ECONSTOR
Make Your Publications Visible.

A Service of

ZBW Leibniz-Informationszentrum Wirtschaft
Leibniz Information Centre for Economics

Rauen, Andre Tortato; Guth, Frederico; Paiva, Bianca Paiva

Working Paper

Proposta de indicador de tema tecnológico emergente

Texto para Discussão, No. 2577

Provided in Cooperation with:
Institute of Applied Economic Research (ipea), Brasília

Suggested Citation: Rauen, Andre Tortato; Guth, Frederico; Paiva, Bianca Paiva (2020) : Proposta de indicador de tema tecnológico emergente, Texto para Discussão, No. 2577, Instituto de Pesquisa Econômica Aplicada (IPEA), Brasília

This Version is available at:
https://hdl.handle.net/10419/240772

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

WWW.ECONSTOR.EU

Mitglied der Leibniz-Gemeinschaft

TEXTO PARA DISCUSSÃO

2577

PROPOSTA DE INDICADOR DE TEMA TECNOLÓGICO EMERGENTE

André Rauen
Frederico Guth
Bianca Paiva

ipea

2577
TEXTO PARA DISCUSSÃO
Brasília, julho de 2020

PROPOSTA DE INDICADOR DE TEMA TECNOLÓGICO EMERGENTE[1]

André Rauen[2]
Frederico Guth[3]
Bianca Paiva[4]

1. Este projeto contou, em parte, com o financiamento daComissão Econômica para a América Latina e o Caribe (Cepal) em acordo de cooperação com o Ipea. Adicionalmente, os autores agradecem os valorosos comentários de Pedro Miranda e Graziela Zucoloto. Evidentemente, os eximindo de qualquer responsabilidade por eventuais erros ou omissões.
2. Diretor da Diretoria de Estudos e Políticas Setoriais de Inovação e Infraestrutura (Diset) do Ipea. *E-mail*: <andre.rauen@ipea.gov.br>.
3. Consultor da Cepal. *E-mail*: <fredguth@fredguth.com>.
4. Pesquisadora do Subprograma de Pesquisa para o Desenvolvimento Nacional (PNPD) na Diset/Ipea. *E-mail*: <bianca.paiva@ipea.gov.br>.

Governo Federal

Ministério da Economia
Ministro Paulo Guedes

ipea **Instituto de Pesquisa Econômica Aplicada**

Fundação pública vinculada ao Ministério da Economia, o Ipea fornece suporte técnico e institucional às ações governamentais – possibilitando a formulação de inúmeras políticas públicas e programas de desenvolvimento brasileiros – e disponibiliza, para a sociedade, pesquisas e estudos realizados por seus técnicos.

Presidente
Carlos von Doellinger

Diretor de Desenvolvimento Institucional
Manoel Rodrigues Junior

Diretora de Estudos e Políticas do Estado, das Instituições e da Democracia
Flávia de Holanda Schmidt

Diretor de Estudos e Políticas Macroeconômicas
José Ronaldo de Castro Souza Júnior

Diretor de Estudos e Políticas Regionais, Urbanas e Ambientais
Nilo Luiz Saccaro Júnior

Diretor de Estudos e Políticas Setoriais de Inovação e Infraestrutura
André Tortato Rauen

Diretora de Estudos e Políticas Sociais
Lenita Maria Turchi

Diretor de Estudos e Relações Econômicas e Políticas Internacionais
Ivan Tiago Machado Oliveira

Assessora-chefe de Imprensa e Comunicação
Mylena Fiori

Ouvidoria: http://www.ipea.gov.br/ouvidoria
URL: http://www.ipea.gov.br

Texto para Discussão

Publicação seriada que divulga resultados de estudos e pesquisas em desenvolvimento pelo Ipea com o objetivo de fomentar o debate e oferecer subsídios à formulação e avaliação de políticas públicas.

© Instituto de Pesquisa Econômica Aplicada – **ipea** 2020

Texto para discussão / Instituto de Pesquisa Econômica Aplicada.- Brasília : Rio de Janeiro : Ipea , 1990-

ISSN 1415-4765

1.Brasil. 2.Aspectos Econômicos. 3.Aspectos Sociais. I. Instituto de Pesquisa Econômica Aplicada.

CDD 330.908

As publicações do Ipea estão disponíveis para *download* gratuito nos formatos PDF (todas) e EPUB (livros e periódicos). Acesse: http://www.ipea.gov.br/portal/publicacoes

As opiniões emitidas nesta publicação são de exclusiva e inteira responsabilidade dos autores, não exprimindo, necessariamente, o ponto de vista do Instituto de Pesquisa Econômica Aplicada ou do Ministério da Economia.

É permitida a reprodução deste texto e dos dados nele contidos, desde que citada a fonte. Reproduções para fins comerciais são proibidas.

JEL: O30; O31; O32.

SUMÁRIO

SINOPSE

ABSTRACT

1 INTRODUÇÃO 7

2 NOVAS TECNOLOGIAS NA "SEGUNDA ERA DAS MÁQUINAS" 8

3 METODOLOGIA 13

4 RESULTADOS 27

5 VALIDAÇÃO 30

6 DISCUSSÃO 32

7 CONSIDERAÇÕES FINAIS 36

REFERÊNCIAS 37

SINOPSE

Este texto para discussão tem por objetivo propor e discutir um indicador de temas tecnológicos emergentes, isto é, novas discussões específicas sobre tecnologias, aplicações tecnológicas e suas implicações.[1] A construção de tal indicador se justifica em razão da necessidade de pautar o governo federal brasileiro nas discussões emergentes no campo da política tecnológica. Em realidade, a vasta gama de informações produzidas a cada momento torna a atividade de monitoramento de temas emergentes extremamente desafiadora. Este texto procura, então, separar temas tecnológicos do ruído que envolve qualquer discussão de política pública no início do século XXI.

Palavras-chave: monitoramento tecnológico; pesquisa e desenvolvimento; indicador.

ABSTRACT

This discussion paper aims to propose an indicator of emerging technological themes. The development of such indicator is justified due to the need to guide the Brazilian federal government in emerging discussions on the field of technological policy. In reality, the wide range of information produced every second makes the activity of monitoring emerging issues extremely challenging. This text, then, seeks to separate technological themes from the noise that involves any public policy discussion at the beginning of the 21st century.

Keywords: technological monitoring; research and development; indicator.

1. Por "emergente" entendemos temas cujo volume de discussão foge ao normal e esperado dado o comportamento no semestre anterior a sua identificação. Procuramos anomalias.

1 INTRODUÇÃO

Este texto para discussão tem por objetivo propor e discutir um indicador de temas tecnológicos emergentes, isto é, novas discussões específicas sobre tecnologias, aplicações tecnológicas e suas implicações.[1]

A construção de tal indicador se justifica em razão da necessidade de pautar o governo federal brasileiro nas discussões emergentes no campo da política tecnológica. Em realidade, a vasta gama de informações produzidas a cada segundo torna a atividade de monitoramento de temas emergentes extremamente desafiadora. Este texto procura, então, separar temas tecnológicos do ruído que envolve qualquer discussão de política pública no início do século XXI.

