

## **ÉPISTEMOLOGIA, MÉTODOS E TEORIAS DA COMUNICAÇÃO NA ERA DO *BIG DATA*: PANORAMA CRÍTICO DA PESQUISA EM MÍDIAS SOCIAIS**

**Ana Thereza Nogueira Soares**

---

### **RESUMO**

O artigo propõe uma reflexão crítica sobre as implicações epistemológicas, metodológicas e teórico-conceituais das pesquisas baseadas no *Big Data* – especialmente em dados de mídias sociais – para o campo científico da comunicação. Do ponto de vista epistemológico, revela a insustentabilidade de modelos de análise baseados em enquadramentos estáticos da comunicação, alegando que os processos sociais que emergem com a influência da internet apresentam-se, inequivocamente, em formatos fluidos e contingentes. Nesse âmbito, destaca que a própria evolução da tecnologia tem condições de impulsionar a construção de ferramentas de coleta e análise de dados capazes de apreender os movimentos comunicacionais, fazendo jus à necessidade de alinhamento entre ontologia, epistemologia e metodologia na pesquisa científica. O texto, ainda, problematiza questões relativas à construção teórico-conceitual da comunicação. Acredita-se que o relevo adquirido pelos dados, nos últimos anos, não deve apontar para um domínio do empírico sobre o teórico. Efetivamente, o fortalecimento da ciência da comunicação passa pela precisão e pelo cuidado com o uso de termos, modelos e referências teóricas historicamente consolidados na problematização e explicação do contemporâneo.

### **PALAVRAS-CHAVE**

*Big Data*; epistemologia; mídias sociais; teoria da comunicação

---

### **ABSTRACT**

This paper proposes a critical reflection on the epistemological, methodological and theoretical implications of the researches based on Big Data – especially on social media data – for the scientific field of communication. From an epistemological point of view, it reveals the unsustainability of analytical models based on static frameworks of communication, claiming that the social processes that emerge with the influence of the internet are unequivocally presented in fluid and contingent formats. In this context, it highlights that the evolution of technology itself has the potential to boost the construction of data collection and analysis tools capable of grasping the communication movements, justifying the need for alignment between ontology, epistemology and methodology in scientific research. The text, also, poses questions about communication theory and its concepts. It is believed that the relevance acquired by data in recent years should not point to a domain of the empirical over the theoretical. Effectively, the strengthening of the communication science demands precision and care with the use of terms, models and theoretical references historically consolidated in the problematization and explanation of the contemporary.

### **KEYWORDS**

Big Data; communication theory; epistemology; social media

---

## INTRODUÇÃO

Pode o *Big Data*, a partir do potencial disruptivo que apresenta para as formas de organização da sociedade e para a ciência, gerar insumos para rediscussão de conceitos e teorias da comunicação? A questão, aparentemente retórica, configura-se como essencial para uma reflexão sobre o futuro da ciência da comunicação. Cerca de um século após as primeiras elaborações teóricas sobre o fenômeno comunicacional e seus efeitos sobre as configurações psíquicas e sociais dos indivíduos na sociedade de massa, a ubiquidade das mensagens e dos meios materializada pela apropriação que fazemos da tecnologia gera intenso debate, que alcança todas as esferas da sociedade.

Cukier e Myer-Schonberger (2013), Boyd e Crawford (2012) Coté, Gerbaudo e Pybus (2016) e Gil de Zúñiga (2015) são apenas alguns dos autores dedicados à reflexão sobre a força do *Big Data* e dos algoritmos na conformação e reprodução das práticas socioculturais. Segundo Kitchin (2017), há uma pressão cada vez maior para que se analise criticamente o modo como os algoritmos têm interferido na produção do conhecimento científico. Torna-se difícil, de fato, identificar um campo que não seja direta ou indiretamente afetado pela velocidade, variedade e volume da produção dos dados e de sua conversão em informações com alto poder de difusão. Da ciência à saúde pública, do sistema bancário à produção de notícias: a quantidade de dados produzidos e registrados espontânea ou compulsoriamente pela sociedade cresce de forma inimaginada, gerando novos regimes políticos que têm potencial para desafiar convenções sociais em vigor. Algo realmente distinto está impresso na realidade social quando, por exemplo, uma empresa como a Google torna-se base confiável de conhecimento para a prevenção e controle de uma gripe com potencial para se alastrar pelo mundo (Cukier & Mayer-Schönberger, 2013; Helbing, 2015).

Assim, o objetivo deste texto é apontar, de forma preliminar, pontos críticos e caminhos para o ofício do pesquisador da comunicação na atualidade, tendo em vista as vastas possibilidades de coleta e análise de dados habilitadas pelas plataformas, ferramentas e aplicativos de internet. O foco da reflexão é o impacto do *Big Data* nas epistemologias, teorias e métodos em que se baseiam as pesquisas que traduzem a realidade da comunicação, por meio de dados oriundos das diversas mídias e aplicativos digitais hoje disponíveis para a sociedade.

### **BIG DATA: CIRCUNSCREVENDO OS SIGNIFICADOS DO FENÔMENO PARA A CIÊNCIA**

O crescimento exponencial de pesquisas e artigos que envolvem a mensuração da vida social através de ambientes digitais como plataformas de mídias sociais e sistemas de dados abertos governamentais revela a premente necessidade de debate sobre questões epistemológicas e teórico-metodológicas que os dados vêm ajudando a forjar.

Coté, Gerbaudo e Pybus (2016), por exemplo, situam o *Big Data* no espectro político que orienta a manipulação de dados algorítmicos pelos agentes sociais. A partir da questão geral sobre que relações de poder influem no conhecimento gerado pela análise de dados digitais, os autores refletem criticamente sobre os significados da apropriação

do *Big Data* pelos cientistas. Para eles, não convencem os argumentos utilizados em favor de um conhecimento isento e de validade universal garantido pelas supostas neutralidade e objetividade técnica das coletas massivas de dados da internet. O poder de mercado do Silicon Valley e seu papel preponderante nessa nova cadeia produtiva da economia, as desigualdades estruturais entre a geração dos dados pelos cidadãos e seu controle pelas corporações, bem como a opacidade dos algoritmos que decidem configurações da rede e percepções dos usuários (Coté, Gerbaudo & Pybus, 2016) são apenas alguns dos sinais de que já exploramos uma nova interface da relação da sociedade com a tecnologia.

