

Big data e direito concorrencial: da necessidade de intervenção regulatória no âmbito das tecnologias voltadas ao processamento de dados

Big data and antitrust law: the need for regulatory intervention on data processing technologies

Leonardo Arêba Pinto*

RESUMO

Busca-se responder a seguinte problemática: as tecnologias conhecidas como *big data analytics* demandam a intervenção do direito concorrencial? Desta feita, estabelece-se uma distinção entre o *big data* e as tecnologias de processamento de dados. Embora parte da doutrina não enxergue um esquadro de atuação do direito concorrencial no que tange ao *big data*, dada até mesmo a abrangência desse conceito, procura-se demonstrar que mostra-se muito mais plausível cogitar a interferência do direito concorrencial nas tecnologias emergentes cujo escopo consiste no processamento de dados. Desta feita, em primeiro momento, os conceitos fundamentais inerentes ao termo "*big data*", bem como a inadequação aparente do direito concorrencial em face ao *big data* serão explorados. Em segundo momento, o potencial impacto das tecnologias emergentes voltadas ao processamento de dados atualmente em perspectiva será analisado. Por fim, pretende-se demonstrar a interseção existente entre o direito concorrencial e as tecnologias de processamento de dados.

Palavras-chave: Direito concorrencial. *Big data*. Processamento de dados. *Big data analytics*. Aprendizado de máquina.

ABSTRACT

This work aims to answer the following question: do technologies known as "big data analytics" demand the intervention of the competition law? Thus, this article establishes a distinction between the big data and data processing technologies. Although part of the doctrine do not relate the competition law to big data, given to the broadness of the concept, the article seeks to demonstrate that it is highly plausible to cogitate the interference of competition law in the emergent technologies which goal consist in data processing. As follows, in the first moment, the basic concepts inherent to the term "big data" will be explored, as well as the apparent inadequacy of the competition law in face of big data technology. In a second moment, the potential impact of the emergent technologies focused on processing data currently in perspective will be analyzed as well. Eventually, the article intends to demonstrate the intersection between the competition law and the technologies of data processing.

Keywords: Antitrust law. Big data. Data processing. Big data analytics. Machine learning.

1 Introdução

O conceito atribuído ao termo *big data*, em realidade, demonstra uma indeterminação, muito em virtude da velocidade com que a tecnologia de armazenamento de processamento de dados se desenvolve, bem como em virtude da abrangência desse conceito.

Além disso, até o presente momento não se estabeleceu uma forma de intervenção regulatória pontual por parte dos órgãos concorrenciais. Em realidade, a maior parte da literatura inerente à presente discussão considera uma atuação concorrencial em abstrato, a partir de possíveis condutas danosas à concorrência de mercado ideal.

Em função disso, parte da doutrina se posiciona no sentido de que o *big data*, já em face das situações previstas em abstrato, demanda inexoravelmente a intervenção estatal em dimensão concorrencial,

* Mestre em direito. Doutorando em direito. Professor de direito do UniProjeção – Centro Universitário e Faculdade Projeção. Advogado e sócio titular do Arêba Pinto Advogados.

como se verá no presente estudo¹. Noutra giro, outra parte da doutrina já se posicionou no sentido de que não existe, atualmente, elementos suficientes para justificar a intervenção regulatória, considerando-se o conhecimento pertinente ao conceito de *big data*².

Em virtude dessa dissidência doutrinária, e considerando-se a relativa imprecisão do conceito de *big data*, pretende-se, no presente estudo, estabelecer fundamentos estruturais da supracitada tecnologia, distinguindo-se as tecnologias voltadas ao processamento de informação insurgentes em mercado dos marcos teóricos relativos ao *big data* a serem retratados nesta oportunidade.

Parte-se da hipótese de que a noção de *big data*, a ser explorada nesta oportunidade, limita-se a um conceito essencialmente voltado à velocidade de circulação, criação e armazenamento de informação, distinguindo-se das tecnologias emergentes voltadas ao processamento de dados vinculados ao *big data*, conhecidas como *big data analytics*, já que as formas de processamento e tratamento da informação podem sofrer incalculáveis variáveis.

A partir de tal hipótese é que se alcança a problemática que direciona o presente estudo: as tecnologias conhecidas como *big data analytics*, voltadas ao processamento de dados, demandam a intervenção do direito concorrencial?

Para que seja possível solucionar a problemática suscitada neste artigo, mostra-se necessário, primeiramente, estabelecer os limites do conceito de *big data*, explicitando suas características a fim de que seja viável distingui-las do conceito que se atribuirá à tecnologia voltada ao processamento de dados.

Em segundo momento, as tecnologias emergentes e as suas repercussões no tratamento

da informação serão exploradas, fortalecendo-se a distinção estabelecida entre tais elementos e o *big data* e viabilizando uma conexão com a temática regulatória.

Em terceiro momento, a intervenção regulatória, tendo em vista o prisma concorrencial no âmbito das tecnologias voltadas ao processamento de dados, será analisada, possibilitando ao menos tatear um horizonte delineado no tocante ao tema.

Por fim, tanto a distinção conceitual estabelecida em momento inicial do artigo como a análise abstrata da intervenção regulatória serão reiteradas, assentando a conclusão de que os órgãos concorrenciais teriam muito mais espaço de atuação no âmbito das tecnologias voltadas ao processamento de dados.

A título de metodologia, o presente artigo se utilizará de embasamentos doutrinários de forma concentrada, tendo em vista a escassez de referenciais casuísticos aproveitáveis.

2 Apresentação sumária do conceito de *big data*

Além de ser necessário estruturar uma noção basilar do que é o *big data* a fim de que seja possível alcançar uma resposta satisfatória para a problemática suscitada neste artigo, trata-se de etapa fundamental para destacar a relevância pontual das tecnologias voltadas ao processamento de informação.