Não faltam, contudo, instrumentos privados de sinalização e monitoramento de temas emergentes, já existindo um grande número de consultorias internacionais dedicadas ao assunto. De fato, no mundo corporativo, o acompanhamento de tendências tecnológicas não é novo e nem restrito.[2] O fato é que o governo brasileiro tem dependido dessas análises, mesmo tendo capacitação interna para produzi-las. Este texto pretende, então, diminuir essa dependência. Evidentemente, o objetivo final não é o de substituir o emprego de outras ferramentas, mas adicionar mais uma, agora, produzida internamente.

Já nesta introdução é importante destacar uma distinção fundamental: a preocupação central deste texto é de encontrar temas tecnológicos emergentes, e não de definir *tecnologias* propriamente ditas. Isso significa que o foco da análise está na discussão, e não na previsão tecnológica.

Para tanto, parte-se da discussão sobre a emergência de um novo paradigma técnico-econômico que, de tão intenso e pervasivo, tem sido creditado como origem de uma nova era, aos moldes do que representou a Revolução Industrial. Nessa discussão se insere a questão do desenvolvimento e gradual difusão de novas tecnologias que permitirão, cada vez mais, a interconexão autônoma de pessoas, objetos, produtos e serviços.

1. Por "emergente" entendemos temas cujo volume de discussão foge ao normal e esperado, dado o comportamento no semestre anterior à sua identificação. Procuramos anomalias.
2. No campo das consultorias privadas internacionais, a ferramenta de acompanhamento tecnológico mais conhecida e empregada diz respeito ao Hype Cycle da Gartner. Disponível em: <https://gtnr.it/2I4K2Ru>.

Esse conjunto de novas tecnologias tem produzido e também consumido um volume de dados sem precedentes. Coloca-se, portanto, um novo desafio nas políticas públicas baseadas em evidências, qual seja, como selecionar as evidências mais relevantes num conjunto crescente de dados. É sobre essa discussão que se desdobra a primeira seção, na qual sustentamos que é característico da contemporaneidade a necessidade de tratar o ruído.

Na segunda seção, com base em referências selecionadas e num esforço exploratório, apresenta-se a metodologia de elaboração de um Indicador de Temas Tecnológicos Emergentes (ITTE) próprio e baseado na rede social Twitter, mais especificamente, em tuítes de contas selecionadas.

Na sequência, a terceira seção apresenta os resultados do indicador, para, na quarta seção, testar-se a validade deles. Finalmente, a quinta seção promove uma discussão dos resultados e procura qualificá-los.

2 NOVAS TECNOLOGIAS NA "SEGUNDA ERA DAS MÁQUINAS"

A literatura sobre o possível surgimento de um novo paradigma técnico-econômico e até de uma nova era aos moldes do que aconteceu com a Revolução Industrial é vasta e extremamente heterógena em termos de escopo e qualidade. Tem-se desde trabalhos no campo da gestão de empresas a trabalhos de cunho mais sociológico, bem como uma mistura de filosofia com autoajuda: "*All lis flux. Nothing is fineshed. Nothing is done. This never-ending change is the pivotal axis of the modern world*" (Kelly, 2017).

Do ponto de vista de uma abordagem mais econômica, destacam-se – em razão da robustez e da qualidade metodológica – os estudos de Brynjolfsson e McAfee (2014; 2017). Esta literatura tem definido esse momento histórico particular como "líquido", "interconectado", "digital", "automático", "tecnológico", "novo", "baseado em dados" e tantas outras qualificações. É evidente, agora, a necessidade de incorporar a característica "ruidosa" a essas qualificações. Ou seja, existe muita informação e a maior parte dela é ruído. Em nenhum momento da história produzimos tanta informação e em tamanha velocidade. Num único dia de 2019, são enviados 500 milhões de tuítes e 65 bilhões de mensagens no WhatsApp.[3] A cada 60 segundos, quase US$ 1milhão são gastos em compras *on-line* e 3,8 milhões de pesquisas são feitas no Google Search (tabela 1).

3. Para mais informações, ver: <https://bit.ly/38t9V8q>.

TABELA 1
Criação de dados por minuto (2019)

Rede social	Indicador	Frequência
Facebook	Loggin	1 milhão
Youtube	Vídeos visualizados	4,5 milhões
Instagram	Rolamentos	347.222
Twitter	Pessoas tuitando	87.500
Emails	Enviados	188 milhões
Facebook, Messenger e WhatsApp	Enviados	41,6 milhões
Google	Pesquisa	3,8 milhões

Fonte: Desjardins (2019).

A todo esse vasto número de informações diariamente criadas e compartilhadas em mídias sociais somam-se informações coletadas automaticamente nos diferentes equipamentos conectados à internet, sejam eles pessoais (através de *wearables*, por exemplo) ou independentes (cada carro autônomo é capaz de criar, por dia, 4 *tera-bytes* de dados).[4]

Considerando-se que "*for the first time in human history, a near-majority of the world's adults are now connected with each other digitally and with a large chunk of the world's accumulated knowledge*" (Brynjolfsson e McAfee, 2017), é de se esperar o vasto volume de informações produzidas atualmente e ilustradas na tabela 1.

De fato, lidar com esse volume inédito de informações é um desafio próprio da modernidade.

> *We are morphing so fast that our ability to invent new things outpaces the rate we can civilize them. These days it takes us a decade after a technology appears to develop a social consensus on what it means and what etiquette we need to tame it. In another five years we'll find a polite place for twittering, just as we figured out what to do with cell phones ringing everywhere* (Kelly, 2016, p. 3).

Do ponto de vista das estratégias microeconômicas, mesmo que Schumpeter (1982) tenha nos alertado para evitar a armadilha da baixa eficiência de tecnologias da

4. Para mais informações, ver: <https://bit.ly/38t9V8q>.

moda, já se observa o surgimento de expressões como *data gluttony*[5] e *data infobesity*,[6] que se referem ao uso inadequado e excessivo de dados pelas empresas:

> *Why does infobesity compromise performance? It's mostly because we human beings can process only so much data. An uncontrollable flood of it overwhelms us, and we feel stressed. Our systems shut down, and our capacity to absorb additional information actually decreases. To cope with the flood, our brains develop tricks and habits. We rely more on information that is closer to home than on information from a distant source. We remember data presented in one format and forget data presented in another. We even squirrel away bits of information for our own private use. (Maybe that's why local offices know things that the corporate center doesn't, and vice-versa.) All such reactions hamper our ability to make decisions based on the best available evidence, and performance suffers accordingly* (Rogers, Puryear e Root, 2013).

Portanto, torna-se cada vez mais evidente o desafio de lidar com a vastidão de dados disponíveis e constantemente produzidos. Ademais, esse não parece ser um desafio apenas das modernas empresas, mas também de indivíduos e governos. De fato, as estratégias de governo digital já deixaram de ser promessas para se tornarem realidades concretas, mesmo em países periféricos. O caso estoniano é um exemplo paradigmático dessa discussão (Björklund, 2016).