Há no espírito de nosso tempo uma urgência em entender como a díade dos algoritmos e *Big Data* podem habilitar novas formas sociais e culturais. É preciso entender como eles são empregados e desenvolver um senso crítico sobre seus limites, capacidades, implicações e possibilidades. Se esses arranjos tecnológicos propiciam um vasto manancial de possibilidades de conhecimento, contudo precisam também ser compreendidos diante do aspecto humano que orienta os procedimentos formais de reconhecimento de padrões, estabelecimento de parâmetros de programação e geração de dados em que eventualmente incorrem lacunas semânticas e preconceitos (Kitchin, 2017; Uricchio, 2017).

Ao mesmo tempo, os mecanismos de dominação exercidos pela dinâmica capitalista de monetização dos dados digitais têm sido constantemente tensionados pelos próprios agentes sociais. Isto é, os usos e apropriações dos dados pelas esferas econômica e política, porque vêm adquirindo essa enorme dimensão e influência, tornam-se também *inputs* para a constituição do conhecimento, no inevitável movimento reflexivo que rege as consciências individuais e coletivas. Milan e van der Velden (2016) propõem o conceito de “data ativismo”, vislumbre de um conhecimento que escape à reificação do futuro, a partir de um desenvolvimento crítico da ciência e da tecnologia, inscrito pelo próprio uso dos dados pelos investigadores. Neste âmbito, a crescente disponibilidade dos dados é vista como uma oportunidade poderosa e sem precedentes para provocar mudanças sociais. O “data ativismo” apoia a emergência de novas culturas epistêmicas que “desafiam as leituras predominantes da realidade” e “moldam a forma como relacionamo-nos com o conhecimento e sua validação” (Milan & van der Velden, 2016, p. 63).

Do ponto de vista da investigação sobre a comunicação política nas mídias sociais, ou das plataformas digitais de redes sociais, também Gil de Zúñiga (2015) introduz um conjunto de trabalhos que oferecem formas únicas de combinar métodos e técnicas relacionados ao *Big Data* e que, além disso, levam em conta o componente ético da utilização de dados pessoais para a formulação de pesquisas. O próprio autor reconhece, porém, que saber como interpretar a imensidão de dados disponibilizados por nossas “pegadas digitais” permanece ainda um enorme desafio (Gil de Zúñiga, 2015, p. 2). No contexto que vivenciamos, estão em jogo noções como a validade de construtos científicos, testes de hipóteses e generalização, ou seja, alguns dos cânones paradigmáticos da ciência moderna. Preocupam os autores questões como a eliminação das teorias como ponto de partida para a produção de conhecimento, a perda de força explicativa

da dimensão estrutural da sociedade (Coté Gerbaudo & Pybus, 2016), assim como os moldes das ferramentas de coleta de dados capazes de reunir potencialmente infinitas bases de dados digitais (Gil de Zúñiga, 2015).

A inversão de procedimentos metodológicos que colocam as coleções e correlações de dados extensivos sobre o social como expressões máximas e aprimoradas do conhecimento é tida como indicador de uma macro revolução científica, no sentido que Kuhn (1998) imprimiu ao termo. Kitchin (2014) explora justamente em que medida o *Big Data* conduz a epistemologias alternativas nas ciências sociais, a partir de uma avaliação também crítica das mudanças nas práticas de pesquisa. O autor desafia a ideia, que emerge nesse contexto, de um *modus operandi* puramente indutivo nas pesquisas, defendendo que, ao invés do fomento ao empiricismo em detrimento da construção teórica, a ciência conduzida pelos dados (*data-driven science*, no original) seja uma versão reconfigurada do método científico tradicional. Para isso, entretanto, é preciso que dados sejam coletados e interpretados à luz de questões relevantes para o conhecimento já existente. Os conjuntos desagregados de “fotografias” relativas a ambientes e contextos temporais específicos da realidade digital, ao ilustrar algum tipo de correlação visualizada pelo investigador sobre rastros sociais dispostos na internet, não respondem, necessariamente, a questões teórico-científicas.

O aspecto sedutor da pesquisa baseada em *Big Data* afirma-se, também, pela construção de uma mitologia, segundo o termo usado por Boyd e Crawford, que se baseia na “crença generalizada de que grandes bases de dados oferecem uma forma maior de inteligência e conhecimento geradora de conhecimentos anteriormente improváveis, com aura de verdade, objetividade e acurácia” (2012, p. 663). Tal dimensão simbólica dos dados atua como reforço para a ilusão de abrangência, representatividade e fidedignidade das informações e postulados disponibilizados tende a ser capturada, assim, como fundamento ontológico que orienta e molda o conhecimento científico e sua difusão pública. Há uma garantia de verossimilhança que subjaz a essa nova epistemologia de viés altamente empiricista e alicerçada por uma profusão de mergulhos exploratórios em conjuntos de dados minerados da internet. Nesse movimento, porém, Kitchin (2014, p. 5) adverte que “os dados são examinados através de lentes particulares que influenciam como são interpretados” e, ademais, que “correlações entre variáveis num conjunto de dados podem ser aleatórias por natureza e possuir nenhuma ou pequena associação causal”.

A naturalização de falácias ecológicas apontada por Kitchin (2014), como risco potencial dessa epistemologia da ciência que emerge com o *Big Data*, não se configura como questão isolada para a reflexão. Entre outros desafios observados para as práticas tradicionais do campo científico está também aquele de uma excessiva individualização dos métodos de busca e coleta de dados. Isso conduz, no extremo, à incomensurabilidade das análises e, portanto, à inviabilização dos procedimentos críticos que constituem a base sobre a qual são testadas e aperfeiçoadas as teorias.