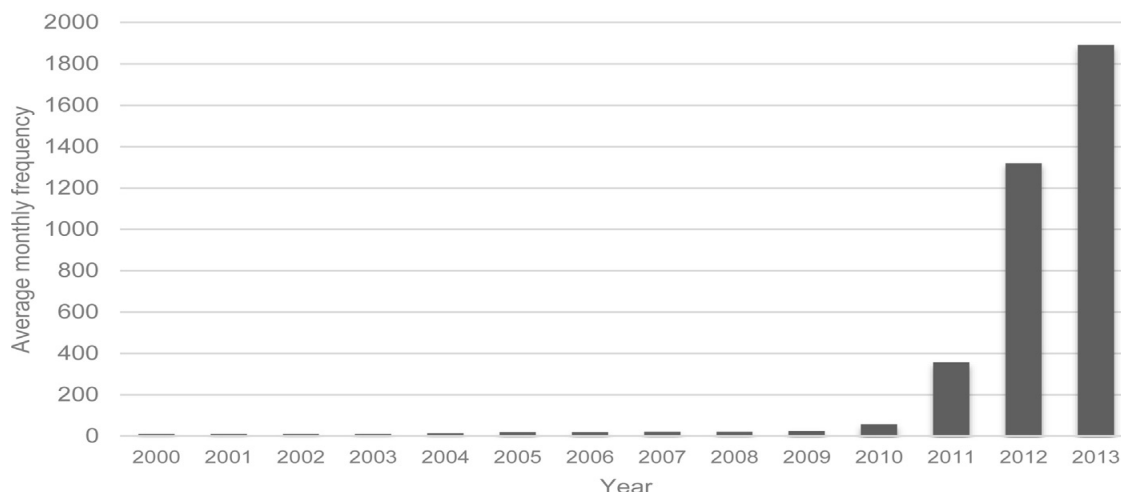
2.1 Das características elementares

Conforme dito, ainda em 2015, o conceito de *big data* já se mostrava crescente, com apresentação de repercussão exponencial em 2011, conforme demonstra o quadro a seguir reproduzido³:

¹ GRUNES, Allen P. and STUCKE, Maurice E., No mistake about it: the important role of antitrust in the era of big data (April 28, 2015). Antitrust source (Apr. 2015), Online; *University of Tennessee Legal Studies Research Paper* n. 269. Trata-se de artigo científico elaborado em 2015 no qual resta evidente que o *big data* demanda intervenção regulatória em prisma concorrencial, tendo em vista diversos argumentos a serem explorados brevemente nesta oportunidade.

² SOKOL, D. Daniel Sokol; COMERFORD, Roisin, *Antitrust and regulating big data*, 23 Geo. Mason L. Rev. 1129 (2016), available at <https://scholarship.law.ufl.edu/facultypub/803/>. O trabalho científico de Daniel Sokol e Roisin Comerford afirma que o *big data*, como se conhece à época de elaboração de seus estudos, não representa por si só um fator que justifique a intervenção de atuação regulatória por parte de órgãos concorrenciais, estabelecendo um conceito objetivamente mais palpável de *big data*.

³ GANDOMI, Amir. HAIDER, Murtaza. Beyond the hype: big data concepts, methods, and analytics. *International Journal of Information Management*, Volume 35, Issue 2, April 2015, Pages 137-144, International Journal of Information Management. "Frequency distribution of documents containing the term "big data" in ProQuest Research Library", p. 3.



O quadro acima reproduzido transmite a frequência com que o termo *big data* foi buscado em sítios eletrônicos voltados à pesquisa, demonstrando que, embora tal termo tenha apresentado inegável aumento de utilização, ainda se mostrava desprovido de um conceito objetivo.

Resta possível, no entanto, garantir a máxima aproximação permitida do conceito de *big data* difundido no âmbito doutrinário, sendo necessário para tanto destacar os *três Vs* inerentes a tal tecnologia⁴.

De fato, o primeiro “v” considerando como um elemento estrutural vinculado à tecnologia da *big data* é o *volume*, elemento vinculado à magnitude de informações e dados a circularem em rede e a serem armazenados por agentes de rede⁵. Vale destacar

que o acesso a uma magnitude sem procedência de dados é disponibilizado tanto aos usuários que estruturam mecanismos de coleta de informação quanto aos usuários que trafegam em rede, adquirindo informações e se vinculando a *sites* e aplicativos numa intensidade altíssima.

Importante ressaltar que a própria noção de volume no tocante ao *big data* admite variações conforme o tipo de dado coletado muda, e conforme o dispositivo de armazenamento desses dados também se moderniza. Sob este prisma, efetivamente a noção de *big data* depende da quantidade de dados que se concebe em determinada rede ou sistema, consistindo este aspecto em elemento inerente ao conceito deste termo tão utilizado atualmente.

O segundo “v” diz respeito à *variedade* da informação circulada, que pode ser classificada em estruturada, semiestruturada e desestruturada. Trata-se do grau de processamento e organização dos dados coletados, que a partir de sua captação recebem um tratamento e forma de armazenamento diferenciados⁶.

A forma como a informação é tratada é extremamente relevante para o presente estudo, sendo que a forma como os dados coletados forem

⁴ DIEBOLD, F. X. (2012). A personal perspective on the origin(s) and development of “big data”: The phenomenon, the term, and the discipline (Scholarly Paper No. ID 2202843). *Social Science Research Network*. Retrieved from http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2202843, p. 8.

⁵ “A survey conducted by IBM in mid-2012 revealed that just over half of the 1144 respondents considered datasets over one terabyte to be big data (SCHROECK, SHOCKLEY, SMART, ROMERO-MORALES, & TUFANO, 2012). One terabyte stores as much data as would fit on 1500 CDs or 220 DVDs, enough to store around 16 million Facebook photographs. BEAVER, KUMAR, LI, SOBEL, and VAJGEL (2010) report that Facebook processes up to one million photographs per second. One petabyte equals 1024 terabytes. Earlier estimates suggest that Facebook stored 260 billion photos using storage space of over 20 petabytes”. Pelo que consta, uma pesquisa do IBM apontou que os bancos de dados com mais de um terabyte de carga correspondem ao que se considera *big data*. GANDOMI, Amir. HAIDER, Murtaza. Beyond the hype: Big data concepts, methods, and analytics. *International Journal of Information Management*, Volume 35, Issue 2, April 2015, pages 137-144, International Journal of Information

Management. “Frequency distribution of documents containing the term “big data” in ProQuest Research Library”, p. 4.

