o mundo, em meio a sinais mistos de rejeição contra os usos dos dados de seus usuários. A pesquisa dos Estados Unidos também está na vanguarda da fronteira do conhecimento em muitas áreas de *big data*, mas a aplicação desse saber para a governança será limitada pelas condições políticas e econômicas prevalentes nos Estados Unidos. Atualmente, as condições políticas são tais que, além das preocupações de segurança, há pouco espaço para usos mais extensivos de *big data* pelo governo para a formulação de políticas. As empresas dos Estados Unidos, por outro lado, continuarão na vanguarda do uso de *big data* na segmentação de consumidores e públicos-alvo.

Finalmente, a Suécia tem uma tradição fortemente desenvolvida de manutenção de registros do governo, mas também de fortes leis de proteção de dados. A Suécia, além disso, tem amplo uso de um sistema de identificadores pessoais. Esse identificador – o *número pessoal* – é obrigatório para todos os que moram na Suécia e está em uso contínuo desde antes do início dos registros informatizados. Eles são rotineiramente usados para tudo, desde o sistema de saúde e estatísticas nacionais até impostos e compras, unindo muitas facetas do comportamento individual e atingindo toda a população. As leis suecas de proteção de dados foram testadas em uma série de casos amplamente divulgados em que os dados foram usados indevidamente pelos pesquisadores, mas ainda há um alto nível de confiança pública nos investigadores e no governo em relação aos usos de dados (Axelsson; Schroeder, 2009). A Suécia pode, portanto, ser vista como um exemplo de como um Estado forte tem pré-condições favoráveis para capitalizar a partir do *big data*.

Assim, podemos ver que os sistemas de mídia, mas também os Estados, moldam o papel do *big data* na elaboração de políticas. As empresas de mídia digital atravessam os contextos nacionais, exceto no caso da China, onde predominam os meios digitais comerciais com fronteiras nacionais (Baidu, Alibaba, Tencent). Além da proteção nacional de dados, o Regulamento Europeu de Proteção de Dados deve, pelo menos em parte, conter o uso de *big data* por parte das empresas dos Estados Unidos. Desse modo, os usos comerciais e de formulação de políticas do *big data* só irão se sobrepor parcialmente: os formuladores de políticas querem usar os dados de mídia digital, já que é a fonte mais importante de *big data*, além das estatísticas e pesquisas coletadas nacionalmente. Mas as empresas de mídia e os pesquisadores fazem mais – embora, como vimos, dentro dos limites das fontes e usos de dados. ■

REFERÊNCIAS

- AXELSSON, A.-S.; SCHROEDER, R. Making it open and keeping it safe: e-enabled data sharing in Sweden. *Acta Sociologica*, Thousand Oaks, v. 52, n. 3, p. 213-226, 2009. DOI: <https://doi.org/10.1177/0001699309339799>
- BAKSHY, E.; MESSING, S.; ADAMIC, L. A. Exposure to ideologically diverse news and opinion on Facebook. *Science*, Washington, DC, v. 348, n. 6239, p. 1130-1132, 2015. DOI: <https://doi.org/10.1126/science.aaa1160>
- BENIGER, J. *The control revolution: technological and economic origins of the information society*. Cambridge, MA: Harvard University Press, 1986.
- BOLSOVER, G. Technology and political speech: commercialisation, authoritarianism and the supposed death of the internet's democratic potential. 2017. 386 f. Tese (Doutorado em Informação, Comunicação e Ciências Sociais) – Oxford Internet Institute, University of Oxford, Oxford, 2017. Disponível em: <<https://goo.gl/UtYq7j>>. Acesso em: 28 jul. 2018.
- BOND, R. et al. A 61-million-person experiment in social influence and political mobilization. *Nature*, Londres, v. 489, p. 295-298, 2012. DOI: <https://doi.org/10.1038/nature11421>
- BORGMAN, C. *Big data, little data, no data*. Cambridge, MA: MIT Press, 2015.
- BRIGHT, J. “Big social science”: doing big data in the social sciences. In: LEE, R.; FIELDING, N.; BLANK, G. (Eds.). *Handbook of online research methods*. Londres: Sage, 2016. p. 125-139.
- CASTELLS, M. *Communication power*. Oxford: Oxford University Press, 2009.
- CHERUBINI, F.; NIELSEN, R. Digital news reports 2016. Editorial analytics: how news media are developing and using audience data and metrics. Oxford: Reuters Institute, 2016. Disponível em: <<https://goo.gl/coKPFw>>. Acesso em: 28 jul. 2018.
- CONOVER, M. et al. Political polarization on Twitter. In: INTERNATIONAL AAAI CONFERENCE ON WEBLOGS AND SOCIAL MEDIA, 5., 2011, Barcelona. *Proceedings...* Palo Alto: AAAI, 2011. v. 133. p. 89-96. Disponível em: <<https://goo.gl/zX3B8r>>. Acesso em: 28 jul. 2018.
- DONNER, J. *After access: inclusion, development, and a more mobile internet*. Cambridge, MA: MIT Press, 2015.
- DORON, A.; JEFFREY, R. *The great Indian phone book: how the mass mobile changes business, politics and daily life*. Cambridge, MA: Harvard University Press, 2013.
- EAGLE, N.; GREENE, K. *Reality mining: using big data to engineer a better world*. Cambridge, MA: MIT Press, 2014.
- EKBIA, H. et al. Big data, bigger dilemmas: a critical review. *Journal of the Association for Information Science and Technology*, Hoboken, v. 66, n. 8, p. 1523-1545, 2015. DOI: <https://doi.org/10.1002/asi.23294>