A autonomia concedida aos pesquisadores para a elaboração de sua própria ferramenta de rastreamento, mineração, coleta e análise de dados tem potencial para

tornar-se, paradoxalmente, aquilo que retira da ciência moderna uma de suas principais forças – a possibilidade de evolução e busca da verdade por meio da comparação de resultados de testes investigativos e da refutação de hipóteses. Porque enfatiza o empírico disposto nas plataformas digitais e o estabelece, *a priori*, como a verdade a ser desvelada, a customização das pesquisas fundamentadas pelo *Big Data* individualiza o fazer científico, transformando enquadramentos múltiplos de realidade em conhecimento válido. Os resultados de pesquisas, particularizados por instrumentos algorítmicos que coletam e sistematizam dados, tendem a perder o potencial explicativo que é marca da generalização e alimenta as teorias.

Hoje, diversos laboratórios e grupos de pesquisa espalhados pelo mundo, dedicados à pesquisa sobre internet e interessados em explorar dados armazenados digitalmente, têm seus próprios modelos e ferramentas digitais de coletas de dados<sup>1</sup>. Os algoritmos criados para a seleção e combinação de dados das mídias sociais são inúmeros, e nem todos têm o código aberto. O interesse crescente na realização de pesquisas sobre as redes de relacionamento social na internet (que ganham tanto relevo nos estudos do campo da comunicação na última década) está refletido na vasta gama de ferramentas e *softwares* disponíveis exclusivamente para o tratamento desse tipo de dados, cada qual com sua funcionalidade e com seu percurso de codificação algorítmica e formulação estética<sup>2</sup>.

Apesar dessa diversidade de abordagens representar uma superação em relação a modelos e técnicas que limitam a compreensão da realidade social, justamente porque construídos a partir de uma perspectiva de controle e de fixidez dessa mesma realidade em enquadramentos interpretativos estáticos (e incongruentes, portanto, com o movimento inerente à comunicação), deve-se atentar para algumas características que singularizam essas ferramentas de pesquisa. A construção desses dispositivos técnicos de captura e análise de dados é feita por meio de algoritmos, cuja forma heterogênea e opaca, eventualmente programada para exercer a função de inteligência artificial, tem potencial para estender as intenções balizadoras de sua concepção. Como salienta Kitchin, os algoritmos trabalham de modos multivariados, “precisam ser reconhecidos como sendo ontogenéticos, performativos e contingentes: isto é, eles não são nunca fixos por natureza, mas sim emergentes” (2017, p. 21). “Muitos algoritmos são desenhados para serem reativos e mutáveis por *inputs*” (Kitchin, 2017, p. 21), desdobrando continuamente, por meio de sua inteligência artificial, novos dados e informações passíveis de interpretação.

<sup>1</sup> Entre alguns dos muitos centros de pesquisa hoje em funcionamento podem ser citados o The Citizen Lab, laboratório da Munk School of Global Affairs, na Universidade de Toronto; o The Social Media Lab, da Ryerson University, também em Toronto, o Centro de Cultura Digital, no México; o Nordic Centre for Internet and Society, da Universidade de Siegen, na Alemanha; o Singapore Internet Research Centre e o Media Lab Research, do MIT (Massachusetts Institute of Technology), nos Estados Unidos. No Brasil, contando apenas o foco de pesquisa em dados de mídias sociais, podemos citar ao menos três centros de pesquisa, o Laboratório de Estudos sobre Imagem e Cibercultura (Labic), da Universidade Federal do Espírito Santo; o Grupo Interdisciplinar de Pesquisa em Análise de Redes Sociais, da Universidade Federal de Minas Gerais e o Laboratório de Mídia, Discurso e Análise de Redes Sociais (MIDIARS), da Universidade Católica de Pelotas. Essa lista não é exaustiva, contudo.

<sup>2</sup> Apenas a título de ilustração, as seguintes ferramentas, entre muitas outras, compõem a lista de alternativas para coleta, tratamento e análise de dados da internet: NodeXL, iGraph, Netlytic, Uberlink, Issue Crawler, Gephi, Cosmos, Chorus, VT Tracker e Netvizz, este último específico para coleta de dados do Facebook. Os recursos e aplicações permitidas são inúmeras, gerando as mais diversas especificidades analíticas.

Pouco conhecemos, todavia, os efeitos oriundos da manipulação do *Big Data* para a ciência, e ainda menos para nossas múltiplas formas de comunicação contemporâneas e para os metadados minerados por cientistas da comunicação a partir delas. Como a codificação algorítmica que impulsiona o *Big Data* tende a ficar obscurecida pela estética dos aplicativos, *softwares* e demais aparatos técnicos, é trabalhoso desvelar os sentidos originários que levam à difusão e ao compartilhamento de dados pelos agentes. Tais sentidos dependem de traduções inversas de códigos previamente programados, afinal. Kitchin (2017) também se refere à questão das traduções em sua revisão crítica acerca dos algoritmos. A programação dos espaços cibernéticos digitais depende de duas traduções, fundamentalmente:

a primeira é traduzir uma tarefa ou problema em uma fórmula estruturada (pseudocódigo). A segunda é traduzir esse pseudocódigo em código fonte que, quando compilado, irá performar a tarefa ou solucionar o problema. Ambas as traduções podem ser desafiadoras, requerindo definição precisa acerca de qual tarefa/problema trata (lógica), transformando-o em um conjunto preciso de instruções a serem desenvolvidas em quaisquer contingências, para que o algoritmo performe o esperado sob diferentes condições (controle). (Kitchin, 2017, p. 17)

Ainda que pareça um processo de construção automático e efetivamente passível de controle, o desafio posto pelos algoritmos e pela profusão de dados que impulsiona é justamente o de que a sociedade possa efetivamente comungar com as traduções realizadas pelos programadores. É a comunicação sobre os sentidos dos algoritmos e sobre as implicações que trazem para a dinâmica de organização da sociedade uma das mais subestimadas necessidades da política contemporânea.