⁶ GANDOMI, Amir. HAIDER, Murtaza. Beyond the hype: Big data concepts, methods, and analytics. *International Journal of Information Management*, Volume 35, Issue 2, April 2015, pages 137-144, International Journal of Information Management. “Frequency distribution of documents containing the term ‘big data’ in ProQuest Research Library”.

tratados e utilizados pode redundar em condutas anticompetitivas sujeitas à regulação concorrencial, foco da terceira seção deste estudo.

No entanto, somente o aspecto relativo à variedade da informação é tido como inerente ao conceito de *big data*, e não a forma como tais informações são processadas. De fato, é sábio desvincular as diversas formas de processamento de informações do elemento *variedade* inerente ao *big data*, de forma a possibilitar maior especificidade científica e até maior objetividade por parte de órgãos concorrenciais em suas intervenções.

A este ponto, logo, já restou iniciada a separação conceitual essencial ao desenvolvimento regular deste estudo, assentando-se o segundo “v”, variedade, como característica relativa aos tipos de informações coletadas e, no máximo, em relação ao grau de organização que tais dados admitem, sendo totalmente distinto do tipo de destinação e processamento que esses dados admitem.

O terceiro “v” diz respeito à *velocidade* com que os dados inerentes à rede de acesso civil são criados e eventualmente processados e direcionados. O processamento de dados dentro deste requisito diverge das tecnologias que serão apresentadas neste estudo, posto que o foco a ser analisado nesta oportunidade se refere a tecnologias específicas cujo objetivo específico é garantir formas privilegiadas de processar informações.

Pelo que se vê da doutrina pertinente, o fator velocidade é sempre considerado como característica que se ressalta pelo pouco tempo de transmissão da informação, o que torna o *big data* um conjunto de dados e informações em volume extremamente abrangente e variado sendo processado e transmitido entre programas, usuários, *hardwares* e estruturas empresariais e governamentais, delineando um conceito tanto jurídico como técnico extremamente abrangente.

Pode-se ainda destacar, neste prisma, outros Vs que foram acrescentados àqueles majoritariamente suscitados no âmbito de um panorama doutrinário. Desta feita, vale mencionar os termos *veracidade*, que põe em cheque a credibilidade da informação repassada em magnitude e volume estratosféricos, *variabilidade*, que diz respeito à inconstância do fluxo de informação repassado entre usuários de rede, bem

como entre mecanismos de captação e processamento de informação⁷.

Com o desenvolvimento das características estruturais do *big data* realizado neste estudo de forma sumária, restou demonstrado que o conceito de tal instituto, ao menos juridicamente/legalmente, ainda é inconsistente e de extrema abrangência, gerando, até mesmo, uma imprecisão quanto à incidência de regulamentação concorrencial.

Para que seja possível, portanto, desenvolver a solução da problemática, analisar-se-ão os aparentes empecilhos da regulação concorrencial do *big data* e distinguir-se-á esse conceito genérico para uma noção de *big data analytics*, que diz respeito às etapas e formas de processamento de dados dotados das características já exploradas.

Vale destacar, neste ponto, que a análise promovida neste estudo é desprovida de qualquer consideração de regulação a partir da natureza da informação e, por consequência, da proteção de dados como direitos humanos, a despeito da criação da Lei Geral de Proteção de Dados⁸, embora se admita desde já que tais fundamentos podem repercutir na análise de viabilidade de uma intervenção regulatória⁹. Pretende-se prosseguir nesta análise tendo como escopo consideração em abstrato dos riscos que o *big data analytics* pode oferecer à concorrência ideal primada pela legislação antitruste.

2.2 Da aparente inaplicabilidade do direito concorrencial ao *big data*

Como dito, é possível afirmar que uma porção da doutrina pertinente considerou inexistir justificativa para a intervenção de uma regulação concorrencial no âmbito do *big data*. No entanto, entende-se que esse posicionamento advém da extensão do conceito atribuído ao instituto do *big data*, que não distingue a

⁷ GANDOMI, Amir. HAIDER, Murtaza. Beyond the hype: Big data concepts, methods, and analytics. *International Journal of Information Management*, Volume 35, Issue 2, April 2015, pages 137-144, International Journal of Information Management. “Frequency distribution of documents containing the term ‘big data’ in ProQuest Research Library”.

⁸ Lei Geral de Proteção de Dados. Acesso em 12/07/2019. Disponível em: http://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/L13709.htm.

⁹ LANCIERI, Filippo Maria, Digital protectionism? Antitrust, data protection, and the EU/US transatlantic rift. *Journal of Antitrust Enforcement*, 2018.

noção de *big data analytics*. Logo, entende-se que os fundamentos como os seguintes se aplicam à noção já explanada de *big data*.

Primeiramente, a título de fundamentação da inexistência de uma intervenção regulatória, suscita-se o argumento de que não existem casos concretos que demandem a intervenção do direito antitruste na regulação do *big data*, ou seja, nenhum desafio que demande a elaboração de um modelo legal que se aplique a um problema fático¹⁰.

De fato, a problemática relativa ao *big data*, até mesmo no âmbito deste estudo que é feito sob um prisma peculiar ao que já se observa no campo doutrinário, é desprovida de casos práticos. Todavia, essa dificuldade metodológica não inviabiliza hipóteses que representam efetivos prejuízos ao modelo concorrencial almejado pelos órgãos antitrustes/concorrenciais.

Suscita-se também a justificativa de que a informação isolada não é suficiente para justificar a interferência da legislação antitruste, sendo totalmente inviável à regulação concorrencial interferir em informações desprovidas de qualquer direcionamento ou processamento específico.

Ora, tal justificativa advém de uma concepção abrangente e inerente à noção de *big data* já desenvolvido até o momento neste estudo, e é totalmente desvinculada da noção de *big data analytics*, a ser desenvolvido no capítulo 3 deste artigo. De fato, a informação nua e crua é totalmente irrelevante para a regulação concorrencial, enquanto a forma como os dados e informações são processados pode repercutir diretamente na concorrência ideal almejada para um mercado.

A partir deste ponto, já é possível constatar a distinção que o presente estudo busca estabelecer para fins de facilitação metodológica: por vezes as características inerentes ao *big data* por si só servem como empecilhos à intervenção até mesmo hipotética de órgãos concorrenciais, sendo que a abrangência do conceito atribuído a esse instituto gera essa imprecisão regulatória.

Vale prosseguir nas justificativas doutrinárias para a não incidência da regulação concorrencial, postura

que tão somente reforçará a força do argumento sendo construído.