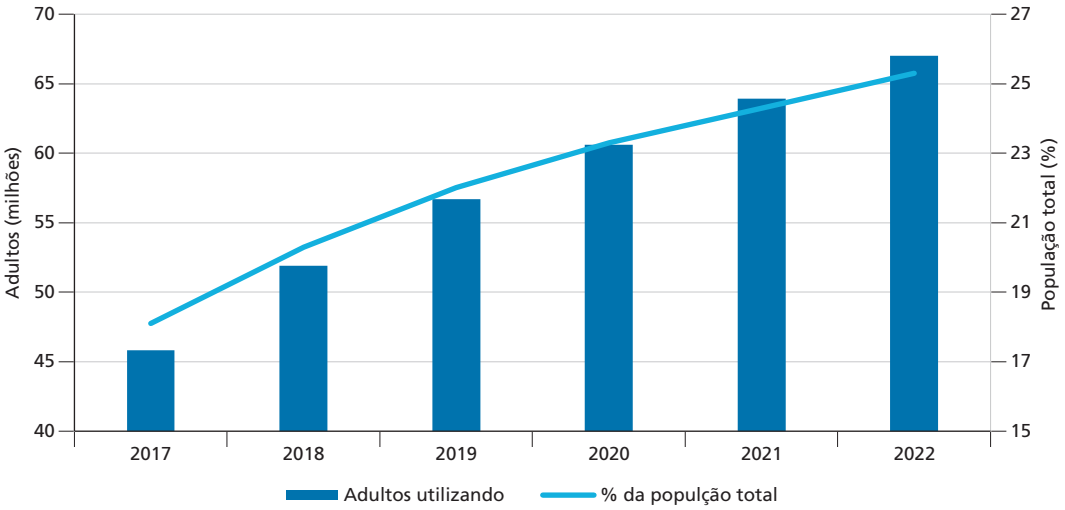
É curioso perceber que, da vasta quantidade de dados produzidos a cada minuto, boa parte é feita sem que tenhamos muita consciência a respeito. De fato, as atividades humanas mais corriqueiras, como caminhar, ou mesmo as mais essenciais, como o bater do coração, agora produzem dados em tempo real, de forma autônoma e completamente interconectada. Os *wearables* fazem justamente isso:

> *Wearables are miniaturized computer and sensor devices, which are worn effortlessly on or in a further development in the body of the wearer. They provide information and interaction with information anytime and anywhere, which is often described as always on and always accessible* (Bauer *et al.*, 2016: 531).

Segundo Wurmser (2019), em 2020, mais de um quarto da população norte-americana estará utilizando algum *wearable*, representando 67 milhões de adultos. Ou seja, essa fonte de dados ainda crescerá no futuro próximo (gráfico 1).

5. Para mais informações, ver:<https://bit.ly/3cg2BjC>.
6. Para mais informações, ver: <https://bit.ly/396MWBe>.

GRÁFICO 1
Previsão de penetração dos *wearables* nos Estados Unidos (2017-2022)



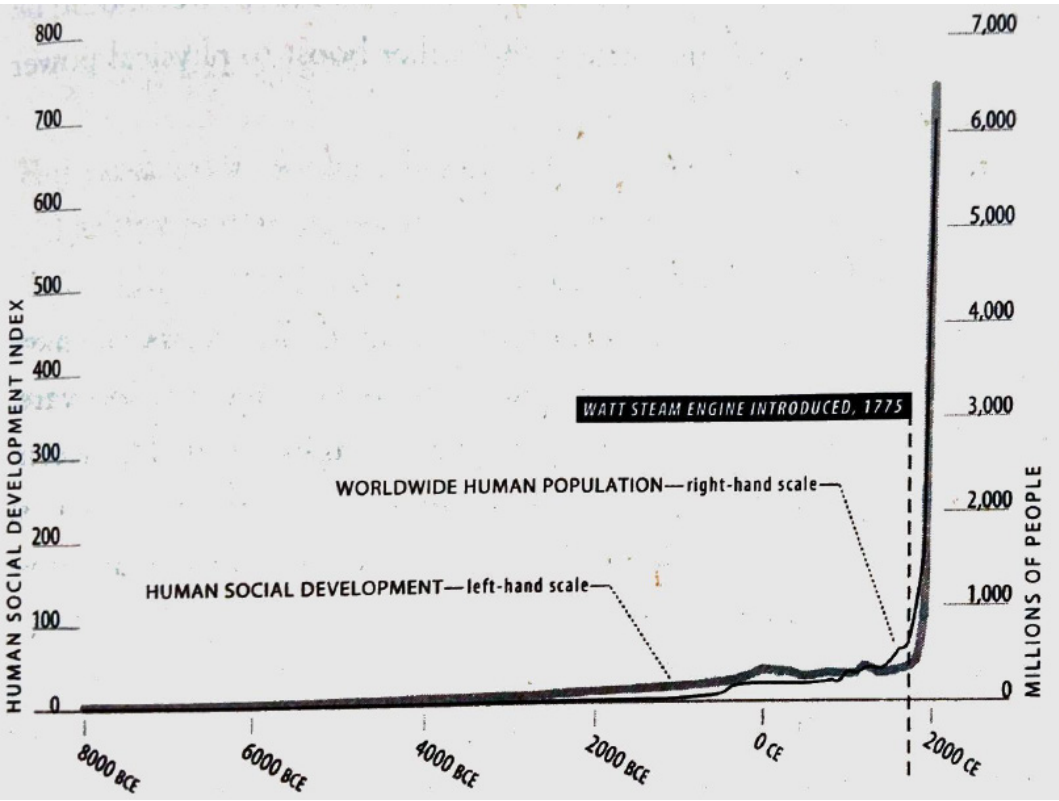
Fonte: Wurmser (2019).
Elaboração dos autores.

Evidentemente, essa produção – quase inconsciente – de dados tem exigido da sociedade que se discuta a segurança e a forma de apropriação destes. Cria-se, assim, uma nova forma de poder cuja realização não é necessariamente virtuosa (O'Neil, 2016).

Toda essa vastidão de dados e a forma autônoma como são coletados e empregados criam um potencial econômico jamais visto. Isto é, ao se equacionar a questão de como dar vazão às informações que hoje possuímos, estaríamos em condição de dar um salto na história tão relevante quanto foi a primeira Revolução Industrial. Estaríamos prestes a entrar numa "segunda era das máquinas" (Brynjolfsson e McAfee, 2014) e, quem sabe, manter a curva da história "dobrada" indefinidamente.

Em outras palavras, esse mesmo volume de dados que cria um elevado "ruído" é também o insumo principal do paradigma que se avizinha. Portanto, o esforço de definição do que é emergente no cenário tecnológico é uma dificuldade própria do atual momento histórico e só pode ser entendida à luz deste momento particular no qual se está prestes a "dobrar" a curva da história novamente.

GRÁFICO 2
O que dobrou a curva da história humana? A Revolução Industrial



Fonte: Brynjolfsson e McAfee (2014).
Obs.: Gráfico reproduzida em baixa resolução e cujos leiaute e textos não puderam ser padronizados e revisados em virtude das condições técnicas dos originais (nota do Editorial).

