

BIG DATA E O CONSUMO DE NOTÍCIAS NAS REDES SOCIAIS

BIG DATA AND THE CONSUMPTION OF NEWS IN SOCIAL NETWORKS

Paulo Pinheiro Gomes Jr. ¹

RESUMO

Em um mundo cada vez mais conectado, as pessoas compartilham os seus dados publicamente na Web e, de forma geralmente consentida, costumam abrir mão de parte da sua privacidade. Surge, dessa forma, uma nuvem de dados *on-line* repleta de vestígios digitais que podem ser rastreados, medidos, quantificados, analisados e mensurados. A organização e a armazenagem de toda essa informação não estruturada em um formato capaz de ser aproveitado em termos jornalísticos pode apontar um novo caminho para o consumo de notícias.

Palavras-chave: *Big Data*. Redes Sociais. Webjornalismo. Métricas. Gestão da Informação.

ABSTRACT

In a world increasingly connected, people share their data publicly on the web, and in a way generally consented, usually giving up some of their privacy. Arises, in this way, an online data cloud full of digital traces that can be traced, measured, quantified, analyzed and valued. The organization and storage of all this unstructured information in a format that can be used in journalistic terms can point a new way for the consumption of news.

Keywords: *Big Data*. Newsmaking. Webjournalism. Metrics. Information management.

¹ Paulo Pinheiro é Mestre em Ciências da Comunicação pela PUCRS, jornalista e professor da ESPM-Sul. Foi editor do Portal Terra e do Portal clicRBS. Atualmente é o coordenador do Portal de Jornalismo da ESPM-Sul. Trata-se de artigo inédito. *E-mail:* ppinheiro@espm.br.

1 REDES SOCIAIS E CONSUMO DE NOTÍCIAS

Atualmente vive-se um tempo no qual as redes sociais e os dispositivos móveis avançam e conquistam mais espaço, tornando-se parte importante da sociedade (RECUERO, 2008). O correto uso de ambos pode contribuir para o aperfeiçoamento da difusão de informações (CHRISTOFOLETTI & LAUX, 2008; ZAGO, 2008). A evolução tecnológica provocou um crescimento capaz de provocar uma constante digitalização dos processos de difusão das informações. Os ambientes virtuais foram inundados com uso de aplicações para dispositivos móveis. Mais importante ainda: a realização de conteúdos audiovisuais digitais nessas plataformas começa a ser tornar uma realidade.

Esse talvez seja, no entendimento de Cruz (2002), um fenômeno da globalização, a qual ocorreu de tal forma, que é possível conceber a forma de funcionamento dos *media* como um grande sistema mundial, por meio do qual fluem quantidades ilimitadas de informações, notícias, propaganda política, publicidade, cinema, jornais, revistas, programas de rádio e televisão, através de diversos suportes tecnológicos.

Nesse sentido, a palavra global ganha um sentido nunca visto antes. Já em 1964, McLuhan foi um visionário ao cunhar o conceito de “aldeia global”, simbolizando o efeito global do aumento da velocidade e do alcance das tecnologias da comunicação.

A experiência humana desta forma torna- -se global e o global torna-se intimamente conhecido. Assim apresentava o novo mundo da aldeia global no qual a circularidade eléctrica acabou por derrubar o regime do tempo e espaço (CASCAIS, 2001, p. 23).

Porém, no nível da comunicação e informação, observa-se, simultaneamente, uma revalorização do local, fazendo com que as informações globais disputem a atenção com as locais. Metaforicamente pode-se afirmar que da “aldeia global” de McLuhan emerge a “nossa aldeia” (CAMPONEZ, 2002).

Essa busca de informações locais numa sociedade globalizada talvez seja uma boa demonstração de que mesmo o encurtamento das distâncias é incapaz de apagar por completo os laços que ligam o indivíduo a determinados territórios (sejam eles físicos ou simbólicos). A “aldeia global” e a tentativa de uniformizar padrões – nesse caso, um dos exemplos mais sintomáticos seria o imperialismo cultural dos Estados Unidos da América – gerou uma ação em sentido contrário como o surgimento de uma defesa das culturas nacionais, as quais erguem as suas identidades com orgulho e não cedem à proclamada fatalidade de imperialização cultural.

Pelo contrário, historicamente, as informações locais sempre tiveram importância e são várias as pequenas comunidades que têm os seus títulos de imprensa e rádios, em que se dão a conhecer as notícias locais, as histórias da sua comunidade, num trabalho que, muitas vezes, é fruto da colaboração com essa mesma comunidade. Os meios de comunicação de âmbito local (tal como os *media* em geral) funcionam como uma espécie de próteses/extensões identitárias, tal como referiu McLuhan.

Tal situação provocou uma mudança de comportamento nos consumidores de notícias. É, de fato, pouco provável acreditar que a maneira como as pessoas se informam permaneça igual, sem alterações significativas, por tanto tempo. Eles, portanto, se tornaram mais conscientes e desconfiados de uma mídia que não oferece espaço para que possam expor suas opiniões e pensamentos.

[...] os cidadãos vêem assim reforçada a possibilidade de uma participação mais activa em processos de deliberação, num quadro de interacção que é agora muito diferente daquele proporcionado pelas

tecnologias da comunicação mais convencionais (rádio e televisão, ou mesmo a imprensa), cujas características evidenciam fortes condicionalismos de unidireccionalidade (ESTEVEZ, 2007, p. 220).