Tem-se também que plataformas diferentes demandam informações altamente distintas entre si, tornando impossível a intervenção regulatória concorrencial objetiva para fins de preservação de uma concorrência ideal¹¹.

Ora, este aspecto tão somente reforça a tese sendo construída, no sentido de que, em função do *big data* ser extremamente abrangente, torna-se necessário focalizar o ponto de atenção das agências e órgãos concorrenciais às tecnologias de processamento de dados, considerando-se em específico as tecnologias que possuem destinações e comportamentos previsíveis, até mesmo em função do setor de mercado em que tecnologias de processamento de dados são implementadas.

Ademais, vale suscitar que existe ainda o argumento de que nem mesmo a regulação das fusões se mostraria pertinente, já que o caso da fusão entre *Facebook* e *Whatsapp* não resultou em barreiras para ingresso no mercado, pois os usuários podem transitar entre plataformas; trata-se de um mercado volátil e a fidelização não é tão estrita¹².

¹¹ SOKOL, Daniel; COMERFORD, Riso. Antitrust and regulating big data. *George Mason Law Review*, v. 199. 2016.

¹² “O *Big Data* assume um papel fulcral no seio da economia digital. O modelo de negócio das plataformas multilaterais online depende da aquisição e tratamento de dados pessoais em larga escala (*Big Data*) e a concorrência nestes mercados envolve cada vez mais características relacionadas com a exploração de dados pessoais, o que significa que a mesma estratégia empresarial (uma operação de concentração) pode criar, simultaneamente, preocupações ao nível da protecção de dados pessoais e do direito da concorrência. Na concentração *Facebook/WhatsApp* a CE reconheceu a privacidade enquanto parâmetro da concorrência não baseado no preço sem que, no entanto, tivesse levado a cabo qualquer apreciação a esse respeito, o que demonstra não apenas a inexistência de métodos de análise de parâmetros como a qualidade, mas também a predominância do preço enquanto parâmetro jusconcorrencial. Nesta decisão, a CE considerou que, no mercado de serviços de comunicações entre consumidores online o *Facebook* e o *WhatsApp* eram aplicações complementares, ignorando que a concorrência ao nível da privacidade é mais sequencial do que simultânea, o que significa que pode ocorrer por dissimelhança porquanto a adopção de determinada tecnologia ou política de privacidade por parte de uma empresa, que seja considerada atraente pelos consumidores, causa um efeito mimético nos concorrentes. A CE subestimou, assim, a pressão concorrencial exercida pelo *WhatsApp* sobre o *Facebook* e vice-versa, não tendo analisado a diminuição da concorrência na circunstância em que a maioria dita as escolhas de privacidade da maioria. Nesta decisão não foi, ainda, analisada a forma como esta concentração poderia aumentar as barreiras à entrada em resultado dos efeitos de rede potenciados pelos dados. Os efeitos

¹⁰ SOKOL, Daniel; COMERFORD, Rison. Antitrust and regulating big data. *George Mason Law Review*, v. 199. 2016.

Por fim, não se verificou um mercado significativo em âmbito doutrinário no qual a informação seja o produto relevante, mas sim as propagandas e a interação entre os usuários, ainda mais considerando-se o efeito de rede na propagação de dados.

Ora, novamente o argumento voltado ao processamento de dados sob o escopo do *big data analytics* se mostra válido, pois mostra que a destinação das informações coletadas bem como a forma de utilização de tais ativos variam conforme a tecnologia que venha a ser aplicada.

Desta feita, estabelece-se desde já uma distinção conceitual que permite o presente estudo reservar os argumentos que se mostram contrários à intervenção regulatória ao termo *big data* de forma independente do que se ora concebe como *big data analytics*, viabilizando uma discussão sobre a tecnologia exclusivamente voltada ao processamento de dados e reservando a abrangência do conhecimento comum tido como *big data*.

Prossegue-se, portanto, à exploração do que seria o *big data analytics*, a fim de que seja possível delinear uma atuação concorrencial mais objetiva.

3 Da destinação das tecnologias voltadas ao processamento de dados: do conceito de *big data analytics*

Atualmente, é possível afirmar que, efetivamente, a informação coletada por si só não representa motivo que justifique a intervenção regulatória por parte

de rede ajudaram o Facebook a manter o poder de mercado que detinha do lado gratuito e do lado pago da plataforma multilateral que opera, não sendo despiendo recordar que actualmente o Facebook e a Google constituem um oligopólio no mercado da publicidade online. A CE apreciou os efeitos sobre a concorrência apenas no lado pago de uma plataforma multilateral, olvidando a circunstância de que muitos operadores de negócios multilaterais concorrem, no lado gratuito do mercado, em muitos parâmetros não relacionados com o preço, tais como a qualidade do serviço e, cada vez mais, na quantidade de dados recolhidos relativamente aos utilizadores. Efectuou, assim, uma análise incompleta tendo demonstrado uma visão redutora, e talvez ingénua, acerca da possibilidade de o Facebook, por via da recolha e utilização de dados pessoais de utilizadores do WhatsApp, controlar uma quantidade colossal de dados pessoais, remetendo para as regras de protecção de dados quaisquer preocupações ao nível da privacidade que pudessem surgir desta concentração por entender que estas não são abrangidas pelo âmbito das regras de concorrência". BIDARRA, Ana Isabel Rodrigues Gomes. *O impacto do Big Data no direito da concorrência: o caso Facebook-WhatsApp e o controlo de concentrações na União Europeia*. Faculdade de Direito da Universidade de Lisboa, junho de 2018, p. 125.

de órgãos concorrenciais, já que não se vislumbra que a grande variedade ou volume de dados gere, obrigatoriamente, condições que prejudiquem a concorrência ideal¹³.

Em realidade, o grande poder transformador do *big data* reside no aumento do poder de decisão do detentor de tais informações e dados, capacidade que, caso seja desenvolvida ou promovida de determinadas formas, pode inegavelmente repercutir no que se espera de uma concorrência ideal de mercado.

Logo, o *big data analytics*, suscitado no presente texto até o momento, consiste no conjunto de tecnologias cujo o escopo é extrair conclusões a partir dos dados coletados, gerando substratos para a tomada de decisões seja por usuários de rede ou por tecnologias que também possuem o escopo de tomar decisões em simulações de poderes decisórios naturalmente humanos, tais como o *machine learning*¹⁴.