- ESLAMI, M. et al. "I always assumed that I wasn't really that close to [her]:" reasoning about invisible algorithms in the news feed. In: CONFERENCE ON HUMAN FACTORS IN COMPUTING SYSTEMS, 33., 2015, Seoul. *Proceedings...* Nova Iorque: ACM, 2015. p. 153-162. DOI: <https://doi.org/10.1145/2702123.2702556>
- _____. First I "like" it, then I hide it: folk theories of social feeds. In: CONFERENCE ON HUMAN FACTORS IN COMPUTING SYSTEMS, 2016, San Jose, CA. *Proceedings...* Nova Iorque: ACM, 2016. p. 2371-2382. DOI: <https://doi.org/10.1145/2858036.2858494>
- FISCHER, C. *America calling: a social history of the telephone to 1940*. Berkeley: University of California Press, 1992.
- GRIMMELMAN, J. The Facebook emotional manipulation study: sources. *The Laboratorium*, Nova Iorque, 30 jun. 2014. Disponível em: <<https://goo.gl/uyXuz6>>. Acesso em: 28 jul. 2018.
- HACKING, I. *Representing and intervening*. Cambridge, UK: Cambridge University Press, 1983.
- _____. The self-vindication of the laboratory sciences. In: PICKERING, A. (Ed.). *Science as practice and culture*. Chicago: University of Chicago Press, 1992. p. 29-64.
- HAENSSGEN, M. Mobile phone diffusion and rural healthcare access in India and China. 2016. 391 f. Tese (Doutorado em Desenvolvimento Internacional) – Hertford College, University of Oxford, Oxford, 2016. Disponível em: <<https://goo.gl/x3X4P2>>. Acesso em: 28 jul. 2018.
- HAMPTON, K. et al. Social networking sites and our lives. *Pew Research Center*, Washington, DC, 16 jun. 2011. Disponível em: <<https://goo.gl/YZs3fr>>. Acesso em: 28 jul. 2018.
- HERSH, E. *Hacking the electorate: how campaigns perceive voters*. Cambridge, UK: Cambridge University Press, 2015.
- KARPF, D. *Analytic activism: digital listening and the new political strategy*. Oxford: Oxford University Press, 2015.
- _____. Schrodinger's audience: how news analytics handed America Trump. *Civicist*, Nova Iorque, 4 maio 2016. Disponível em: <<https://goo.gl/3jpFrr>>. Acesso em: 28 jul. 2018.
- KRAMER, A.; GUILLORY, J.; HANCOCK, J. Experimental evidence of massive-scale emotional contagion through social networks. *Proceedings of the National Academy of Sciences*, Washington, DC, v. 111, n. 24, p. 8788-8790, 2014. DOI: <https://doi.org/10.1073/pnas.1320040111>
- KREISS, D. *Prototype politics: technology-intensive campaigning and the data of democracy*. Oxford: Oxford University Press, 2016.