De acordo com Recuero (2009), as redes sociais foram capazes de ampliar as possibilidades de conexões, porém também conseguiram aumentar a capacidade de difusão de informações.

No espaço offline, uma notícia ou informação só se propaga na rede através das conversas entre as pessoas. Nas redes sociais online, essas informações são muito mais amplificadas, reverberadas, discutidas e repassadas (RECUERO, 2009, p. 25-26).

Vale notar que, em um primeiro momento, tamanha quantidade de informação circulando poderia ser considerada prejudicial para o jornalismo. Afinal, ficaria difícil separar a informação verdadeira no meio de uma avalanche de informações desencontradas. Mas Kovach e Rosenstiel (2003) discordam dessa visão:

À medida que os cidadãos encontram um fluxo cada vez maior de dados, eles têm mais necessidade – não menos – de fontes reconhecidas dedicadas a verificar aquela informação, salientando o que é importante saber e filtrando o que não é. O papel da imprensa nesta nova era torna-se trabalhar para responder a questão ‘onde está o bom material?’. A verificação e a síntese tornam-se a espinha dorsal do novo papel do gatekeeper do jornalismo, o de fazedor de sentido (sensemaker) (KOVACH; ROSENSTIEL, 2003, p. 64).

Percebe-se que a questão do *newsmaking* permanece atual, mesmo em um tempo no qual as redes sociais funcionam como ferramentas de informação. O *newsmaking*, como se sabe, busca revelar o que está por trás da escolha de um fato como notícia. O que leva um acontecimento a possuir o carimbo, a classificação de jornalístico? O que o torna tão relevante, tão especial? Como ocorre o deslocamento de algo cotidiano para o *status* de notícia?

A hipótese comunicacional do *Newsmaking* é um estudo mais ligado à teoria do jornalismo (HOHLFELDT, 2001), mas, por tradição, está relacionado a uma perspectiva da Comunicação. Durante um período, aproximadamente 50 anos, várias teorias ligadas aos processos de comunicação foram criadas (WOLF, 2003). Da teoria hipodérmica aos *cultural studies*, muito foi realizado nessa área de pesquisa.

Porém, a partir dos anos 60, nos Estados Unidos, surge o que se costuma chamar de *communication research*, que buscava compreender de maneira mais ampla o processo comunicacional por meio do cruzamento de informações provenientes das diversas teorias existentes. Basicamente, eram estudos que buscavam perceber o impacto social dos meios de comunicação de massa.

Portanto, no seu conjunto, as teorias dos mass media, ligadas à abordagem sociológica empírica, defendem que a eficácia da comunicação de massa está largamente associada e depende de processos de comunicação não provenientes dos mass media e que existem no interior da estrutura social em que o indivíduo vive (WOLF, 2003, p. 50).

Os estudos de *newsmaking* começaram através do conceito de *gatekeeper*. A primeira menção à ideia ocorre em um estudo elaborado, em 1947, por Kurt Lewin. Ele concebeu um estudo sobre as dinâmicas interativas nos grupos sociais. O pesquisador analisou problemas ligados à mudança de hábitos alimentares.

Lewin, que era psicólogo, identificou canais nos quais fluíam certa sequência de comportamentos relativos a determinado campo. Tal atitude provocava a criação de uma cancela ou porteira. Além

disso, essa cancela também poderia ser encontrada na sequência de informações por meio de canais de comunicação por parte de um grupo. Ele investigava sobre as decisões domésticas relativas à aquisição de alimentos para a casa.

Mais tarde, em 1950, David Manning White ampliou o conceito criado por Lewin trazendo a ideia para o jornalismo.

[...] o fluxo de notícias tem de passar por vários gates, isto é, portões, que não são mais do que áreas de decisão em relação às quais o gatekeeper, o jornalista, tem de decidir se vai escolher essa notícia ou não. Se a decisão for positiva a notícia acaba de passar pelo portão; se não for, sua progressão é impedida, o que na prática significa a sua morte, porque significa que a notícia não será publicada (TRAQUINA, 1993, p. 69).

O estudo de White foi feito com base nas informações de Mr. Gates. Este era um jornalista de meia-idade de um periódico dos Estados Unidos. Gates anotou durante uma semana os motivos que o levaram a recusar determinadas matérias em detrimento de outras.

De posse das informações de Mr. Gates, a pesquisa de White revelou que, das 1.333 explicações de recusa de uma notícia, quase 800 foram devidas à falta de espaço, aproximadamente 300 citavam a sobreposição com matérias já previamente selecionadas, a falta de interesse jornalístico ou de qualidade de escrita. Outros 76 eventos foram considerados por demais distantes do público leitor do jornal, por isso, sem relevância.

As decisões do gatekeeper são realizadas menos em uma base de avaliação individual de noticiabilidade do que em relação a um conjunto de valores que incluem critérios tanto profissionais quanto organizacionais, como a eficiência, a produção de notícias, a velocidade (ROBINSON, 1981, p. 97).