Comparar a atual onda de inovações com a Revolução Industrial iniciada em fins do século XVIII é evidentemente uma atitude extremamente impactante, uma vez que ela transformou, por completo, a história humana. Portanto, essa nova onda de inovações faria a mesma coisa, isto é, mudar de forma contundente a realidade no planeta.

É interessante perceber que o virtuoso uso de todo esse potencial dá origem, inclusive, a um novo tipo de tecnologia. Allen *et al*. (2020) e Davidson, Filippi e Potts (2018) apresentam o conceito de tecnologias institucionais. Ou seja, tecnologias que diminuem custos de transação não sendo apropriadas por nenhum agente em particular, mas, sim, pelo conjunto do sistema econômico. Ambos os estudos apresentam, como exemplo, as tecnologias de *blockchain,* que diminuem tempo e custo das transações ao mesmo tempo em que garantem verificação suficiente.

Do ponto de vista da economia, Brynjolfsson e McAfee (2017) afirmam que as mudanças em curso promovem uma modificação em direção a máquinas, plataformas e multidões. Em outros termos, pessoas são substituídas por máquinas (inclusive em decisões complexas), produtos dão lugar às plataformas (prover o serviço é mais rentável que vender o produto) e os esforços individuais são substituídos pelas multidões *on-line* (pergunta-se à rede qual seria a solução de determinado problema).

De certa forma, o presente esforço de pesquisa é uma maneira de "ouvir" essa multidão e, a partir de múltiplas opiniões, tentar encontrar um caminho em meio ao elevado ruído.

O problema todo é que só entraremos nessa possível nova era se soubermos lidar com essa vastidão inédita de dados. É isso que tentamos fazer aqui e, ao fazê-lo, pretendemos ajudar a dar melhor uso às informações diariamente consumidas pelo Estado brasileiro.

Portanto, questiona-se: num paradigma de política pública baseada em evidências, como fornecer à política tecnológica – em todo o seu ciclo[7] – informações de qualidade sobre temas emergentes? Assim, não estamos preocupados em encontrar novas tecnologias e nem em definir qual é a percepção dos agentes sobre elas, mas, sim, filtrar as discussões que importam.

3 METODOLOGIA

A intenção final deste esforço de pesquisa é de criar uma ferramenta reproduzível, capaz de alimentar a discussão da política pública brasileira, em especial a tecnológica e a de inovação, sobre temas emergentes. Em outras palavras, o objetivo é o de criar um indicador de discussões emergentes e não de afirmar quais tecnologias nascentes serão relevantes no futuro próximo.

Este texto não faz nenhum exercício de previsão. Nosso intuito é o de identificar novas discussões no campo da tecnologia. Assim, não existe nenhuma garantia de que as tecnologias envolvidas nas discussões aqui realizadas serão, de fato, transformadoras

7. Identificação do problema, inclusão na agenda, planejamento, implementação, monitoramento e avaliação.

da realidade econômica. Aqui, estamos interessados em retirar do enorme ruído que é a internet aqueles temas realmente novos. Separar o que ainda é ficção científica das tecnologias que já estão próximas de transformar o mercado é uma tarefa posterior, que dependerá do sucesso do atual esforço de pesquisa.

Além disso, este texto não tem por objetivo apresentar uma métrica de temas classificados como *trending* ou *hype*. A questão central aqui é, tão somente, identificar temas que, de forma não esperada, surgiram no debate tecnológico.

Dada a velocidade da mudança técnica, não seria eficiente criar tal ferramenta a partir de bases tradicionais de informação tecnológica, a saber, artigos científicos e patentes. Por definição, tais bases trariam resultados fundamentado sem temas já discutidos em algum momento anterior. Por isso, optamos por utilizar as discussões feitas nas redes sociais.

É verdade, contudo, que tais discussões são superficiais, do ponto de vista técnico, e efêmeras, mas não aleatórias. Elas estão associadas a um determinado contexto técnico-econômico que queremos acompanhar. Novamente, nosso objetivo específico não é o de definir tecnologias emergentes que, necessariamente, irão transformar a realidade no futuro, tampouco apresentar uma lista de tecnologias "quentes" (*hype*), mas, sim, evidenciar discussões tecnológicas emergentes que nos permitam assessorar o governo federal, evitando que nos percamos no ruído de informações atualmente disponível.

Em um mundo com inéditos níveis de produção de informação, é preciso ter uma ferramenta que permita observar o que, de fato, é emergente e o que deveríamos estar discutindo no Brasil.

O ciclo *hype* de tecnologias emergentes da consultoria Gartner é, atualmente, a ferramenta mais empregada nos meios empresariais e de gestão tecnológica para separar a ficção científica do real insumo produtivo. Ela se baseia numa representação gráfica da maturidade tecnológica que relaciona expectativas e tempo (gráfico3).

GRÁFICO 3
Interpretando o ciclo *hype* de tecnologias emergentes

Fonte: Linden e Fenn (2003).
Elaboração dos autores.
Obs.: Gráfico cujos leiaute e textos não puderam ser padronizados e revisados em virtude das condições técnicas dos originais (nota do Editorial).

A referida consultoria considera que, de forma geral, as tecnologias possuem um ciclo de vida parecido, que se inicia com provas de conceito que fomentam casos e pautas na imprensa. Nessa primeira fase, não existe qualquer expectativa de comercialização (gatilho da inovação, no gráfico 3).

A soma de casos de sucesso e a exploração deles pelas mídias, inclusive sociais, elevam consideravelmente as expectativas, mesmo que não haja nenhuma evidência concreta sobre a aplicação comercial da tecnologia (pico das expectativas, na figura 3).

Como as expectativas são elevadas, mas torna-se evidente a necessidade de vencer inúmeros desafios, tanto tecnológicos como econômicos, surge grande desapontamento (vale da desilusão, no gráfico 3).

Justamente no auge do desapontamento, poucas instituições e indivíduos, guiados pela curiosidade, presos pela inércia ou talvez conscientes das potencialidades, percebem aplicações potenciais próximas. Novas gerações de produtos e serviços surgem.

Cria-se uma nova expectativa, agora muito mais realista. Contudo, apenas as empresas mais dinâmicas acabam por introduzir comercialmente a tecnologia (ladeira do esclarecimento, no gráfico 3).

Finalmente, com a introdução e difusão massivas da tecnologia, aparecem os primeiros ganhos reais de produtividade e a tecnologia passa a ser paradigma de mercado (platô de produtividade, no gráfico 3). Depois dessa fase, segue-se a obsolescência da tecnologia.

Com base nessa dinâmica geral – altamente criticável, como veremos em seguida –, a consultoria elabora anualmente, mediante entrevistas e análises com especialistas internos, o ciclo *hype* anual. Ou seja, todo ano, por volta do início do segundo semestre, as tecnologias emergentes são classificadas dentro de alguma destas fases.

Diferentemente do que se poderia imaginar, os anos não estão relacionados e tecnologias podem aparecer e desaparecer, ou mesmo podem estar em fases completamente distintas, dependendo do ano. O fato é que essa ferramenta é utilizada sobretudo para classificar uma dada tecnologia emergente dentro das expectativas daquele ano específico.