- LEWIS, K. et al. Tastes, ties, and time: a new social network dataset using Facebook.com. *Social Networks*, Amsterdã, v. 30, n. 4, p. 330-342, 2008. DOI: <https://doi.org/10.1016/j.socnet.2008.07.002>
- LICOPPE, C. “Connected” presence: the emergence of a new repertoire for managing social relationships in a changing communication technoscape. *Environment and Planning D: Society and Space*, Thousand Oaks, v. 22, n. 1, p. 135-156, 2004. DOI: <https://doi.org/10.1068/d323t>
- LING, R. et al. Small circles: mobile telephony and the cultivation of the private sphere. *The Information Society*, Abingdon, v. 30, n. 4, p. 282-291, 2014. DOI: <https://doi.org/10.1080/01972243.2014.915279>
- MAYER-SCHOENBERGER, V.; CUKIER, K. *Big data*: a revolution that will transform how we live, work and think. Londres: John Murray, 2013.
- MCCHESENEY, R. W. *Digital disconnect*: how capitalism is turning the internet against democracy. Nova Iorque: The New Press, 2013.
- MEISSNER, M. China’s social credit system: a big-data enabled approach to market regulation with broad implications for doing business in China. *China Monitor*, Berlim, 24 maio 2017. Disponível em: <<https://goo.gl/TkgvxH>>. Acesso em: 28 jul. 2018.
- MEYER, E. T.; SCHROEDER, R. *Knowledge machines*: digital transformations of the sciences and humanities. Cambridge, MA: MIT Press, 2015.
- MEYER, M. Misjudgements will drive social trials underground. *Nature*, Londres, v. 511, n. 265, 17 jul. 2014. DOI: <https://doi.org/10.1038/511265a>
- MILLER, C. C. The Plus in Google Plus? It’s Mostly for Google. *The New York Times*, Nova Iorque, 15 Feb. 2014, p. A1.
- NAPOLI, P.; OBAR, J. The emerging mobile internet underclass: a critique of mobile internet access. *The Information Society: an international journal*, Abingdon, v. 30, n. 5, p. 323-334, 2015. DOI: <https://doi.org/10.1080/01972243.2014.944726>
- O’NEIL, C. *Weapons of math destruction*: how big data increases inequality and threatens democracy. Londres: Allen Lane, 2016.
- PAN, J. How market dynamics of domestic and foreign social media firms shape strategies of internet censorship. *Problems of Post-Communism*, Abingdon, v. 64, n. 3-4, 2017. DOI: <https://doi.org/10.1080/10758216.2016.1181525>
- PARISER, E. *The filter bubble*: what the internet is hiding from you. Harmondsworth: Penguin, 2011.
- PASQUALE, F. *The black box society*: the secret algorithms that control money and information. Cambridge, MA: Harvard University Press, 2015.
- PETRIE, C. The traffic factories: metrics at Chartbeat, Gawker Media, and The New York Times. *Tow Center for Digital Journalism*, Nova Iorque, 7 maio 2015. Disponível em: <<https://goo.gl/Qf1HRR>>. Acesso em: 28 jul. 2018.