A conclusão de White mostrou que o processo de seleção de notícias era subjetivo e arbitrário. Além disso, a teoria do *gatekeeper* analisa as notícias levando em conta somente quem as produz, ou seja, o jornalista. É, nesse sentido, uma teoria restritiva.

Atualmente, para Bruns (2003), existe um “*Gatewatching*” nas redes sociais. Canavilhas (2010) complementa e salienta que o “*gatewacher* emerge assim como um elemento central num ecossistema mediático onde a fragmentação motivada pela multiplicação de fontes e o excesso de informação obrigam os media a disputarem a atenção dos leitores”.

Tal afirmativa pode ser comprovada por meio de um estudo da Retrevo² (2010), que confirma que 42% dos norte-americanos começam o dia consultando o *Facebook* ou o *Twitter* em busca de informações. Outros 16% garantem que é por meio dessas redes sociais que procuram saber sobre as notícias matutinas.

Um relatório do *Pew Research Center's*³ (2010) aponta para resultados semelhantes: cerca de metade dos americanos opta pela leitura das notícias nas redes sociais. O motivo? Eles confiam na seleção noticiosa que os seus amigos fazem. O estudo mostra ainda que 75% dos consumidores de notícias *on-line* recebem *links* para essas notícias via *e-mail* ou redes sociais e 52% partilham igualmente notícias por essas plataformas. Mais ainda, a pesquisa destaca que 51% dos usuários de redes sociais, como, por exemplo, o *Facebook*, garantem que todos os dias leem notícias disponibilizadas por algum dos seus amigos. Além disso, cerca de 23% seguem notícias disponibilizadas nessas redes por empresas de mídia tradicionais ou jornalistas.

² Para mais detalhes, acessar: <http://www.retrevo.com/content/node/1324>.

³ Para mais detalhes, acessar: <http://www.pewinternet.org/Reports/2010/Online-News.aspx>.

Brambilla (2005) destacou que existem três vertentes de como as redes sociais podem ser incorporadas no dia a dia de quem trabalha com a notícia: na apuração (busca por fontes, personagens, pautas, testemunhos, opiniões); na veiculação (linguagem adequada aos *medias* sociais, grupos e momentos certos para divulgação de determinadas notícias); e no *feedback*/relacionamento com o público (aproveitar a quantidade de informação espontânea e gratuita para melhorar o trabalho)⁴.

Dessa forma, é interessante perceber que existem muitas pessoas utilizando as redes sociais para se informar, só que, ao mesmo tempo, esses indivíduos passam a disponibilizar seus dados e suas preferências de forma pública, também se tornando fontes para uma pesquisa. E, nesse sentido, surge a importância de ferramentas de *Big Data*.

2 **BIG DATA E O FUTURO DA COMUNICAÇÃO**

O futuro pertence ao *Big Data*. Os cientistas da computação, físicos, economistas, matemáticos, cientistas políticos, comunicadores sociais, sociólogos e muitos outros estão clamando para ter acesso às grandes quantidades de informações produzidas por e sobre pessoas, coisas e suas interações. Diversos grupos argumentam sobre os potenciais benefícios e custos da análise de informações do *Twitter*, *Google*, *Verizon*, *Facebook*, *Wikipedia* e outros espaços nos quais grandes grupos de pessoas deixam algum vestígio digital.

Questões importantes surgem. Uma análise em larga escala de DNA poderá ajudar a curar doenças? Ou será que o *Big Data* será capaz de inaugurar uma nova onda de desigualdade no tratamento médico, na qual somente quem tiver dinheiro para um cálculo tão complexo será beneficiado?

A grande questão levantada pelo *Big Data* é a seguinte: será que a análise de dados ajuda a tornar o acesso das pessoas à informação mais eficiente e eficaz? Ou será que vai ser usada para controlar os manifestantes nas ruas das grandes cidades que fazem protestos contra o governo? Será que vai transformar a maneira como estudamos a comunicação humana e a cultura, reduzir a paleta de opções de pesquisa ou até mesmo alterar o que significa “pesquisa”?

Antes de mais nada, é preciso entender melhor o conceito. *Big Data* é, em muitos aspectos, um termo pobre. Como Lev Manovich (2011) observa, tem sido utilizado no campo das ciências para se referir a um conjunto de dados suficientemente grande para necessitar de supercomputadores. Atualmente, não há a menor dúvida de que existem quantidades de dados gigantescas disponíveis.

Pode-se considerar um *Big Data* a Receita Federal, o Hospital das Clínicas de São Paulo, a Companhia de Engenharia de Tráfego de São Paulo (CET), o Controle de Tráfego Aéreo, o New York Times entre outros. No caso da Receita Federal, por exemplo, é um *Big Data* não disponível, ou seja, com acesso livre ao público. Mas existem repositórios abertos de dados denominados de Open Data, que possuem dados públicos e podem ser manuseados por quem se interessar (LIMA Jr., 2011, p. 5).

Como prova dessa situação, em 2011, pesquisadores da Universidade do Sul da Califórnia fizeram uma estimativa de quanta informação foi produzida entre os anos de 1986 e 2007. O resultado foram impressionantes 296 exabytes. Para se ter uma ideia, esse valor equivale a uma pilha de CDs

⁴ Para mais detalhes, acessar: <http://webmanario.com/2009/12/15/quem-ignora-o-que-o-publico-diz-em-midias-sociais-nao-pode-ser-jornalista/>.

de 400 metros de diâmetro, capaz de ultrapassar a altura da Lua⁵. Vale notar que 1 exabyte equivale a cerca de 1 bilhão de gigabytes.