Como a história de diferentes tecnologias nos mostra, é evidente que essa representação é uma excessiva simplificação da realidade. O percurso entre o desenvolvimento de uma dada tecnologia e sua aplicação comercial depende de uma variada gama de vetores que envolvem, além de outros fatores, elementos comerciais, sociais, climáticos, pessoais, governamentais e o próprio acaso. Mesmo com apoio na opinião de especialistas e empregando as mais modernas técnicas de aprendizado de máquina, hoje ainda é impossível definir as chances de aplicação de uma tecnologia recém-lançada.

Mullany (2016) comprova, empiricamente, que o *hype cycle* simplesmente não consegue prever o desenvolvimento de tecnologias e que, não raro, inúmeras tecnologias simplesmente morrem. Além disso, não existem dados – pelo menos disponíveis – que sustentem, empiricamente, a forma como a curva é proposta.

Em contrapartida, é verdade que muitas tecnologias que chegaram ao platô de produtividade, de alguma forma, passaram pelas fases descritas pela consultoria. Isso não quer dizer que elas sejam fruto de um processo linear de maturação histórica, isto

é, que apareçam na sequência proposta, nem que seja necessário passarem por todas as fases. O fato é que é possível inserir as tecnologias atualmente existentes em algum lugar deste ciclo e isso nos ajuda na qualificação dos temas tecnológicos que queremos encontrar (diferentemente do caso aqui explorado, o *hype cycle* trata de tecnologias específicas e não de temas de discussão).

O que é interessante para nós aqui é o fato de que, a cada fase da história da tecnologia, haverá temas específicos que podem ou não se repetir. Assim, se nas fases iniciais se discutem possibilidades disruptivas, nas fases finais, espera-se que os temas estejam mais associados a desafios específicos e problemas de implantação. Ou seja, o ciclo de tecnologias emergentes é relevante para ajudar a explicar a evolução dos temas de cada período, mas não mais que isso.

É importante destacar que, neste texto, não estamos preocupados em determinar quais são as tecnologias emergentes, nem em definir sua maturidade ao longo do tempo. Nossa preocupação central é separar discussões emergentes sobre tecnologias e suas aplicações. Isso porque não podemos perder de vista nosso objetivo, que é criar uma ferramenta para reduzir o ruído de informação e dar subsídios de maior qualidade à política tecnológica. A base de dados selecionada para dar conta desse objetivo é aquela formada pelos tuítes de usuários selecionados.

O Twitter é uma rede social destinada a divulgar *microblogs* de até 280 caracteres. Cada postagem é chamada de tuíte (*tweet*). Trata-se de um serviço gratuito no qual pessoas físicas e jurídicas podem se inscrever, postar e seguir contas selecionadas. Os temas de postagem são livres e podem trazer *links* para assuntos relacionados. De fato, é comum que um tuíte seja uma forma de divulgar uma discussão mais abrangente localizada em outros lugares da rede.

Segundo dados da própria plataforma, em 2019, o Twitter continha cerca de 330 milhões de usuários, sendo a maior parte deles dos Estados Unidos (aproximadamente 20% do total).

O uso do Twitter como ferramenta de análise de sentimentos já é bem conhecido e documentado, já tendo sido empregado, por exemplo: na previsão do resultado de eleições (Skoric *et al*., 2012;Songe Kim, 2013); na identificação do comportamento

de mercados financeiros (Daniel, Neves e Horta, 2017; Oliveira, Cortez e Areal, 2017); na avaliação de destinos turísticos (Ramanathan e Meyyappan, 2019); na definição de preferências por marcas de vestuário (Rasool *et al*., 2019); bem como na análise das expectativas tecnológicas ao longo do tempo (Li *et al*., 2019).

Nesse já vasto arcabouço de referências, destaca-se o trabalho de Shen, Urquhart e Wang (2019), que encontraram uma relação entre o volume de tuítes associados às *bitcoins* e à volatilidade do preço destas no dia posterior. Para avaliar a interação entre *bitcoins* e *tweets*, os autores estimaram um modelo de vetores autorregressivos (VAR) e testaram a causalidade tanto linear quanto não linear, resultando numa metodologia complexa, porém robusta.

O fato é que a base de dados do Twitter, disponível através de uma *application programming interface* (API), constitui-se numa das fontes mais dinâmicas e robustas sobre o comportamento das pessoas, instituições e Estados. A partir dela é possível observar tanto grandes tendências (preferências políticas nacionais) como tendências muito específicas (opinião quanto à qualidade de um produto de uma dada empresa). Nesse sentido, seu emprego mostra-se fundamental na moderna economia de mercado.

A análise de sentimento, que é a mais comum metodologia a empregar o Twitter como fonte de informação, pode ser definida como:

> *Sentiment analysis (also called opinion mining, review mining, or attitude analysis) is the task of detecting, extracting, and classifying opinions, expectations, and emotional reactions about different topics, as expressed in textual input (Ravi and Ravi, 2015). Sentiment analysis has become one of the most popular research approaches used for practical applications and academic research* (Li *et al*., 2019).

Tal análise pode empregar uma série de ferramentas, com ou sem supervisão humana, isto é, já é possível fazer análise de opiniões e atitudes a partir de aprendizado de máquinas. Contudo, o mais comum tem sido o emprego de técnicas mistas.

Seja como for, este texto encontra-se numa fase anterior da análise de sentimentos. Nosso objetivo é outro: identificar temas tecnológicos emergentes e não o sentimento das instituições sobre eles. Ou seja, queremos primeiro compreender qual é o debate emergente que determinadas instituições estão realizando. Num segundo momento, talvez, seja necessário definir e qualificar a opinião destas mesmas instituições

sobre os temas identificados. Portanto, este não é um estudo de sentimento a partir de dados do Twitter, mas, sim, um esforço de pesquisa voltado a identificar debates emergentes no campo tecnológico.

Nesse sentido, nosso esforço de pesquisa se assemelha mais ao de Naaman, Becker e Gravano (2011). Esses autores, que estão preocupados em classificar e tipificar os tipos de *trends* (principalmente entre exógenos e endógenos à plataforma), utilizam como definição de *trend* a frequência do determinado termo em *t*, considerando sua frequência em *t-1*. Isto é, procuram termos que apareceram com frequência superior ao que seria esperado:

> *Specifically, to identify terms that appear more frequently than expected, we will assign a score to terms according to their deviation from an expected frequency. Assume that M is the set of all messages in our Tweets dataset, R is a set of one or more terms to which we wish to assign a score, and h, d, and w represent an hour of the day, a day of the week, and a week, respectively. We then define M(R, h, d, w) as the set of every Twitter message in M such that (1) the message contains all the terms in R and (2) the message was posted during hour h, day d, and week w. With this information, we can compare the volume in a specific day/hour in a given week to the same day/hour in other weeks* (Naaman, Becker e Gravano, 2011, p. 906).