O *machine learning*, por exemplo, viabiliza ao seu usuário programar um computador de forma que este tome decisões baseadas num conjunto de dados, ultrapassando a capacidade de raciocínio e velocidade de pensamento humano, podendo gerar, pelo menos em hipótese, condições avessas à concorrência ideal, que pode se fundar em qual usuário de rede dispõe de maior acesso à tecnologia.

A meta do *big data analytics*, por sua vez, consiste em desenvolver métodos confiáveis e eficientes para o futuro, viabilizando melhores processos de tomada de decisão, e ao mesmo tempo aprimorar os mecanismos de resposta e interação entre as fases da análise e interpretação de dados¹⁵.

¹³ "Big data are worthless in a vacuum. Its potential value is unlocked only when leveraged to drive decision making". GANDOMI, Amir. HAIDER, Murtaza. Beyond the hype: Big data concepts, methods, and analytics. *International Journal of Information Management*, Volume 35, Issue 2, April 2015, pages 137-144, International Journal of Information Management. "Frequency distribution of documents containing the term 'big data' in ProQuest Research Library", p. 3.

¹⁴ "Machine learning is a programming computers to optimize a performance criterion using data or past experience. We need learning in cases where we cannot directly write a computer program to solve a given problem, but need example data or experience. Once case where learning is necessary is when human expertise does not exist, or when humans are unable to explain their expertise". ALPAYDIN, Ethem. *Introduction to machine learning*. 2nd ed., 2009.

¹⁵ "What are the goals of analyzing Big Data? According to [3], two main goals of high-dimensional data analysis are to develop effective methods that can accurately predict the future observations and

Neste prisma, tem-se que o *big data analytics* demanda fases de processamento de dados, etapas que se iniciam a partir da aquisição e armazenamento, para em sequência prosseguir com atos de extração, limpeza e anotação, e subsequente integração, agregação e representação. Essas etapas dizem respeito à fase de gestão de dados, que permite ao usuário de rede organizar as informações coletadas para em fase posterior promover uma estratégia de análise e eventual tomada de decisão.

A fase de análise, por conseguinte, pressupõe as etapas de modelagem e análise, com posterior interpretação, momento no qual a tomada de decisão, seja por um usuário humano ou por um usuário mecânico, se mostrará pertinente. Pode-se reproduzir as fases de *big data analytics* da seguinte forma:

Data management

Aquisição e armazenamento → extração,
limpeza e anotação → integração, agregação e
representação

Análise

Modelagem e análise → interpretação¹⁶

Logo, tem-se que todo ato voltado à gestão e processamento de dados perpassa pelas etapas acima reproduzidas, com uma fase inicial, voltada à captação e gerenciamento de dados, e uma posterior, cujo objetivo é interpretar e viabilizar a tomada de decisão, verdadeiro poder inerente ao *big data*, seja em âmbito de análise ou sob sua forma genérica já explorada.

O presente artigo concebe, neste âmbito, que as tecnologias voltadas ao processamento de dados inerentes às etapas acima reproduzidas é que podem ensejar uma intervenção regulatória concorrencial. Isso porque as fases de aquisição e armazenamento, bem como a extração, limpeza e anotação podem gozar de tecnologias diferenciadas que venham a gerar diferenças competitivas.

Ora, mesmo em abstrato, não é inconcebível cogitar que a velocidade e qualidade de extração de limpeza e anotação, por exemplo, de dados adquiridos

em determinado setor de mercado aos níveis de excelência podem gerar vantagem ilícita entre mesmos agentes de mercado, dado que um agente pode se encontrar defasado na capacidade de tomada de escolhas, verdadeiro poder que o *big data* gera aos seus usuários.

A partir da explicitação das fases de gestão e análise da informação, alcança-se um cenário mais palpável à regulação concorrencial, que já partiria do pressuposto de que a informação nua e crua não pode se sujeitar a uma intervenção, mas uma tecnologia vislumbrada em alguma das etapas reproduzidas poderia facilmente ensejar uma conduta anticompetitiva e sujeita a uma regulação.

A fim de facilitar, nesta senda, a intervenção regulatória, e analisar os campos de exploração do *big data analytics*, cumpre explicitar em que contextos tal processo de análise se aplica a serviços e áreas distintas entre si, até mesmo porque os objetivos e formas de processamento variam da área da tecnologia para a área das redes sociais, por exemplo.

Por conseguinte, pode-se registrar diferentes tecnologias de captação e processamento de dados e informações em âmbito de redes sociais, em medidas preventivas de forma genérica, mediante análise de vídeos, áudios e de textos. O chamado *text analytics* se refere à extração e processamento de dados a partir de dados textuais, incluindo nesse leque as redes sociais, *e-mails*, *blogs*, pesquisas, fóruns, documentos, livros e todo tipo de informação constante de veículo textual¹⁷.

As tecnologias aplicadas à análise e processamento de textos partem do *information extraction*, que angaria informações desestruturadas ao longo de um texto presente em rede, enquanto que o *text summarization* gera uma síntese de uma unidade documental ou de um conjunto de textos, aumentando a eficiência na absorção de informações e, conseqüentemente, o processo de tomada de decisões por parte dos usuários de rede¹⁸.

at the same time to gain insight into the relationship between the features and response for scientific purposes". FAN, J., HAN, F.; LIU, H. (2014). Challenges of big data analysis. *National Science Review*, 1(2), 293–314.F.

¹⁶ LABRIDINIS, A.; JAGADISH, H. V. (2012). Challenges and opportunities with big data. *Proceedings of the VLDB Endowment*, 5(12), 2032–2033. Quadro que reproduz as fases do big data analytics de forma adaptada.

¹⁷ GANDOMI, Amir; HAIDER, Murtaza. Beyond the hype: big data concepts, methods, and analytics. *International Journal of Information Management*, Volume 35, Issue 2, April 2015, pages 137-144, International Journal of Information Management. "Frequency distribution of documents containing the term 'big data' in ProQuest Research Library", p. 4.

¹⁸ GANDOMI, Amir; HAIDER, Murtaza. Beyond the hype: big data concepts, methods, and analytics. *International Journal of Information Management*, Volume 35, Issue 2, April 2015, pages 137-144, International Journal of Information Management.