Os pesquisadores estimaram o quanto de informação é guardado em todos os computadores e dispositivos tecnológicos, papéis, livros, gravações e até mesmo o *chip* de memória do seu cartão de crédito. A seguir, eles calcularam quanto dessa informação pode ser transmitida, seja pela internet, pela televisão, pelo rádio e por celulares. Por último, foi calculada a quantidade final de informação que pode ser armazenada, chegando ao resultado final.

Mais ainda: nos EUA, as empresas com mais de 1.000 funcionários, em média, produzem, em um ano, 235 terabytes de dados. Esse número de informações supera todo o conteúdo existente na Biblioteca do Congresso dos EUA⁶. Basta pensar sobre o que poderia estar acontecendo na sua empresa agora: sensores embutidos em máquinas podem coletar dados de operações. Ao mesmo tempo em que comerciantes usam as mídias sociais e os dados de localização de *smartphones* para entender peculiaridades de compra dos adolescentes.

Sim, os números são realmente grandes, mas essa não é a característica mais importante desse novo ecossistema de dados. *Big Data* não é notável por causa de seu tamanho, mas por ser capaz de fazer relações entre os dados existentes. Ou seja, informação por si só não é poder, mas, sim, as relações, os padrões que revelam a compreensão sobre o motivo pelo qual determinado fato ocorreu.

Por exemplo, a rede de varejo norte-americana *Wal-Mart* recentemente contratou a *Hewlett Packard* para a construção de um *data warehouse* capaz de armazenar quatro petabytes (4000 trilhões de bytes) de dados, representando cada compra gravada por seus terminais de ponto de venda (isso significa cadastrar cerca de 267 milhões de transações por dia) em suas lojas em todo o mundo⁷. Através da aplicação dos conceitos do *Big Data*, eles podem detectar padrões que indicam a eficácia de suas estratégias de preços e campanhas publicitárias e gerenciar melhor seus estoques e as cadeias de suprimentos.

Além disso, *Big Data* é importante porque se refere a um fenômeno analítico. É graças a algoritmos complexos que se vislumbram padrões antes desconhecidos, porque grandes quantidades de dados podem oferecer conexões que se irradiam em todas as direções. Por isso, é fundamental para começar a fazer perguntas sobre os pressupostos analíticos e metodológicos incorporados no *Big Data*.

O futuro aponta para o surgimento de uma era de grandes volumes de dados. Essas análises ocorrem em um ambiente de incerteza e mudanças rápidas. Dessa forma, as decisões atuais terão um impacto considerável no futuro. Com o aumento da automação da coleta e análise de dados – bem como algoritmos que podem extrair e informar determinados padrões de massa no comportamento humano –, é preciso perguntar quais sistemas estão levando essas práticas e que se regulam. Lawrence Lessig (1999) argumenta que os sistemas são regulados por quatro forças: o mercado, a lei, as normas sociais e arquitetura – ou, no caso da tecnologia, o código. Quando se trata de *Big Data*, essas quatro forças estão juntas, ainda que, muitas vezes, em desacordo.

⁵ Para mais informações, acessar: <http://www.tecmundo.com.br/curiosidade/8567-pesquisadores-estimam-a-quantidade-de-informacao-existente-no-mundo.htm>.

⁶ Para mais informações, acessar: <http://www1.folha.uol.com.br/fsp/mercado/me0209201029.htm>.

⁷ Para mais informações, acessar: <http://www.informationweek.com/software/information-management/walmart-extends-teradata-mega-deal/228800661>.

O mercado enxerga o *Big Data* como oportunidade pura: comerciantes podem usá-lo para direcionar publicidade, seguradoras querem otimizar suas ofertas e os banqueiros veem a possibilidade de fazer melhores leituras sobre o mercado financeiro.

Nos Estados Unidos, chegou a ser proposta uma legislação para coibir a coleta e a retenção de dados, geralmente, devido a preocupações sobre a privacidade. Características como a personalização permitem o acesso rápido à informação mais relevante, mas apresentam questões éticas difíceis de serem resolvidas (PARISER, 2011).

Apesar de todas essas precauções, muitas empresas têm investido profundamente em TI e possuem grandes conjuntos de dados para explorar. Indústrias também podem usar esses dados de forma inovadora para competir através da adoção de técnicas analíticas sofisticadas.

O setor público é o terreno mais fértil para a mudança. Governos coletam grandes quantidades de dados, realizam transações comerciais com milhares de cidadãos. O uso do *Big Data* pode ser uma maneira de melhorar as ações governamentais e produzir um estado mais eficiente e eficaz. Os benefícios potenciais da utilização dessa técnica são grandes.

3 O JORNALISMO NA ERA DO *BIG DATA*: O QUE MUDA?

A discussão sobre a adoção pelo jornalismo das tecnologias é uma questão antiga. Remonta ao conceito de jornalismo de precisão expressado por Meyer (1973). Apesar de revolucionário, o método não trouxe uma alteração efetiva no modo de produção jornalística.