O objetivo do referido estudo era verificar se os *trends* gerados por discussões surgidas na própria plataforma eram distintos – segundo categorias propostas – daqueles gerados exogenamente, por exemplo, em razão de eventos históricos. Os autores concluem que os *trending topics* gerados a partir de discussões surgidas na própria plataforma são, segundo as categorias propostas, distintos daqueles gerados a partir de fatos externos, como notícias de *sites* ou de jornais. Assim, encontrou-se, dentre outras coisas, que a difusão de termos endógenos entre autores depende de reciprocidade entre eles, o que não se verifica no caso de eventos externos.

Além desses esforços independentes de identificação de temas emergentes, a própria plataforma (Twitter) possui um algoritmo que seleciona os *trending topics*. Acontece que tal algoritmo é calculado para todo o conjunto de tuítes da plataforma, independentemente do autor, e, de forma geral, fundamenta-se numa relação tempo/volume. Ou seja, um *trend topic* surge a partir de um crescimento rápido num curto espaço de tempo. Como veremos neste texto, tal método não é adequado aos objetivos aqui propostos, principalmente porque um crescimento vertiginoso, mesmo de um tema

inédito, corresponde mais a um efeito efêmero (*hype*) do que à introdução de um tema de discussão emergente (é importante destacar que aqui não estamos identificando *trending topics,* e sim temas *emergentes*).

É importante perceber que, em todos os estudos citados, a questão central não é identificar os temas emergentes, mas, a partir de tal identificação, realizar análises e avaliações. Acontece que, dado o nosso objetivo, a identificação do tema é a parte da pesquisa mais relevante e, por isso, exige um esforço customizado de desenvolvimento. Isso porque: *i*) nosso objeto é muito específico; *ii*) o tempo de análise é longo; e *iii*) é preciso intervenção humana, tanto para selecionar os autores das discussões quanto para agregar tópicos tecnologicamente semelhantes com grafias distintas (por exemplo, os termos *Alpha Go* e *IoT improving* são parte de uma mesma discussão e só seriam unidos sem a intervenção humana mediante um grande esforço inédito de aprendizado de máquina).

Sendo assim, e na medida em que este texto para discussão desenvolveria um esforço de pesquisa exploratório a partir de uma base de dados não usual, as possibilidades metodológicas eram vastas. Por outro lado, também é preciso considerar que um indicador feito por *think tank* público deveria ser inteligível e facilmente reproduzível. Por isso, optamos por uma metodologia baseada na identificação de anomalias entre um semestre e outro, identificadas mediante cálculo de probabilidades, aglutinação de termos em temas e posterior ranqueamento.

Esse esforço exigiu realizar sete passos, sendo que a maioria deles foi executada com o uso da linguagem computacional Python, tal como descrito a seguir.

É importante destacar que, considerando termos tecnológicos, estamos preocupados em encontrar anomalias no conjunto de tuítes. Ou seja, queremos observar o que de fato não era esperado, mas ocorreu. Assim, um termo recorrente só é identificado quando sua frequência se torna elevada e fora do comum frente ao semestre anterior.

Diferentemente das metodologias apresentadas, o que este texto propõe é um método muito mais simples – baseado na distribuição de probabilidades –, com a presença de alguma subjetividade. Ou seja, propõe-se uma métrica simples e reproduzível, muito embora se reconheça alguma arbitrariedade em passos específicos, notadamente, na escolha das contas e na agregação de temas.

3.1 Passo 1

Seleção de contas de Twitter de instituições (revistas especializadas, instituições governamentais e grupos de mídia) de reconhecida competência na discussão sobre tecnologia e seus impactos na sociedade.

A seleção foi feita apenas para contas em língua inglesa e foi baseada na experiência do consultor e do coordenador do projeto. Portanto, trata-se de uma lista não exaustiva e que contém grande subjetividade na seleção. Sendo assim, ela deve ser revista sempre que surgirem novas contas e/ou que as contas selecionadas deixem de tratar dos temas de interesse (tabela 2).

A partir da tabela2, é possível observar que não foi realizada nenhuma separação entre ofertantes e demandantes de tecnologias. Tal fato se justifica em razão da necessidade de se captar ambas as visões de forma a ter um conjunto minimamente completo dos possíveis agentes que atuam sobre a produção tecnológica. Ou seja, partiu-se de uma abordagem neoschumpeteriana da mudança técnica na qual o processo inovativo é resultado da interação social tanto entre agentes quanto destes com seu meio.

TABELA 2
Contas selecionadas segundo objetivo declarado e número de seguidores

Conta	Endereço	Objetivo	Seguidores (mil)	Criação
Nautilus	@NautilusMag	*Science, culture, and philosophy for the intellectually curious.*	83,8	2/2012
The Futurist	@TheFuturist007	*Reporting on the latest in cutting-edge technology, scientific endeavors, and futuristic innovation.*	19,5	11/2018
Reuters Tech News	@ReutersTech	*Technology news from Reuters.*	175,9	3/2009
Bloomberg Technology	@technology	*Breaking news & analysis on the global technology industry.*	333,4	2/2009
FT Technology News	@fttechnews	*Financial Times technology news.*	230,4	12/2007
BI Tech	@SAI	*The latest tech news from @BusinessInsider.*	1,3	9/2007
FutureShift	@futureshift	*Tech news from @Mashable Obsessed w/ the innovations, products, and people who shape our connected lives.*	666,5	2/2010
HuffPostTech	@HuffPostTech	*Where technology is anthropology.*	604,2	9/2009
MIT Technology Review	@techreview	*A media company making technology a greater force for good.*	1	8/2008
BBC News Technology	@BBCTech	*The official account for the BBC News technology team with our latest stories and tweets from @BBCRoryCJ @leokelion @zsk @ BBCDaveLee @thisisfoxx @cmg_BBC*	759,7	1/2007
Guardian Tech	@guardiantech	*News and comment from the @Guardian's technology team*	2,5	8/2007
Engadget	@engadget	*Engadget is the original home for technology news and reviews.*	2,4	4/2008
The Verge	@verge	*Covers life in the future.*	2,6	4/2011
Forbes Tech	@ForbesTech	*Tech news and insights from @Forbes.*	2,7	5/2008

(Continua)

(Continuação)

Conta	Endereço	Objetivo	Seguidores (mil)	Criação
TechCrunch	@TechCrunch	*Breaking technology news, analysis, and opinions from TechCrunch. tips@techcrunch.com*	10,1	3/2007
WIRED UK	@WiredUK	*The latest from WIRED UK*	276,4	3/2009
WIRED	@WIRED	*WIRED is where tomorrow is realized.*	10,4	3/2007
DARPA	@DARPA	*Official account of the Defense Advanced Research Projects Agency.*	207,9	7/2009
Hacker News	@hntweets	*Tweets the stories on the front page of Hacker News.*	17,4	2/2010

Elaboração dos autores.

3.2 Passo 2

Download dos tuítes das dezenove contas, publicados entre 1º de janeiro de 2010 e 30 junho de 2019. Para tanto, construímos um robô que realiza o procedimento. A base final contém 674.750 tuítes. Os dados foram separados em semestres. O número total de tuítes varia ao longo dos semestres, assim como o número de contas (uma vez que a data de criação das contas também varia).