Pode-se mencionar a tecnologia *question answering (QA)*, aparato tecnológico que responde perguntas de usuários assim como o proprietário de iPhone obtém respostas de seu celular, ou de qualquer sistema análogo¹⁹.

O *sentiment analysis (opinion mining)* já consiste em técnica cujo escopo é a coleta de opiniões de usuários quanto aos serviços prestados pela plataforma utilizada, estruturando uma rede de dados que permite aos empreendedores alcançar uma excelência em seus serviços e produtos²⁰.

Embora as tecnologias apresentadas não aparentem representar uma afronta a fundamentos concorrenciais, é preciso destacar que o *text analytics* se mostra capaz até de processar e sumarizar notícias de porte econômico e financeiro, podendo gerar severa repercussão num grupo de agentes que atuam em mesmo mercado relevante²¹.

Por conseguinte, é importante ressaltar que o aprofundamento nas tecnologias voltadas ao processamento de dados permite aos órgãos reguladores, em âmbito concorrencial, delimitar pontos de intervenção mais objetivos, de forma a frisar uma metodologia ou meio (no caso o *text analytics*), e o objetivo e forma de funcionamento da tecnologia empregada em meio ao *big data analytics*.

É preciso frisar que, de fato, muitas das tecnologias também não representam por si só fundamentos plausíveis para uma interferência regulatória, como pode se ver com uma tecnologia do tipo *question answering*, muito embora a delimitação mais objetiva da tecnologia e do seu escopo auxilie os órgãos regulatórios a observarem condutas anticompetitivas.

A partir do posicionamento ora apresentado, *data venia*, pode-se ao menos considerar que a forma

de processamento é escrita e, em ato subsequente, delimitar o alcance

Noutro giro, pode-se vislumbrar que o avanço das tecnologias e a automatização dos procedimentos pode gerar, por si só, um disparate enorme entre agentes de mercado, já que a aplicação de procedimentos de gestão e processamento de dados amplamente operacionalizada pode repercutir de forma incalculável na competição entre dois agentes de mercado, no qual um desses *players* se vê em notável vantagem ao se mostrar capaz de tomar decisões e implementar práticas comerciais de forma muito mais eficiente e célere que seus concorrentes.

Prossiga-se com a áreas de incidência da função analítica do *big data*.

A análise de dados pode se dar também em meio a dados de áudio (*audio analytics*), sendo possível cogitar o processamento de informações localizados em chamadas gravadas, materiais de áudio de aplicativos sociais e atendimentos eventualmente gravados, providenciando *feedbacks* instantâneos possivelmente realizados em ligações cruzadas²².

O *video analytics* é estruturado de forma a captar e monitorar as informações trazidas em vídeos ao longo de sua transmissão, sendo capaz de, a partir de tais dados, otimizar a tomada de decisão e formadores de opiniões em meio aos materiais audiovisuais²³.

"Frequency distribution of documents containing the term 'big data' in ProQuest Research Library", p. 4.

¹⁹ GANDOMI, Amir; HAIDER, Murtaza. Beyond the hype: big data concepts, methods, and analytics. *International Journal of Information Management*, Volume 35, Issue 2, April 2015, pages 137-144, International Journal of Information Management. "Frequency distribution of documents containing the term 'big data' in ProQuest Research Library", p. 4.

²⁰ LIU, B. (2012). Sentiment analysis and opinion mining. *Synthesis lectures on human language technologies*, 5(1), 1–167.

²¹ CHUNG, W. (2014). BizPro: extracting and categorizing business intelligence factors from textual news articles. *International Journal of Information Management*, 34(2), 272–284.

²² GANDOMI, Amir; HAIDER, Murtaza. Beyond the hype: big data concepts, methods, and analytics. *International Journal of Information Management*, Volume 35, Issue 2, April 2015, pages 137-144, International Journal of Information Management. "Frequency distribution of documents containing the term 'big data' in ProQuest Research Library", p. 4.

²³ "In terms of the system architecture, there exist two approaches to video analytics, namely server-based and edge-based: Server-based architecture. In this configuration, the video captured through each camera is routed back to a centralized and dedicated server that performs the video analytics. Due to bandwidth limits, the video generated by the source is usually compressed by reducing the frame rates and/or the image resolution. The resulting loss of information can affect the accuracy of the analysis. However, the server-based approach provides economies of scale and facilitates easier maintenance. Edge-based architecture. In this approach, analytics are applied at the 'edge' of the system. That is, the video analytics is performed locally and on the raw data captured by the camera. As a result, the entire content of the video stream is available for the analysis, enabling a more effective content analysis. Edge-based systems, however, are more costly to maintain and have a lower processing power compared to the server-based systems". Trecho que explicita as formas de processamento dos dados audiovisuais. GANDOMI, Amir; HAIDER, Murtaza. Beyond the hype: big data concepts, methods, and analytics. *International Journal of Information Management*, Volume 35, Issue 2, April 2015, pages 137-144, *International Journal of Information Management*.

A promoção de análise de dados em meio às redes sociais obedece, como dito, pressupostos e dinâmicas próprias. O *social media analytics* que busca focalizar o processamento de dados em meio às redes sociais nas quais os usuários constituem perfis e trocam materiais e dados entre si, participando coletivamente de comunidades e intermediando encontros, reuniões ou eventos sociais.

Trata-se de plataforma que, mais do que outras áreas, atravessa informações e dados relativos à sociologia, psicologia, economia e demais campos de seus usuários, tendo em vista a variedade de objetivos e abrangência de atuação dos usuários de tais programas. As redes sociais e as tecnologias inerentes ao *big data analytics* partem da premissa de análise de conteúdo (*content-based analytics*) e de estrutura (*structure-based analytics*)²⁴, estabelecendo pontos de focos distintos de forma a abranger de forma satisfatória a magnitude e diversidade das redes sociais e da interação entre os seus membros²⁵.

Nessa análise das redes sociais, também consideram-se as tecnologias voltadas à *community detection*, que delinea as comunidades internas a uma rede social, bem como o *social influence analysis*, que averigua a relevância individual dos usuários de determinada rede social²⁶.

Por fim, essencial distinguir as análises necessariamente *preditivas* (*predictive analysis*), cujo escopo exclusivo é prever tendências futuras a partir de registro de dados e de informações correntes.