Quem descobriu essa possibilidade foi Philip Meyer, um jornalista que decidiu trabalhar a informação por intermédio de uma base de dados. Na época, Meyer fazia parte da equipe da *Detroit Free Press*. Com o auxílio de um computador *mainframe*, produziu uma matéria sobre distúrbios raciais em Detroit e foi capaz de traçar a demografia dos negros na cidade. Nesse ponto, surgiu a “integração de computadores e a ciência social e ele foi o precursor na área de *Computer Assisted Reporting* (CAR) (ROYAL, 2010)”.

Porém, quase três décadas depois – como consequência da evolução tecnológica no campo dos computadores (*hardware*) e da computação (*software*) –, Meyer revisitou o conceito e observou que “com modesta melhora quantitativa em rapidez e veracidade, o jornalismo de precisão preserva a função clássica do jornalista” (MEYER, 1973, p. 231). E, justamente por isso, segue como uma alternativa importante:

As normas básicas não diferem das que nós sempre temos trabalhado. Descobrir os fatos e contar o significado deles sem perda de tempo. Se existem instrumentos que nos autorizam a cumprir esta tarefa com maior poder, veracidade e luzes, nós deveríamos utilizá-los ao máximo” (MEYER, 1973, p. 15)

Os avanços nesse sentido, no campo da Comunicação Social, foram introduzidos de forma paulatina nos grupos de mídia. De acordo com Vianna (1992), na década de 80, os PCs começaram a ser introduzidos nas redações brasileiras. O primeiro jornal a adotar terminais de computador em suas redações foi a Folha de São Paulo. O projeto do periódico paulista começou a ser elaborado em 1968, mas somente foi posto em prática em 1983.

Porém, levou tempo até que os equipamentos fossem efetivamente parte da rotina dos jornalistas brasileiros. Vale notar que, no final dos anos 60 e início dos 70 (ou seja, no século passado), o *The*

New York Times estruturou o primeiro banco de dados. Esse foi inserido diretamente nas etapas de produção da notícia. Em 1968, a agência de notícias Reuters já utilizava máquinas computacionais nas conexões da sua rede interna para o gerenciamento da demanda de notícias recebidas.

O Brasil, como nota Lima Jr., levou mais tempo para aderir à mudança tecnológica provocada pela adoção de PCs nas redações:

Assim, a cultura de apropriação tecnológica, no caso da produção do Jornalismo brasileiro, possui uma história de atraso em relação aos principais centros. O intervalo de implementação tecnológica, no que tange aos investimentos, pelos parques brasileiros muitas vezes é explicado pela diferença entre a economia estadunidense e a brasileira nas quatro últimas décadas do século XX (LIMA JR., 2012, p. 209).

Essa situação despertou o interesse dos pesquisadores. Para Barbosa e Ribas (2007), Nilson Lage, professor titular aposentado da UFSC, foi o primeiro a discutir no Brasil a questão da aplicação das bases de dados como uma ferramenta útil para o trabalho jornalístico em dois artigos publicados em 2002⁸.

Apesar do interesse acadêmico, ainda existe um certo receio de aplicar com mais frequência o uso de ferramentas de banco de dados nas redações:

O descompasso entre as demandas da prática profissional e o modelo de adoção dos computadores nas redações impediu uma modificação no sistema de produção de conteúdos nas organizações jornalísticas. Muito desta defasagem se deve a difusão de conceitos como jornalismo de precisão ou reportagem assistida por computador, capazes de reduzir a tecnologia a um uso instrumental porque aperfeiçoa o trabalho sem desestabilizar os fundamentos da prática. A consolidação no jornalismo digital pressupõe a compreensão de que a tecnologia representa a possibilidade de criação de um formato distinto de jornalismo em que todas as etapas do sistema de produção de conteúdos – desde a apuração a circulação – são circunscritas aos limites do ciberespaço (MACHADO, 2003, p 10).

Como nota Lima Jr. (2010), a mídia centralizada (*broadcasting*) enfrenta a concorrência de múltiplas plataformas digitais, que possuem conexões através das redes telemáticas descentralizadas e de baixa hierarquia, permitindo trocas de dados por meios físicos (cabos) ou sem fio. Essa estrutura proporciona novas possibilidades de consumo de conteúdo jornalístico, alterando a relação estabelecida, pelo modelo *broadcasting*, entre a audiência e as suas preferências informacionais.

É fácil perceber essa alteração quando se leva em conta os índices de audiência. No caso da TV, existe uma queda no número de televisores ligados, por consequência, a diminuição da audiência. E essa queda tem correlação direta com a Web devido, por exemplo, ao acesso às redes sociais, na substituição da atenção televisiva⁹.

Justamente pelo crescimento em importância das redes sociais, surge uma massa de dados não estruturados, de vestígios digitais prontos para serem analisados. Contudo, a quantidade espantosa de dados disponível não quer dizer muito se não puder ser relacionada, transformada em informação estruturada e, no caso do Jornalismo, utilizada para construir conteúdo de relevância social, aproveitando a Era do *Big Data* para criar valor em diversos caminhos: criando transparência (órgãos públicos); habilitando descobertas experimentais, criando segmentações (exemplo: dados personalizados); substituindo/auxiliando processos de decisão (algoritmos) e inovando nos modelos de negócio.