3.3 Passo 3

Na base total de tuítes, retiraram-se termos não relevantes, como contrações, palavras de ligação, emojis, elementos temporais e monetários, nomes de pessoas e entidades, bem como termos de elevada frequência (acima de 500). Essa limpeza teve por objetivo reduzir o volume de termos tratados e, com isso, reduzir o ruído na base (diminuindo, consideravelmente, o tempo necessário para a *mining*). As palavras excluídas encontram-se listadas no código disponível no GitHub.[8]

3.4 Passo 4

Após a limpeza, criamos uma matriz de dados com as frequências de todos os termos ao longo dos semestres de análise, atribuindo frequência zero nos semestres em que o termo não foi observado. Os termos foram separados palavra por palavra, denominados *unigrams*, e a cada duas palavras em sequência no texto, *bigrams*. As matrizes passaram por um processo de normalização para retirarmos o viés de diferença de volumes dos tuítes nos diferentes semestres.

8. Disponível em: <https://bit.ly/2T8tL4w>.

Procuramos, então, aplicar um modelo estatístico para identificar frequências anômalas entre semestres, em especial, estamos interessados no caso em que a frequência de ocorrência de um termo em um semestre é consideravelmente inferior frente ao semestre seguinte. Como modelo, usamos a distribuição de Poisson, que é um modelo matemático para eventos aleatórios, habitualmente utilizado em teoria das filas.

Uma distribuição de Poisson com parâmetro λ no intervalo ([0, ∞) é um gerador aleatório em que:

- o número de eventos que ocorrem por determinada unidade de medida (no nosso caso, um período *t,* que representa um semestre) é uma variável aleatória com distribuição de Poisson (λ); e
- o número de eventos em intervalos disjuntos (sem sobreposição) é variável aleatória independente.

De acordo com a distribuição de Poisson, a probabilidade de um termo aparecer *k* vezes (frequência observada em um tempo *t*, no nosso caso, um semestre) é dada por:[9]

$$'(X = \ k) = pmf(\mathrm{k}; \lambda) = \ \frac{\lambda^k e^{-\lambda}}{k!} \quad (1)$$

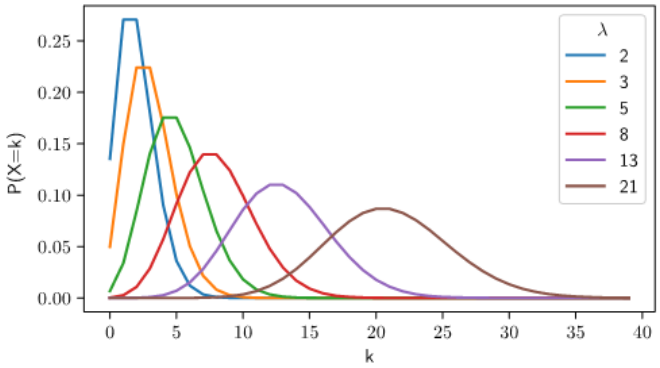
em que:

- *pmf* é a função massa de probabilidade;
- e é a base do logaritmo natural (e = 2, 71828...);
- λ é o parâmetro do modelo que indica o número esperado de ocorrências no intervalo de tempo *t*; e
- *k* é o número de ocorrências em um intervalo de tempo *t*.

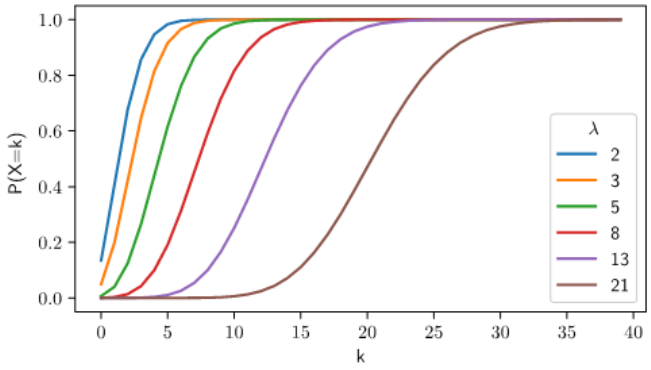
O parâmetro λ altera a assimetria da distribuição de Poisson. Saliente-se que a distribuição não é simétrica, como podemos observar no gráfico 4, a seguir:

9. Sobre a distribuição de Poisson ver, por exemplo: <https://bit.ly/398sujF>.

GRÁFICO 4
CDFs e PMFs para diferentes valores de λ
4A – Distribuição de Poisson (CDF)



4B – Distribuição de Poisson (PMF)



Elaboração dos autores.
Obs.: 1. CDF – função distribuição acumulada; PMF – função massa de probabilidade.
2. Gráfico cujos leiaute e textos não puderam ser padronizados e revisados em virtude das condições técnicas dos originais (nota do Editorial).

A função distribuição acumulada de Poisson traz o somatório de todas as realizações da variável aleatória X até um número k de ocorrências:

$$cdf(k;\lambda) = P(X \leq k) = \sum_i^k \frac{\lambda^i}{i!} e^{-\lambda} \quad (2)$$

Geralmente, distribuições de Poisson são utilizadas para, dado o conhecimento de uma certa frequência λ em um período *t-1* no passado, estimar a probabilidade de o mesmo evento ocorrer um número k de vezes em um período t no futuro. Este, entretanto, não é o nosso caso.

O que estamos fazendo, ao contrário, é usar apenas dados passados para caracterizar uma relação entre a frequência com que um termo aparece em dois semestres diferentes (que podem ser considerados variáveis aleatórias independentes), classificando se é uma relação anômala em função de crescimento ou não. Ou seja, estamos observando a frequência passada (*t-1*) e questionando se ela seria esperada, dada a frequência observada em (*t*).

A opção por uma abordagem menos intuitiva se deve à necessidade de dar tratamento adequado à incerteza. Isto é, numa abordagem mais intuitiva, no caso de um termo que saísse de uma frequência no tempo *t-1*, *f*(termo, *t-1*), igual a 0, estaríamos impossibilitados de calcular a probabilidade da frequência observada no tempo subsequente *t*, *f*(termo, *t*), utilizando *f*(termo, *t-1*) como parâmetro da Poisson, pois tanto *pmf*(λ=0) quanto *cdf*(λ=0) são indefinidas. Em outras palavras, como calcular a probabilidade de um termo em *t*+1 se o termo nunca ocorreu?

Assim, a utilização mais intuitiva do método impossibilitaria identificar termos inéditos, que são a essência do conceito de emergente, invalidando a metodologia. É verdade, contudo, que poderíamos somar valores aos termos com frequência zero em *t*-1, mas isso seria ignorar que a incerteza, por definição, não pode ser medida. Ou seja, qualquer tentativa *ad hoc* de adicionar um valor arbitrariamente pequeno ε a λ pode modificar o resultado de maneira imprevisível. O resultado prático dessa estratégia seria a contaminação das probabilidades de termos que não foram citados no semestre anterior, mas tiveram alta frequência no semestre observado.

Aproveitando que a ocorrência de um termo em diferentes semestres é independente, ao invés de calcular (como seria mais intuitivo) a probabilidade de k = *f*(termo, *t*), usando como parâmetro λ = *f*(termo, *t-1*), fazemos o contrário:

- λ = *f*(termo, t); e
- k = *f*(termo, t-1).