Diante do panorama apresentado, assim, a pesquisa no campo da comunicação depara-se com pelo menos duas grandes categorias de questões acerca de seus desenvolvimentos atual e futuro. A primeira refere-se a um alinhamento necessário entre a ontologia e a metodologia da pesquisa, no sentido em que Hall (2003) empregou para discutir os fundamentos da pesquisa em política comparada. Trata-se de refletir acerca do que parece emergir como uma orientação epistemológica das ciências da comunicação, cuja máxima, apesar de situar-se no desenvolvimento técnico dos meios, não necessariamente contempla os pressupostos e implicações de seu uso para a organização teórico-metodológica das pesquisas.

A segunda categoria, inevitavelmente associada à primeira, engloba o receio do declínio da teoria em favor da dataficação da realidade na pesquisa científica, conforme descrito acima. Nesse âmbito, é pertinente repensar a tendência à utilização indiscriminada de concepções teóricas e conceitos deslocados dos contextos originários de suas produções, que descrevem e designam de forma anacrônica os conjuntos de dados e seus significados comunicacionais. Isso, defende-se aqui, é uma das garantias para a evolução do conhecimento científico em comunicação.

As experiências de imersão social na internet têm demonstrado que estamos diante de um novo contexto, de novas demandas e processos, permeados em todos os níveis

pela produção, difusão e apreensão de dados, por comunicações entre essas esferas, enfim. A Web 3.0 configura-se como uma espécie de palco para essas conexões entre dados, um novo tipo de espaço em que as relações acontecem de forma diversa, dispersa e múltipla (Amaral, 2012). Sugere-se que não existam muitos pontos em comum entre a comunicação conforme empreendida no âmbito do domínio das comunicações de massa e aquela que emerge nas vias digitais. Diante dessa perspectiva, seria errôneo interpretar essa última realidade com base em artefatos teórico-conceituais constituídos para explicar um outro padrão comunicacional da sociedade.

Sobre os conteúdos gerados pelos agentes sociais nas redes digitais, por exemplo, argumenta-se que “sua potencial classificação, até pela variedade da forma e diversidade das plataformas, ultrapassa categorizações standardizadas” (Amaral, 2012, p. 142). Se o *Big Data* e os metadados que permitem configurar são o novo suporte dos laços sociais, como também defende Amaral (2012), é preciso avançar velozmente na compreensão sobre sua semântica, sobre os sentidos cravados em sua concepção e evolução.

Essa preocupação remete também àquela expressa por Shaldbolt, Hall e Berners-Lee (2006) a partir da crescente necessidade de integração dos dados digitais em denominadores comuns de sentido. A Web Semântica defendida pelos autores é vislumbrada para além das chamadas “folksonomias”, linguagens compartilhadas por um número considerável de pessoas interessadas em replicar determinado dado por meio de informação, a repercuti-lo.

Na ciência da informação, essa organização do fluxo informativo e construção compartilhada de semânticas na web é caracterizada pela descentralização do acesso, e pela facilidade com que os dados são também difundidos através das mídias sociais. Os ícones dessas “folksonomias”, entretanto, são vários. Desde *hashtags* a *memes*, várias formas de representação estética podem representar as dinâmicas relacionais dos dados na internet. Entretanto, resta saber quais são as semelhanças e os paralelos, as comparações mais apropriadas a serem feitas entre esses conjuntos semânticos, e que viabilizam sua indexação a ontologias, conforme a definição de Shaldbolt, Hall e Berners-Lee (2006). As ontologias, nesse sentido, são capazes de remover as ambiguidades existentes entre os incontáveis caminhos pelos quais os dados transitam e são difundidos no espaço digital. São semânticas mais perenes, sobre as quais estão assentadas taxonomias para a compreensão do social no contexto da internet.

Por esses meandros, o *Big Data* parece configurar-se como um fenômeno promissor para a compreensão da realidade da comunicação social contemporânea e, além disso, aciona a comunicação e seus processos informacionais, definitivamente, como um eixo teórico e interpretativo relevante e necessário para as ciências humanas e sociais. A emergência dos dados como paradigma metodológico dominante revela a convergência entre as ciências exatas e as ciências sociais. Nesse escopo de investigação com o qual se defronta o fazer científico, não há como prescindir da comunicação como componente vital de explicação sobre as conexões hoje habilitadas pelas interrelações entre indivíduos e dados. Porém, cabe indagar como o fenômeno comunicação é concebido e difundido pelos investigadores em suas construções sobre o que se delineaia nas e pelas mídias e aparatos digitais.

## ONTOLOGIAS DA COMUNICAÇÃO E DESENHOS METODOLÓGICOS DAS PESQUISAS – ALGUNS APONTAMENTOS

De que realidade comunicacional os pesquisadores estão falando? Que perguntas conduzem as investigações científicas que se sustentam nos dados de internet?<sup>3</sup> Que ontologia sustenta as escolhas metodológicas de quem realiza as pesquisas? Isto é, defende-se que o ângulo desde o qual se orientam os trabalhos baseados no *Big Data* ajuda a conformar os caminhos metodológicos e teóricos das investigações. Por exemplo, a prevalência de uma visão da comunicação como fenômeno estático, como composto por retratos fixos da realidade, localizados em planos pré-determinados e controlados de observação direta e imparcial induz a um desenho metodológico de pesquisa que privilegia mecanismos lineares de circulação de informações.