⁸ *Jornalistas-Robôs. A era das máquinas inteligentes e No futuro do jornalismo, a integração com os computadores.*

⁹ Fantástico perde audiência para Redes Sociais, DVD, Games e TV a Cabo. Disponível em: <http://www.cidademarketing.com.br/2009/blog/mercadologia/50/fantstico-perde-audincia-pararedes-sociais-dvd-games-e-tv-a-cabo.html>. Acessado em: 26 jun. 2013.

De certa forma, pode-se afirmar que o *Big Data* pode, em um futuro próximo, provocar uma mudança na composição das redações jornalísticas. Com uma cada vez mais provável aproximação com cientistas da computação.

Ainda hoje, ser um jornalista que mexe com dados pode significar coisas muito distintas. O profissional pode ser um processador de estatísticas. Pode ser um designer de interação criativo. Pode ser um repórter que usa o traquejo com dados – extraíndo a notícia desse conteúdo e/ou explicando o viés ali contido – para cobrir sua área. Embora os papéis ainda estejam sendo definidos, é raro encontrar equipes de cientistas da informação e matemáticos em organizações jornalísticas (BELL, 2012, p. 56).

Como nota Fuller (2010), o mundo da informação, em que a notícia é apenas uma parte, tem sido produzido pelos instrumentos tecnológicos. Para ele, os jornalistas necessitam:

[...] conseguir ir além da nostalgia e da utopia pelo entendimento do que está acontecendo no nível mais fundamental, que é frequentemente escondido. Útil como as diferentes explicações para as mudanças que inundam o jornalismo algumas vezes, forças mais profundas estão em ação. Existe mais do que a Internet. (FULLER, 2010, p. 4).

Vale destacar que o *Big Data* depende de três V's: grande *Volume* de dados, grande *Velocidade* possível e a capacidade necessária para processá-los, além da sua *Variedade*, devido ao fato de que os dados são extraídos e obtidos de fontes diversas, tais como *e-mails*, *blogs*, *redes sociais*, etc.

Todavia, ainda são necessários mais dois V's importantes. A *Veracidade*: afinal de contas, é necessário que os dados sejam autênticos e façam sentido (HURWITZ *et al.*, 2013, p. 16). Além disso, também é necessário *Valor*. Ou seja, é fundamental que os dados acrescentem valor ao seu utilizador (BEULKE, 2011). Somente dessa forma o enorme investimento necessário para o *Big Data* não se transformará em uma despesa inútil.

Mas, a partir do exposto, o que se pode concluir em relação ao futuro do jornalismo? O *Big Data* é uma ferramenta que tem potencial para mudar o modo como o público consome notícias?

4 CONSIDERAÇÕES FINAIS

Não há dúvida de que o *Big Data* tem potencial para revolucionar o jornalismo, mas existem alguns mitos que devem ser discutidos como, por exemplo, a falsa 'garantia' de que os dados serão mantidos anônimos antes de serem processados, uma vez que já foi demonstrado que geralmente é possível torná-los públicos justamente por fazer o cruzamento com outras bases de dados (CRAWFORD, 2013).

Também é questionável se o *Big Data* vai mudar nossas vidas para melhor. Muito do que foi trabalhado com a ferramenta até o momento diz respeito a aperfeiçoar mecanismos de incitação ao consumo. Deve-se, nesse mesmo sentido, ter um cuidado para não cair na falácia de afirmar que "os números falam por si". Isso é parcialmente verdade, mais importante do que mostrar estatísticas é compreender o que os dados estão indicando. Não se deve descartar a necessidade de modelos, teorias ou de uma análise crítica.

O mercado de *Big Data* já movimentava 26 bilhões de dólares em todo o mundo. Das 500 maiores companhias globais, 450 têm projetos nessa área. Como, a partir de agora, empresas de menor porte devem passar a usar essa ferramenta, estima-se que o mercado de *Big Data* chegue a 38 bilhões de dólares em 2015¹⁰.

¹⁰ Para mais detalhes, acessar: <<http://exame.abril.com.br/revista-exame/edicoes/1025/noticias/para-nao-se-afogar-em-numeros?page=3>>.

Ainda assim, como o jornalismo pode fazer parte deste mercado de *Big Data*? Não é difícil imaginar que a coleta de dados não estruturados pode ser utilizada para saber as preferências do leitor de modo a descobrir padrões de interesse. Ou seja, seria possível para o consumidor de notícias receber informações sobre questões que possam vir a preocupar o leitor. Se os algoritmos identificam que a pessoa está com um problema de colesterol alto, podem ser oferecidas notícias sobre tratamentos para combater a enfermidade.

Mais ainda: a partir do momento em que os algoritmos identificam qual é o trajeto que o leitor faz para se deslocar de casa para o trabalho, ele passa a enviar notícias e alertas personalizados no caso de haver um acidente que dificulte o trânsito ou mesmo uma possível alteração de rota em função de uma obra.