Trata-se de viés inerente ao *big data*, que estabelece formas de tomada de escolhas mediante a análise de padrões de condutas e fenômenos, orientando-se ao ponto de até mesmo substituir as decisões humanas por decisões de mérito providas

de máquinas, contexto no qual nasce o *machine learning*²⁷.

É neste aspecto que a distinção traçada entre *big data* e *big data analytics* toma força, pois focando-se no processamento e gestão de dados voltados à predição de tendências futuras, pode-se cogitar a intervenção de órgãos regulatórios no âmbito do direito concorrencial, ainda que de forma independente do próprio conteúdo dessa informação manipulada e processada.

Uma vez alcançado tal patamar, cumpre finalizar expondo os argumentos regulatórios típicos para o *big data analytics*, também suscitado por parte da doutrina pertinente e ora reproduzido.

4 Interseção entre o direito concorrencial e as tecnologias voltadas ao processamento de dados

Como já referendado, firma-se o posicionamento neste estudo que a distinção entre o *big data* como instituto genérico e o *big data analytics* permite que seja vislumbrado um horizonte mais específico de intervenção regulatória, contanto que o objeto de ação seja o processo relativo à análise de dados e à tomada de decisões por parte de agentes de mercado.

Com isso, resta possível delinear que argumentos suscitados pela doutrina se mostram aplicáveis, mesmo que em tese, ao direito concorrencial.

Primeiramente, o argumento de que as estratégias fundamentadas em escolhas sustentadas em processamento de dados gerariam vantagem competitiva²⁸ é um ponto de fato pertinente, que justificaria a intervenção regulatória. Como se viu, as tecnologias inerentes ao *big data analytics* possuem potencial exponencial, e podem efetivamente

"Frequency distribution of documents containing the term 'big data' in ProQuest Research Library", p. 4.

²⁴ BARBIER, G., LIU, H. (2011). Data mining in social media. In C. C. Aggarwal (Ed.), *Social network data analytics* (p. 327–352). United States: Springer, p. 2.

²⁵ AGGARWAL, C. C. (2011). An introduction to social network data analytics. In C. C. Aggarwal (Ed.), *Social network data analytics* (p. 1–15). United States: Springer.

²⁶ GANDOMI, Amir; HAIDER, Murtaza. Beyond the hype: big data concepts, methods, and analytics. *International Journal of Information Management*, Volume 35, Issue 2, April 2015, pages 137-144, International Journal of Information Management. "Frequency distribution of documents containing the term 'big data' in ProQuest Research Library", p. 4.

²⁷ "Based on the underlying methodology, techniques can also be categorized into two groups: regression techniques (e.g., multinomial logit models) and machine learning techniques (e.g., neural networks). Another classification is based on the type of outcome variables: techniques such as linear regression address continuous outcome variables (e.g., sale price of houses), while others such as Random Forests are applied to discrete outcome variables (e.g., credit status)". GANDOMI, Amir. HAIDER, Murtaza. Beyond the hype: big data concepts, methods, and analytics. *International Journal of Information Management*, Volume 35, Issue 2, April 2015, pages 137-144, International Journal of Information Management. "Frequency distribution of documents containing the term 'big data' in ProQuest Research".

²⁸ GRUNES, Allen P.; STUCKE, Maurice E. No mistake about it: the important role of antitrust in the era of big data (April 28, 2015). Antitrust Source (Apr. 2015), Online; *University of Tennessee Legal Studies Research Paper* n. 269, p. 2.

servirem de base para a constituição de uma vantagem competitiva entre agentes equiparados em determinado mercado.

Vale destacar que a maior objetividade do ponto de intervenção de regulação por partes dos órgãos concorrenciais permite afastar o argumento de que a informação em si é inofensiva, já que o ponto fulcral a partir da premissa ora estabelecida é a tomada de decisão a partir de um processo tecnológico de processamento de dados e tomada de decisões.

Desta perspectiva, as operações de fusões e aquisições de agentes de mercado estruturados numa prática de aquisição e processamento de dados pode ser analisada previamente por órgãos concorrenciais, dado que o pareamento de determinadas tecnologias pode gerar vantagem anticompetitivas, conforme suas práticas aprimorem a prestação de serviço em determinado campo ao ponto de inviabilizar o ingresso de outros concorrentes²⁹.

O argumento de que os agentes de mercado podem vir a obstar o acesso dos demais concorrentes também é plausível, já que a informação pode ser disposta em rede em diferentes graus de organização³⁰. A informação crua poderá ser encontrada de forma desestruturada, mas a informação processada e organizada, tendo em vista a aplicação de tecnologia exclusivamente voltada a essa finalidade, poderia ser efetivamente obstada por parte de agentes com maior aparato tecnológico, seja em qualquer uma das áreas já tratadas neste estudo.

²⁹ "Second, companies undertake data-driven strategies to obtain and sustain competitive advantages. Companies increasingly strive to gain a 'big data advantage' over their rivals. One MIT-led study showed that the more companies characterized themselves as data-driven, the better they performed on objective measures of financial and operational results. "[C]ompanies in the top third of their industry in the use of data-driven decision making were, on average, 5% more productive and 6% more profitable than their competitors." GRUNES, Allen P.; STUCKE, Maurice E. No mistake about it: the important role of antitrust in the era of big data (April 28, 2015). Antitrust Source (Apr. 2015), Online; University of Tennessee Legal Studies Research Paper n. 269, p. 3.

³⁰ "Fourth, when data-driven businesses incur significant costs to obtain, store, and analyze data (as well as provide 'free' services to collect data), they may have strong incentives to limit their competitors' access to these datasets, prevent others from sharing the datasets, and could likely be averse to data-portability policies that threaten their data-related competitive advantage." GRUNES, Allen P.; STUCKE, Maurice E. No mistake about it: the important role of antitrust in the era of big data (April 28, 2015). Antitrust Source (Apr. 2015), Online; University of Tennessee Legal Studies Research Paper n. 269, p. 4.

Mais uma vez, o *big data analytics* demonstra relevância ao se cogitar a intervenção concorrencial, posto que o conhecimento relativo à informação processada adquire significado também diferenciado.