Analogamente ao que Hall (2003) observou para as pesquisas em política comparada internacional, podemos dizer que, também no campo da comunicação, o valor dos métodos empregados depende de sua congruência com o modo como o pesquisador entende o funcionamento das causalidades dos processos que se propõe analisar. Assim, a captura de dados digitais num dado momento do tempo e num contexto específico de atuação representa o quê, exatamente, dentro do universo de explicações comunicacionais possíveis? Se a perspectiva de uma realidade linear da comunicação e do controle de seus resultados é o que predomina ontologicamente na pesquisa, o *Big Data* pouco terá a oferecer em termos de construção do conhecimento. Essa forma de representação da realidade, como composta por entidades fixas e organizadas aprioristicamente, em formatos lineares de causas e consequências, é limitada em sua capacidade explicativa. Ao denotar um modo de pensar sobre como a realidade social funciona, tal visão de mundo incorpora modelos lineares como padrões representativos da realidade (Abbott, 1998). Assim, nessa perspectiva dificilmente haverá avanços epistemológicos em termos de ampliação das possibilidades de explicação do fenômeno comunicacional. As pesquisas que enxergam a comunicação como fenômeno pontual, situado no tempo, e mais, veem seu propósito como instrumental e passível de controle pelas intenções do emissor, criam uma espécie de paradoxo na forma como interpretam os dados digitais.

Na internet, dados e informações parecem poder ser identificados, definitivamente, como unidades de uma complexidade social recortada pelas mais diversas variáveis e multilinear, portanto. Na medida em que são traduzidos por meio de expressões textuais, gráficas, imagéticas, auditivas e também por meio de estruturas e arquiteturas de relacionamento entre nós, os dados coletados em ambientes digitais permitem vislumbrar – e mesmo representar – a potência da comunicação humana.

Não se trata aqui de defender que o que ocorre na internet em termos de ação e expressão sociais é fundamentalmente distinto dos processos já experimentados pela “sociedade analógica”. Porém, se tomamos o plano digital da existência social como ponto de partida para explicá-la, é preciso cuidar para que um ajuste adequado seja feito

<sup>3</sup> Aqui, importante é reiterar que os dados de plataformas de redes sociais, ou das mídias sociais (conforme o conceito popularizado em inglês), são apenas um tipo de *Big Data*. O crescimento dessas pesquisas não tem sido tarefa exclusiva de cientistas da comunicação, mas também de cientistas políticos, antropólogos, psicólogos, geógrafos, cientistas da computação, enfim, acadêmicos de vários campos de conhecimento.



entre nossos estoques interpretativos até então acumulados como campo científico e essa realidade reflexiva e cotidianamente tecida por indivíduos e organizações. Sugere-se, portanto, a impossibilidade de sustentação das mesmas concepções sobre a realidade comunicacional do século XX neste momento da história, em que a internet compõe parte significativa do funcionamento da vida social.

Uma epistemologia baseada na complexidade é muito mais coerente com a dinâmica das mídias sociais, calcada no não-linear, nas multicausalidades e nos movimentos perpétuos observados por meio da expressão de tensões e de conexões em disputa. Redes de relacionamento, como o Facebook e o LinkedIn, combinam-se às plataformas informativas, como o Twitter. Redes sociais de *blogs*, como a Medium, têm substituído o papel dos tradicionais *websites* como locais de apresentação pública e identitária de pessoas e organizações, propiciando o surgimento e construção de uma nova ecologia das mídias que resulta da proliferação crescente de aplicativos identificada por Amaral (2012) em sua análise sobre a Web 3.0. Conteúdos audiovisuais ganham com o Vimeo e o YouTube significados que variam em função das percepções e apropriações que os usuários deles fazem. As mídias sociais deixam mais evidente, enfim, toda uma gama de variáveis que interpelam os processos de comunicação, denotando de forma mais clara a insuficiência de representações de mundo que se aprisionam exclusivamente em manifestações pontuais e autocentradas de emissões informativas. Em casos de plataformas e aplicativos móveis como o Snapchat ou as funcionalidades do tipo *stories* do Instagram e WhatsApp, as mensagens tornam-se, efetivamente, uma representação literal da vida cotidiana, num *streaming* editado e personalizado de acordo com contextos de produção adequados à visão de mundo do emissor e transformado continuamente pelas relações que estabelecem com interlocutores externos.

Circunscrever as mídias sociais como objeto de investigação da comunicação converte-se, dessa forma, no grande desafio de refinar os métodos e ferramentas disponíveis para que sejam capazes de apreender as mais variadas dinâmicas de expressão, as apropriações das tecnologias e de seus sentidos por emissores e receptores e, também, as improbabilidades comunicacionais (Luhmann, 2006) que permeiam a realidade digital. Nesse sentido, iniciativas como a de Procter, Vis e Voss (2013) são inspiradoras, pois propõem métodos capazes de incorporar de maneira mais explícita, na interpretação e representação da realidade virtual, o movimento inerente às disputas e às confluências de sentido dos processos de comunicação social. Os autores conseguiram elaborar uma ferramenta de coleta de posts do Twitter (*tweets* e *retweets*), em cuja codificação algorítmica está previsto o rastreamento da dinâmica de distúrbios, rumores e boatos na plataforma. Mais do que isso, a representação dos resultados das análises, numa linha temporal de ocorrências de mensagens emitidas e repercutidas no Twitter, preocupa-se com a fluidez das ações de compartilhamento de informações (as quais compõem uma parcela do *Big Data* disposto nas mídias sociais) com sua capacidade de difusão e com as disputas sobre sua veracidade, seus sentidos, enfim.

Cientistas da comunicação podem encontrar na internet, dessa maneira, um espaço privilegiado para a aplicação de ferramentas tecnológicas inovadoras, capazes de

apreender os percursos que as mensagens emitidas percorrem, e especialmente sua apropriação e ressignificação por instâncias receptoras. Os dispositivos interacionais de crítica midiática analisados por Braga (2006) podem, finalmente, de acordo com os novos escopos metodológicos e as ferramentas e técnicas de coleta e análise de dados aperfeiçoados com o *Big Data*, tornar-se visíveis em suas distintas camadas de construção coletiva.