Todavia, essa customização do noticiário pode revelar um lado sombrio e perverso. Os padrões detectados pelo *Big Data* podem levar a uma minimização do mundo jornalístico. Leitores podem manifestar apreço por um tipo específico de informação em detrimento de outra. Por exemplo, pode haver um desejo maior por notícias de celebridades e de esportes e, em contrapartida, fatos políticos e econômicos podem terminar esquecidos. Essa distorção vai permitir que muitos se aprofundem no raso e saibam cada vez mais sobre menos.

Como se vê, as possibilidades de aplicação do *Big Data* no jornalismo são reais e indicam que a profissão pode estar se aproximando de uma era cada vez mais computacional.

REFERÊNCIAS

BARBOSA, Suzana; RIBAS, Beatriz. **Mapeamento conceitual e metodológico preliminar sobre as bases de dados no ciberjornalismo**. Artigo apresentado no I Colóquio Internacional Brasil-Espanha sobre Cibermeios. Faculdade de Comunicação, Universidade Federal da Bahia, Salvador (Brasil), 3 a 7 de Dezembro, 2007.

BELL, Emily. Números do jornalismo. **Revista de Jornalismo da ESPM**, ed. 2, p. 55-57, jan./fev./mar. 2013.

BEULKE, Dave. **Big Data Impacts Data Management: The 5 Vs of Big Data** [Blog post]. Disponível em: <<http://davebeulke.com/big-data-impacts-data-management-the-five-vs-of-big-data/>>. Acesso em: 24 jun. 2013.

BOLTER, Jay David; GRUSIN, Richard. **Remediation**. Understanding New Media. Cambridge: The MIT Press, 1999.

BOYD, Danah; CRAWFORD, Kate. **Six Provocations for Big Data**. Social Science Research Network. Working Paper Series, 2011. Disponível em: <<http://ssrn.com/abstract=1926431>>. Acesso em: 27 jun. 2013.

ELLISON, Nicole B. et al. Social network sites: Definition, history, and scholarship. **Journal of Computer-Mediated Communication**, v. 13, n. 1, p. 210-230, 2007. Disponível em: <<http://jcmc.indiana.edu/vol13/issue1/boyd.ellison.html>>. Acesso em: 27 jun. 2013.

BRAMBILLA, Ana Maria. Jornalismo open source em busca de credibilidade. **Intercom**, 2005.

BRUNS, Axel. Gatewatching, not gatekeeping: Collaborative online news. **Media International Australia Incorporating Culture and Policy**: quarterly journal of media research and resources, v. 107, p. 31-44, 2003.

CAMPONEZ, Carlos. **Jornalismo de Proximidade**. Coimbra: Minerva Coimbra, Coleção Comunicação, 2002.

CANAVALHAS, João. Do gatekeeping ao gatewatcher: o papel das redes sociais no ecossistema mediático. In: **Congresso Internacional Comunicación**. 2010.

CASCAIS, Fernando. **Dicionário de Jornalismo**: As palavras dos media. Lisboa: Editorial Verbo, 2001.

CHRISTOFOLETTI, Rogério; LAUX, Ana Paula França. Confiabilidade, credibilidade e reputação: no jornalismo e na blogosfera. **Intercom – Revista Brasileira de Ciências da Comunicação**, São Paulo, v. 31, n. 1, p. 29-49, 2008. Disponível em: <<http://www.portaldemidia.ufms.br/wp-content/uploads/2011/03/4809-12999-1-PB.pdf>>. Acesso em: 27 jun. 2013.

CORREIA, João Carlos. Novo jornalismo, CMC e esfera pública. **Recensio – Revista Eletrônica de Recensões de Comunicação e Cultura**, 2002. Disponível em: <<http://www.bocc.ubi.pt/pag/correia-joao-jornalismo-cmc-esfera-publica.pdf>>. Acesso em: 9 jun. 2013.

CRAWFORD, Kate. **Think Again**: Big Data. *Foreign Policy*, 9. May. 2013. Disponível em: <http://www.foreignpolicy.com/articles/2013/05/09/think_again_big_data>. Acesso em: 25 jun. 2013.

CRUZ, João Cardoso. **Introdução ao Estudo da Comunicação**: Imprensa, Cinema, Rádio, Televisão e Redes Multimédia. Lisboa: ISCSP, 2002.

DAVENPORT, Thomas H.; BECK, John C. **The attention economy**: Understanding the new currency of business. Harvard Business Press, 2001.

DOROGVTSEV, Sergei N.; MENDES, José F. **Evolution of networks**: From biological nets to the Internet and WWW. Oxford University Press, 2003.

ESTEVES, João Pissarra. Os novos media na perspectiva da democracia deliberativa: sobre redes e tecnologias de informação e comunicação. **Espaços públicos, poder e comunicação**. Porto: Edições Afrontamento, p. 209-224, 2007.

FULLER, Jack. What is happening to news?. *Dædalus*, v. 139, n. 2, p. 110-118, 2010.

HOHLFELDT, Antônio; MARTINO Luiz C.; VEIGA, Vera; (Org.). **Teorias da Comunicação**. Petrópolis: Vozes: 2001.

HURWITZ, Judith et al. **Big Data for Dummies**. Hoboken, NJ: John Wiley & Sons, 2013.

KOVACH, Bill; ROSENSTIEL, Tom. **Os elementos do jornalismo**. São Paulo: Geração Editorial, 2003.