- PORTER, T. Statistics and statistical methods. In: PORTER, T.; ROSS, D. (Eds.). *The modern social sciences*. Cambridge, UK: Cambridge University Press, 2008. p. 238-250.
- QIN, B.; STROEMBERG, D.; WU, Y. Why does China allow freer social media? Protests versus surveillance versus propaganda. *Journal of Economic Perspectives*, Pittsburgh, v. 31, n. 1, p. 117-140, 2017. DOI: <https://doi.org/10.1257/jep.31.1.117>
- RAINIE, L. et al. 2012. Social media and political engagement. *Pew Research Center*, Washington, DC, 19 out. 2012. Disponível em: <<https://goo.gl/d5DDDR>>. Acesso em: 28 jul. 2018.
- RAO, U.; GREENLEAF, G. Subverting ID from above and below: the uncertain shaping of India's new instrument of e-governance. *Surveillance and Society*, Chapel Hill, v. 11, n. 3, p. 287-300, 2013. DOI: <https://doi.org/10.24908/ss.v11i3.4496>
- RULE, J. *Theory and progress in social science*. Cambridge, UK: Cambridge University Press, 1997.
- _____. *Privacy in peril: how we are sacrificing a fundamental right in exchange for security and convenience*. Nova Iorque: Oxford University Press, 2007.
- SCHROEDER, R. *Rethinking science, technology and social change*. Stanford: Stanford University Press, 2007.
- _____. Big data and the Brave New World of social media research. *Big Data and Society*, Thousand Oaks, p. 1-11, Jul.-Dec. 2014a. DOI: <https://doi.org/10.1177/2053951714563194>
- _____. Big data: towards a more scientific social science and humanities? In: GRAHAM, M.; DUTTON, W. H. (Eds.). *Society and the internet*. Oxford: Oxford University Press, 2014b. p. 164-176.
- _____. Big data and communication research. *Oxford Research Encyclopedia of Communication*, Oxford, Nov. 2016. DOI: <https://doi.org/10.1093/acrefore/9780190228613.013.276>
- _____. *Social theory after the internet: media, technology and globalization*. Londres: UCL Press, 2018.
- SETTLE, J. et al. Quantifying political discussion from the universe of Facebook status updates. *SSRN Electronic Library*, Amsterdã, 2013. DOI: <https://dx.doi.org/10.2139/ssrn.2307685>
- SMITH, M.; PATIL, D.; MUÑOZ, C. Big Risks, big opportunities: the intersection of big data and civil rights. White House, Washington, DC, 4 maio 2016. Disponível em: <<https://goo.gl/QnwVKd>>. Acesso em: 7 ago. 2018.
- STOCKMANN, D. *Media commercialization and authoritarian rule in China*. Cambridge, UK: Cambridge University Press, 2013.

- STOCKMANN, D.; LUO, T. Which social media facilitate online public opinion in China? *Problems of Post-Communism*, Abingdon, v. 64, n. 3-4, p. 189-202, 2017. DOI: <https://doi.org/10.1080/10758216.2017.1289818>
- THELWALL, M.; BUCKLEY, K.; PALTOGLOU, G. Sentiment strength detection for the social web. *Journal of the American Society for Information Science and Technology*, Hoboken, v. 63, n. 1, p. 163-173, 2012. DOI: <https://doi.org/10.1002/asi.21662>
- THOMSON, D. *Television: a biography*. Nova Iorque: Thames & Hudson, 2016.
- TUROW, J. *The aisles have eyes: how retailers track your shopping, strip your privacy, and define your power*. New Haven: Yale University Press, 2017.
- WATTS, D. J. Should social science be more solution-oriented? *Nature Human Behaviour*, Londres, v. 1, 2017. Artigo 0015. DOI: <https://doi.org/10.1038/s41562-016-0015>
- WHITE HOUSE. *The big data and privacy review*. Washington, DC: White House, 2014. Disponível em: <<https://goo.gl/S9BPWA>>. Acesso em: 28 jul. 2018.
- WHITLEY, R. *The intellectual and social organization of the sciences*. 2. ed. Oxford: Oxford University Press, 2000.
- WILSON, R.; GOSLING, S.; GRAHAM, L. A review of Facebook research in the social sciences. *Perspectives on Psychological Science*, Thousand Oaks, v. 7, n. 3, p. 203-220, 2012. DOI: <https://doi.org/10.1177/1745691612442904>
- ZENG, J. China's date with big data: will it strengthen or threaten authoritarian rule? *International Affairs*, Oxford, v. 92, n. 6, p. 1443-1462, 2016. DOI: <https://doi.org/10.1111/1468-2346.12750>

Artigo recebido em 9 de julho de 2018 e aprovado em 25 de julho de 2018.