Também as atualidades mediáticas, constituídas como objeto revelador de comunicações, ou de temas de interesse público, conforme a elaboração de Martino (2009), atravessam, com velocidade e sentidos por vezes inesperados, o cotidiano da internet. Entender os mecanismos pelos quais as atualidades emergem nos ambientes digitais, seja como expressão de volições individuais e organizacionais, seja como acontecimentos espontâneos e intrínsecos à dinâmica fluida das relações sociais, converte-se em relevante programa de pesquisa do campo da comunicação. O aprimoramento continuado das tecnologias de mineração e análise de dados digitais tem o potencial para criar instrumentos que mapeiem quadros mais completos da comunicação mediática, as trajetórias que abarcam as emissões de informações e seus direcionamentos subsequentes através das mídias sociais. Os caminhos abertos e trilhados pelas mensagens representariam, nesse sentido, uma arquitetura digital do espaço-tempo comunicacional. Por sua vez, as mensagens, com todas as suas nuances interpretativas, naturalmente emergentes e mutáveis, completam o imbricado quadro analítico da comunicação nas mídias sociais.

Diante desse panorama, a pesquisa sobre a comunicação pode superar, enfim, seu viés instrumental, na medida em que incorpore em suas descrições e análises não só a intencionalidade da emissão, mas também os pontos de fuga que escapam às tentativas de regulação de sentidos, conforme analisado por Fausto Neto (2011). A instrumentalização dos meios – que conformou a ortodoxia do campo da comunicação e ainda domina as esferas de produção de conhecimento direcionadas por e para instituições do mercado e do Estado – não é congruente, enfim, com os processos espontâneos de comunicação nitidamente observáveis nos ambientes da internet.

### TEORIAS E CONCEITOS DA COMUNICAÇÃO SOB O ÂNGULO DAS MÍDIAS SOCIAIS

Nesta seção do texto, retoma-se a ideia de Kitchin (2014), para quem a revolução paradigmática propiciada pelo *Big Data* não pode justificar a ausência da reflexão teórica. Pelo contrário, defende-se que a ascensão de uma sociedade dataficação engendra um momento propício para o aprofundamento e rediscussão sobre a validade dos conceitos e de seus pressupostos epistêmicos. Os problemas e perguntas de pesquisa embasados pelas teorias devem prevalecer, desse modo, como bússolas que norteiam o olhar empírico, e não serem descartados em nome de uma soberania de “dados que falam por si mesmos”.

O exercício contínuo de revisão e do debate sobre as bases teórico-conceituais do campo da comunicação é garantia de validade do conhecimento científico, afinal.

Supõe-se que disso depende a precisão da interpretação de dados que expressam sentimentos, visões de mundo, assertivas, tentativas deliberadas de persuasão e ações individuais e coletivas. Afinal, as mídias sociais e as informações que dispõem sobre a realidade dizem respeito a uma vasta gama de expressões da sociedade.

Dessa forma, torna-se apropriado e urgente debater como os esquemas conceituais e teóricos clássicos da comunicação podem contribuir para pensarmos questões de pesquisa que envolvam *Big Data*, ou como, numa relação inversa, os mesmos repertórios teóricos podem ser repensados a partir da realidade social apresentada pela internet. Um dos pontos de partida para reflexão, conforme argumentação anterior, é a própria natureza volátil, efêmera e contingente que permeia a produção dos dados utilizados para a consecução das pesquisas. As inúmeras plataformas e dispositivos em que os agentes sociais podem projetar suas vozes, suas imagens e ações concertadas são as bases dos dados que manipulamos. São atravessadas e construídas por motivações e interesses dos mais diversos, aparecendo e desaparecendo, como informações passíveis de análise, em movimentos ininterruptos e interconectados de expressões individuais e organizacionais.

Conforme esboçado na seção anterior, as mais diversas mídias sociais competem pela atenção de usuários, num universo repleto de possibilidades comunicacionais, cada qual com sua materialidade e simbolismo. Esse cenário demonstra que estilos de vida, interesses prioritários de comunicação social, localização espacial, condições econômicas, entre outros fatores, compõem o rol de variáveis que influem nas dinâmicas da sociedade digital. Contextos culturais e nacionais podem, por exemplo, induzir ao uso majoritário de uma plataforma em detrimento de outra. O Twitter é um caso muito interessante nesse sentido. Enquanto que em países como os Estados Unidos a plataforma é tida como um termômetro da política na esfera pública, no Brasil seu impacto não é tão abrangente. Por outro lado, estudar o tema da política contemporânea no Brasil sem incluir ao menos uma menção ao papel de destaque do Facebook como plataforma de debate público pode significar negligência.

Portanto, nesse contexto multifacetado, trabalhar de forma estanque com os conceitos de emissor, meio, mensagem e receptor, entre outros, desconsiderando o movimento inerente das conexões de sentido nas mídias digitais, parece ser contraproducente. Uma apropriação conceitual que exemplifica bem os dilemas enfrentados pela teoria é a da ideia de *broadcasting*, hoje aplicada para se compreender os dados gerados numa rede social digital unidirecional e centralizada de disseminação de informações. Acredita-se que tal conceito entrega pouco valor à compreensão da comunicação na internet, uma vez que os dados que normalmente designa não conseguem revelar, necessariamente, a apreensão de informações por uma suposta audiência, e muito menos o que é feito com base nessas informações. A retomada da concepção de audiência para fazer referência a grupos de seguidores ou amigos de um emissor do tipo *broadcasting* também recoloca a instância receptora numa condição de passividade que remonta a teorias como a da agulha hipodérmica. Mas, de fato, podemos interpretar as configurações da internet a partir de mecanismos conceituais que remontam ao contexto histórico da comunicação

do século XX, e portanto se valer de conceitos originários do campo para explicar a realidade atual?