LEMO, André. Cibercultura: alguns pontos para compreender a nossa época. **Olhares sobre a cibercultura**. Porto Alegre: Sulina, p. 11-23, 2003. Disponível em: <<http://www.facom.ufba.br/ciberpesquisa/andrelemos/cibercultura.pdf>>. Acesso em: 21 jun. 2013.

LESSIG, Lawrence. **Code**: And other laws of cyberspace. NY: Basic Books, 1999.

LÉVY, Pierre. **Cibercultura**. São Paulo: Ed. 34, 1999.

LIMA JUNIOR, Walter Teixeira. Big Data, Jornalismo Computacional e Data Journalism: estrutura, pensamento e prática profissional na Web de dados. **Estudos em Comunicação**, n. 12, 207-222, dez. 2012. Disponível em: <<http://www.ec.ubi.pt/ec/12/pdf/EC12-2012Dez-11.pdf>>. Acesso em: 26 jun. 2013.

_____. Jornalismo computacional em função da “Era do Big Data”. **Libero – São Paulo**, v. 14, n. 28, p. 45-52, dez. 2011. Disponível em: <http://www.casperlibero.edu.br/rep_arquivos/2011/12/12/1323717839.pdf>. Acesso em: 25 jun. 2013.

MACHADO, Elias. **O ciberespaço como fonte para os jornalistas**. Salvador: Calandra, 2003.

MANOVICH, Lev. **Trending**: The promises and the challenges of big social data. *Debates in the digital humanities*, p. 460-75, 2011. Disponível em: <http://www.manovich.net/DOCS/Manovich_trending_paper.pdf>. Acesso em: 15 mai. 2013.

MCLUHAN, Marshal. **Os meios de comunicação como extensões do homem**. 10. ed., São Paulo: Cultrix, 1995

MEEKER, Mary. **Internet Trends**. Disponível em: <<http://www.businessinsider.com/mary-meechers-latest-incredibly-insightfulpresentation-about-the-state-of-the-web-2012-5#-1>>. Acesso em: 17 mai. 2013.

MEYER, Philip. **Precision Journalism**. Bloomington: Indiana University Press, 1973.

OLIVEIRA, Nelson; HOLANDA, André. Jornalismo participativo e informação hiperlocal: o papel de mashups e hashtags na construção da notícia em redes sociais. In: **Iniciacom - Revista Brasileira de Iniciação Científica em Comunicação**, v. 2, n. 1. 2010.

PALÁCIOS, Marcos. Cotidiano e Sociabilidade no Cyberespaço: apontamentos para discussão. **O indivíduo e as mídias**. Rio de Janeiro: Diadorim, p. 87-104, 1996.

PARISER, Eli. **The filter bubble: What the Internet is hiding from you**. Penguin UK, 2011.

PAVLIK, John V. **Journalism and new media**. Columbia University Press, 2001.

POSTER, Mark. **A Segunda Era dos Média**. Oeiras: Celta, 2000.

RECUERO, Raquel. Comunidades em redes sociais na internet: um estudo de caso dos fotologs brasileiros. **Liinc em Revista**, v. 4, p. 63-83, 2008.

_____, Raquel. **Redes sociais na Internet**. Porto Alegre: Sulina, 2009.

REICHHELD, Frederick F. **The loyalty effect: The hidden force behind growth, profits, and lasting value**. Harvard Business Press, 2001.

RETREVO (2010). **Is Social Media a New Addiction?** Disponível em: <<http://www.retrevo.com/content/node/1324>>. Acesso em: 27 jun. 2013.

RHEINGOLD, Howard; ARANHA, Helder; BUESCU, Jorge. **A comunidade virtual**. Lisboa: Gradiva, 1996.

ROBINSON, Gertrude Joch. **News agencies and world news: in Canada, the United States and Yugoslavia: methods and data**. University Press, 1981.

ROSEN, Jay. **The people formerly known as the audience**. 2006. Disponível em: <http://journalism.nyu.edu/pubzone/weblogs/pressthink/2006/06/27/ppl_frmr.html>. Acesso em: 17 ago. 2013.

ROYAL, Cindy. The journalist as programmer: A case study of the New York Times interactive news technology department. In: INTERNATIONAL SYMPOSIUM IN ONLINE JOURNALISM, Austin. **Anais...** Austin, TX, USA: The University of Texas at Austin, 2010.

THE PEW INTERNET & AMERICAN LIFE PROJECT (2010). **Understanding the Participatory News Consumer**. 2010. Disponível em <<http://www.pewinternet.org/Reports/2010/Online-News.aspx>>. Acesso em: 23 jun. 2013.

TRAQUINA, Nelson (Org.). **Jornalismo: questões, teorias e estórias**. Lisboa: Veja, 1993.

VIANNA, Ruth Penha Alves. **A informatização da imprensa brasileira**. Loyola, 1992.

WOLF, Mauro. **Teorias da comunicação de massa**. Martins Fontes, 2003.

ZAGO, Gabriela da Silva. O Twitter como suporte para produção e difusão de conteúdos jornalísticos. **Ciberlegenda**, n. 21, 2009. Disponível em: <<http://www.scribd.com/doc/5887184/O-Twitter-como-suporte-para-producao-e-difusaode-conteudos-jornalisticos>>. Acesso em: 27 jun. 2013.