Por fim, o argumento de que o *big data* por si só não justifica a intervenção de uma regulação concorrencial pode obscurecer as verdadeiras intenções de agentes de mercado que queiram se fundir com base em estratégias fundamentadas em dados e informações processadas³¹. Deveras, a precisão teórica e o alcance das tecnologias voltadas ao processamento de dados poderão embasar novas formas interventivas no âmbito do *big data analytics*.

Ora, o processo de *machine learning*, aliado a uma ampla base de dados bem geridos e processados, pode edificar um investimento em proporções sem escala, indubitavelmente demandando uma intervenção regulatória e dando espaço ao direito concorrencial.

Portanto, de uma perspectiva voltada somente ao potencial hipotético das tecnologias de processamento e análise de dados, pode-se sumarizar que o direito concorrencial encontra pontos de intervenção nos seguintes aspectos:

1) Estudos das tecnologias inerentes ao *big data analytics*, de forma a compreender o comportamento de mercado e a destinação das práticas estabelecidas neste espectro.

2) Compreender a acessibilidade de informações estruturadas e a vantagem competitiva que tais recursos proporcionam aos agentes de mercado.

3) Estabelecer a partir de campos de *big data analytics* os mercados relevantes, compreendendo dinâmicas diferenciadas embasadas em tecnologias de processamento de dados.

4) Averiguar até que nível a tomada de decisões por humanos e máquinas geram uma vantagem anticompetitivas em mercado.

³¹ "Sixth, as companies undertake data-driven business strategies, one might expect them to raise data-driven efficiencies as a defense to justify potentially anticompetitive mergers. In closing its investigation of the agreement between Microsoft and Yahoo!, the DOJ found that the transaction would create a more viable competitive alternative to Google because of the importance of scale to competitive performance in search and search advertising, and suggested that the transaction would enable more rapid improvements in Microsoft's search and search advertising technology". GRUNES, Allen P.; STUCKE, Maurice E. No mistake about it: the important role of antitrust in the era of big data (April 28, 2015). Antitrust Source (Apr. 2015), Online; University of Tennessee Legal Studies Research Paper n. 269, p. 4.

5) Considerar o impacto que a junção de determinadas tecnologias pode acarretar no mercado a partir de fusões e aquisições.

5 Conclusão

Buscou-se, ao longo deste estudo, traçar parâmetros minimamente objetivos para que o direito concorrencial pudesse vislumbrar campos de intervenção regulatória. Logo, estabelecendo-se uma distinção entre o *big data*, como conceito abrangente, e o *big data analytics*, acredita-se que um parâmetro minimamente delineado, ainda que em abstrato, restou asseverado.

Elencou-se, também, pontos de intervenção regulatória por parte dos órgãos voltados à defesa da concorrência, ainda que inexista relevantes precedentes aplicáveis.

Finaliza-se a presente exposição com a assertiva que declara que: *“Big data is neither inherently good, evil, nor neutral. Its social value depends on the industry and the purpose and effect of the data-driven strategy”*³². O *big data* em si não é bom nem ruim, o seu valor social depende do propósito e dos efeitos das decisões tomadas a partir de processamento e análise de dados, perspectiva compartilhada a todo tempo pelo presente artigo.

6 Referências

ALPAYDIN, Ethem. *Introduction to machine learning*. 2nd ed., 2009.

BARBIER, G.; LIU, H. (2011). Data mining in social media. In C. C. Aggarwal (Ed.), *Social network data analytics* (p. 327–352). United States: Springer, p. 2.

BIDARRA, Ana Isabel Rodrigues Gomes. *O impacto do Big Data no direito da concorrência: o caso Facebook-WhatsApp e o controlo de concentrações na União Europeia*. Faculdade de Direito da Universidade de Lisboa, junho de 2018, p. 125.

Schroeck, Shockley, Smart, Romero-Morales, & Tufano, 2012.

CHUNG, W. (2014). BizPro: Extracting and categorizing business intelligence factors from textual news articles. *International Journal of Information Management*, 34(2).

SOKOL, D. Daniel; COMEFORD, Roisin. Antitrust and regulating big data. *23 Geo. Mason L. Rev.* 1129 (2016), available at <https://scholarship.law.ufl.edu/facultypub/803/>. Acesso em 10 julho, 2019.

DIEBOLD, F. X. (2012). A personal perspective on the origin(s) and development of “big data”: The phenomenon, the term, and the discipline (Scholarly Paper n. ID 2202843). *Social Science Research Network*. Retrieved from http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2202843, p. 8. Acesso em 27 junho, 2019.

FAN, J., Han, F., & Liu, H. (2014). Challenges of big data analysis. *National Science Review*, 1(2), 293–314.F.

LANCIERI, Filippo Maria Lancieri. “Digital protectionism? Antitrust, data protection, and the EU/US transatlantic rift”. *Journal of Antitrust Enforcement*, 2018.

GANDOMI, Amir. HAIDER, Murtaza. Beyond the hype: Big data concepts, methods, and analytics. *International Journal of Information Management*, Volume 35, Issue 2, April 2015, pages 137-144, *International Journal of Information Management*. “Frequency distribution of documents containing the term ‘big data’ in ProQuest Research Library”, p. 4.

GRUNES, Allen P. and STUCKE, Maurice E., No Mistake About It: The Important Role of Antitrust in the Era of Big Data (April 28, 2015). *Antitrust Source* (Apr. 2015), Online; University of Tennessee Legal Studies Research Paper n. 269. Acesso em 2 jul., 2019.

LABRIDINIS, A; JAGADISH, H. V. (2012). *Challenges and opportunities with big data*. *Proceedings of the VLDB Endowment*, 5(12), 2032–2033.

Lei Geral de Proteção de dados. Disponível em: http://www.planalto.gov.br/ccivil_03/ato2015-2018/2018/lei/L13709.htm. Acesso em 12 jul., 2019.

LIU, B. (2012). Sentiment analysis and opinion mining. *Synthesis Lectures on Human Language Technologies*, 5(1), 1–167.

STUCKE, Maurice E.; GRUNES, Allen. Debunking the myths over big data and antitrust. *CPI Antitrust Chronicle* May 2015 (2).

³² STUCKE, Maurice E., GRUNES, Allen. Debunking the Myths over Big Data and Antitrust. *CPI Antitrust Chronicle* May 2015 (2).