Recorda-se que das informações, em seu estado puro, não se pode deduzir comunicação, cuja existência está atrelada à conexão compartilhada de significados e sentidos, mediados pela técnica e/ou por indivíduos. Seguidores de uma página no Facebook não deveriam ser equacionados, portanto, às audiências das sociedades de massa, muito embora a atuação de empresas como a Cambridge Analytica, embasada na psicométrica, adquira bastante relevo, em relação aos tradicionais meios de massa, no processo de conquista da atenção e persuasão de indivíduos por meio de dados pessoais. Conforme observado acima, uma visão de comunicação linear, mecânica e centrada na emissão não parece compatível com os processos fugidios e impermanentes que são passíveis de compreensão a partir do *Big Data*.

Por isso, defende-se que o paradigma teórico predominante para a condução das pesquisas de comunicação social nesse complexo cenário que interliga os contextos tradicionais da comunicação de massa aos contextos digitais de produção da realidade seja aquele proposto por Luhmann (1995) (Figura 1). A comunicação, nesta perspectiva, só se completa no momento em que a informação [information] emitida [mitteilung] é sistemicamente selecionada e compreendida [verstehen] podendo, a partir daí, originar novas informações e, potencialmente, novos ciclos comunicacionais.



Figura 1: A unidade comunicacional na teoria de Niklas Luhmann (1995)

É justamente essa ideia da comunicação como um elemento central de coordenação do social – como unidade materializada por ciclos ininterruptos de produção e desconstrução de sentidos – que melhor se compatibiliza com o intercâmbio instantâneo dos dados e informações nas mídias sociais ou com a concomitância das esferas de emissão e recepção, se utilizamos conceitos tradicionais do campo.

Sbardelotto, por outros caminhos teóricos, apresenta ideia convergente à perspectiva luhmanniana quando defende a comutabilidade produção-recepção nos processos

de circulação em rede e “que tais polos só existem reciprocamente, constituindo-se mutuamente mediante a dinâmica da circulação” (2017, p. 4). Não haveria respaldo empírico, nesse âmbito, para a reiteração de modelos teóricos de comunicação em que instâncias de produção e assimilação de informações estejam isoladas uma da outra. “No fluxo circulatório em redes comunicacionais, especificamente, os polos de produção e de recepção não desaparecem, mas não é mais possível fixá-los em um sujeito social específico, como as corporações midiáticas ou as instituições sociais tradicionais” (Sbardelotto, 2017, p. 19). A realidade da vida digital hoje é de que é possível para um agente revestir-se de emissor e receptor até de forma simultânea. Além disso, este agente pode ser falso, artificialmente controlado por algoritmos, ou representante de uma organização, de um coletivo ideológico, etc. São muitas as facetas de quem produz e quem consome informações.

Sem a pretensão de esgotar a reflexão sobre como teorias e conceitos da comunicação podem orientar as pesquisas no contexto de complexidade crescente que a internet revela, vale uma última reflexão sobre o significado conceitual da mensagem. Falamos das mensagens, em seu sentido convencional, quando optamos por perguntas de pesquisa que privilegiam o olhar sobre *hashtags*, “conversações”, compartilhamentos, curtidas, fotografias, registros audiovisuais ou *retweets* nas mídias sociais? Tal questão retoma o argumento acima expresso, acerca das ontologias da web semântica (Shaldbolt, Hall & Berners-Lee, 2006). Sem que estejam equacionados os parâmetros sobre os quais se constroem os espaços digitais, é difícil compreender como, efetivamente, podem favorecer a comunicação social. A hipótese é de que a escolha da plataforma de comunicação em que se exprime uma mensagem predispõe tipos específicos de conteúdo e estéticas também coerentes com sua utilização social. Isso tem consequências não apenas para a formulação de categorizações mais abrangentes sobre os significados das mensagens e seus conteúdos, seu acoplamento a meios (aplicativos seriam hoje apenas um dos tipos de meio), bem como a articulação entre essa variável que “carrega o sentido” e as dimensões sociais de sua circulação.

Fundamental é notar, além disso, que os sentidos coletivamente construídos em torno das mensagens podem ser interpretados à luz de distintos ângulos, desde as próprias mensagens originais que alcançam o mundo, passando pelo tipo de perfil digital que a tenha difundido inicialmente e chegando à configuração da rede social que a sustentou por um tempo em evidência. Essas escolhas assumem, sem dúvida, implicações epistemológicas sobre visões mais individualistas ou holísticas da realidade da internet.

#### NOTA CONCLUSIVA

Por último, mas não menos importante, cabe uma nota reflexiva sobre o papel da inteligência artificial, e também da presença dos chamados *bots*, na produção e circulação de informações e, portanto, na potência comunicacional da vida digital. O assunto emerge principalmente por meio de constatações da influência desses sistemas em processos político-eleitorais (Bessi & Ferrara, 2016), mas tem implicações ainda pouco

debatidas no âmbito científico. Seguindo a orientação filosófica de Dusek (2006), de acordo com a qual todos os meios tecnológicos de comunicação moldam, em alguma medida, as experiências sociais, propõe-se repensar, de forma radical, a visão clássica de uma exclusividade da agência humana nos processos comunicacionais. Questiona-se a visão objetificada da tecnologia, a partir da compreensão de que os *bots* têm adquirido, no contexto atual, capacidade de expressar e de processar informações, visto que estão constituídos como algoritmos. Conforme sugerido anteriormente neste texto a partir da necessidade da tradução de códigos algorítmicos para contingências, “desenvolvimentos em inteligência artificial habilitaram algoritmos para otimizar suas próprias decisões, gerando melhorias para eles mesmos” (Uricchio, 2017, p. 127). Autoria e agência, reitera-se, tornam-se conceitos relativizáveis. A heterodoxia de uma concepção teórica que atribui à técnica papéis potencialmente ativos nos processos comunicacionais é necessária, contudo, no movimento de transcendência do conhecimento científico balizado pela vida social digital. A validade deste pressuposto só pode ser mensurada no curso de procedimentos de coleta, sistematização e análise de *Big Data*, porém. O trabalho de análise sobre os rastros digitais mal começou. Se as ciências que se fundam nos dados quiserem fazer jus ao *status* de base para a democracia, como advoga Berners-Lee em recente entrevista (Baker, 2018), será preciso correr, preferencialmente na velocidade da internet. ✍

#### REFERÊNCIAS BIBLIOGRÁFICAS

- Abbott, A. (1998). Transcending general linear reality. *Sociological Theory*, 6, 169-186.
- Amaral, I. (2012). Participação em rede: do utilizador ao “consumer 2.0” e ao “prosumer”. *Comunicação e Sociedade*, 22, 131-147.
- Baker, D. (2018). Is the internet broken? WIRED asks Tim Berners-Lee, Jimmy Wales, James Ball, Wendy Hall and more about how we could – and should – reset the net. Start again. *Wired.*, Jan., UK Edition, 120-129.
- Bessi, A. & Ferrara, E. (2016). Social bots distort the 2016 Presidential election online discussion. *First Monday*, 21(11).
- Boyd, D. & Crawford, K. (2012). Critical questions for Big Data. *Information, Communication & Society*, 15(5), 662-679. doi: 10.1080/1369118X.2012.678878
- Braga, J. L. (2006). *A sociedade enfrenta sua mídia. Dispositivos sociais de crítica midiática*. São Paulo: Paulus.
- Coté, M., Gerbaudo, P. & Pybus, J. (2016). Introduction: politics of big data. *Digital Culture & Society*, 2(2), 5-15. doi: 10.14361/dcs-2016-0202
- Cukier, K. & Mayer-Schönberger, V. (2013). *Big Data: a revolution that will transform how we live, work and think*. Nova Iorque: Houghton Mifflin Harcourt.
- Dusek, V. (2006). *Philosophy of technology: an introduction*. Malden, MA: Blackwell Publishing.

- Fausto Neto, A. (2011). Comunicação das organizações: da vigilância aos pontos de fuga. In I. Oliveira & A. Soares (Eds.), *Interfaces e tendências da comunicação no contexto das organizações* (pp. 43-67). São Caetano do Sul, SP: Difusão Editora; Rio de Janeiro: Editora Senac Rio.
- Gil de Zúñiga, H. (2015). Citizenship, social media, and big data: current and future research in the social sciences. *Social Science Computer Review*, 1-7. doi: 10.1177/0894439315619589
- Hall, P. A. (2003). Aligning ontology and methodology in comparative research. In J. Mahoney & D. Rueschemeyer (Eds.), *Comparative historical analysis in the Social Sciences* (pp. 373-404). Cambridge: Cambridge University Press.
- Helbing, D. (2015). *The automation of society is next. How to survive the digital revolution*. Charleston: CreateSpace Independent Publishing Platform.
- Kitchin, R. (2014). Big Data, new epistemologies and paradigm shifts. *Big Data & Society*, April-June, 1-12. doi: 10.1177/2053951714528481
- Kitchin, R. (2017). Thinking critically about and researching algorithms. *Information, Communication & Society*, 20(1), 14-29. doi: 10.1080/1369118X.2016.1154087
- Kuhn, T. (1998). *A estrutura das revoluções científicas*. São Paulo: Perspectiva.
- Luhmann, N. (1995). *Social systems*. Stanford, CA: Stanford University Press.
- Luhmann, N. (2006). *A improbabilidade da comunicação*. Lisboa: Vega.
- Martino, L. C. (2009, junho). *Atualidade mediática: o conceito e suas dimensões*. Comunicação apresentada no XVIII Compós - Encontro da Associação Brasileira de Programas de Pós-graduação em Comunicação, Belo Horizonte.
- Milan, S. & van der Velden, L. (2016). The alternative epistemologies of data activism. *Digital Culture & Society*, 2(2), 57-74. doi: 10.14361/dcs-2016-0205
- Procter, R., Vis, F. & Voss, A. (2013). Reading the riots on Twitter: methodological innovation for the analysis of big data. *International Journal of Social Research Methodology*, 16(3), 197-214. doi: 10.1080/13645579.2013.774172
- Shaldbolt, N., Hall, W. & Berners-Lee, T. (2006). The semantic web revisited. *IEEE Intelligent Systems*, 96-101.
- Sbardelotto, M. (2017, junho). *Circulação em rede: a comutabilidade dos polos de produção e recepção no fluxo comunicacional digital*. Comunicação apresentada no XXVI Compós - Encontro da Associação Brasileira de Programas de Pós-graduação em Comunicação, São Paulo.
- Uricchio, W. (2017). Data, culture and the ambivalence of algorithms. In M. T. Schäfer & K. van Es. (Eds.), *The datafied society, studying culture through data* (pp. 125-137) Amsterdão; Amsterdam University Press.

## NOTA BIOGRÁFICA

Ana Thereza Nogueira Soares é doutoranda em Comunicação na Universidade de Brasília – UnB, no Brasil. Está vinculada à linha de pesquisa de Teorias e Tecnologias da Comunicação do Programa de Pós-Graduação e aos grupos de pesquisa Comunicação Organizacional e Pensamento Sistêmico – COMSiS e Comunicação, Tecnologia e Política – CTPol. Seus interesses de pesquisa estão focados nos fenômenos de comunicação

empreendidos no contexto de organizações e movimentos da sociedade civil, considerando, especialmente, suas formas de expressão e comunicação no espaço digital.

E-mail: [anatsoares@gmail.com](mailto:anatsoares@gmail.com)

Morada: SQN 411, Bloco E, ap. 304 CEP: 70866-050. Brasília-DF. Brasil.

\* **Submetido: 30.11.2017**

\* **Aceite: 15.03.2018**