Ou seja, esperamos que o passado tenha sido igual ao que sabemos do último semestre.

Dessa forma, nos casos em que consideramos λ= *f*(termo, t) = 0, se anteriormente ele possuía frequência positiva, então o termo "desapareceu", e este definitivamente não é o nosso objeto de estudo; assim, não é um problema que não possamos calcular a probabilidade da frequência em *t*-1, *f*(termo, *t*-1).

Sendo assim, como estamos interessados pelas anomalias da cauda esquerda, ou seja, valores muito baixos na distribuição, basta selecionarmos os mil termos com menores probabilidades acumuladas, os quais classificamos como emergentes.

3.5 Passo 5

De posse dos mil termos com menores probabilidades, construímos duas listas (*unigram* e *bigram*) com os cem termos com as menores probabilidades para cada semestre (2015-1 a 2019-1).[10]

3.6 Passo 6

Os cem termos selecionados nas duas listas foram analisados um a um e apenas aqueles que diziam respeito a uma dada tecnologia (*software*, *hardware* ou sistema) foram selecionados para a etapa seguinte. Assim, essa primeira seleção buscou referências a: *i*) tecnologias específicas, inclusive abreviadas; *ii*) tipos de tecnologias; e *iii*) produtos ou empresas novos e ligados diretamente a uma dada tecnologia (isto é, quando determinado produto é a própria tecnologia ou quando a tecnologia é exclusiva de determinada empresa e dela não pode ser dissociada).

Os critérios de seleção se basearam em análises de textos técnicos e na busca *on-line* por aplicações. Ou seja, mesmo depois dos procedimentos iniciais, que visavam diminuir o ruído, este ainda era muito grande, mesmo nos cem termos selecionados (apareciam muitos termos ligados a eventos conjunturais e não tecnológicos).

Cada semestre então possui duas listas de termos relacionados a determinadas tecnologias, uma para termos simples e outra para termos compostos. Com isso, o número de termos tecnológicos retirados dos cem termos inicialmente selecionados nas duas listas varia conforme o semestre e a interpretação da equipe.

10. Optamos por considerar os tuítes repetidos, uma vez que acreditamos que eles significam uma ratificação da instituição da relevância do tuíte original. Isto é, ao repetir um tuíte, a instituição está reafirmando a relevância do que afirmou antes e, portanto, evidenciando a discussão específica.

3.7 Passo 7

Os termos das listas de simples e compostos foram agrupados, quando possível, em torno de temas comuns. Cada vez que foi identificada a relação de um termo com um tema, o último ganhava um ponto. A somatória dos pontos permitiu criar um *ranking* de temas tecnológicos emergentes num semestre.[11] No sentido de selecionar ainda mais os termos, definiu-se, de maneira *ad hoc* que, para ser considerado um tema tecnológico emergente era preciso ter, no mínimo, 3 pontos (ou seja, três termos, independentemente de serem simples ou compostos).

Por exemplo, no primeiro semestre de 2019, dos cem termos selecionados, apenas 10 termos simples e 17 termos compostos atendiam aos critérios estabelecidos. Destes, 8 termos (simples e compostos) puderam ser agrupados em torno de um tema que foi denominado de "aplicações de IA (inteligência artificial) " e 4 no tema "nova geração de tecnologias de comunicação". Portanto, a aglutinação de termos simples e compostos em torno de temas comuns dá origem ao ITTE, que depois é qualificado e detalhado por análise qualitativa.

Em resumo, a metodologia aqui proposta tem por objetivo encontrar crescimentos não esperados (anomalias) na frequência de termos, aglutinados em grandes temas e que finalmente são ranqueados. Tem-se, assim, um indicador que mede a emergência de um dado tema no debate tecnológico internacional.

4 RESULTADOS

Seguindo a metodologia proposta, os quadros1 a 9 apresentam os resultados semestrais do ITTE para o período de 2015 a 2019.

11. Na medida em que este passo carrega grande subjetividade, é relevante que futuros estudos possam fazer uma análise de sensibilidade, não para validar a análise, mas para compreender os níveis de influência das decisões humanas.

QUADRO 1
Resultados para o primeiro semestre de 2015

Ranking	Grande tema	Termos encontrados (grafia em inglês)
1	Serviços de *streaming* de música e dados	*Tidal, music service, streaming music, live streaming, music streaming, streaming service, sling tv.*
2	*Wearableas*	*Pebble, pebble time, watch launch, watch apps, watch app.*
3	Aplicativos como modelo de negócio	*App-store, app-store marketing, photos app.*
	Aplicações em realidade virtual/realidade aumentada	*Hololens, holographic, oculustouch.*
	Tecnologias automobilísticas digitais	*Connected car, self-driving car, carplay auto.*

Elaboração dos autores.

QUADRO 2
Resultados para o segundo semestre de 2015

Ranking	Grande tema	Termos encontrados
1	Aplicações em internet das coisas	*Iot, iot affect, examining iot, internet examining.*
2	Energias limpas	*Clean energy, solar panels, smart battery.*
3	Inteligência artificial	*Robot ai, ai reads, recommended robot.*

Elaboração dos autores.

QUADRO 3
Resultados para o primeiro semestre de 2016

Ranking	Grande tema	Termos encontrados
1	Atendimento virtual autônomo	*Bots, chatbot, bot, chatbots, chat bot, messenger bots.*
	Avanços em inteligência artificial	*Deepmind, alphago, deepmind ai, ai beats, go world, game go.*
2	Criptografia e privacidade na rede	*Encryption, encryption fight, fbi encryption, encryption battle, end-to-end encryption.*
3	Comercialização em realidade virtual	*Vr headset, future vr, oculus rift, vr ar.*

Elaboração dos autores.

QUADRO 4
Resultados para o segundo semestre de 2016

Ranking	Grande tema	Termos encontrados
1	Compartilhamento de carros e de carros autônomos	*Self-driving, ride-hailing, investing self-driving, ride-hailing companies, ubers self-driving.*
2	Comercialização em realidade virtual	*Niantic, daydream, daydreamvr.*

Elaboração dos autores.

QUADRO 5
Resultados para o primeiro semestre de 2017

Ranking	Grande tema	Termos encontrados
1	Ataques cibernéticos/segurança na rede	*Ransomware, wannacry, phishing, cyber-attack, ransomware attack, wannacry ransomware, ransomware attacks, cyber attack, global ransomware, cia hacking, global cyber, phishing attack.*
2	Avanços em inteligência artificial	*Alphago, iot improving, use ai, alphago ai, bixby.*
3	Privacidade na rede	*Essential internet, internet privacy, privacy rules, broadband privacy.*
4	Internet das coisas	*Iot, internet iot, iot essential.*

Elaboração dos autores.

QUADRO 6
Resultados para o segundo semestre de 2017

Ranking	Grande tema	Termos encontrados
1	Mercado de *criptocurrency*	*Cryptocurrency, cryptocurrencies, coin, ethereum, coinbase, bitcoins, bitcoin futures, initial coin, coin offerings, bitcoin cash, cryptocurrency trading, bitcoin bubble.*
2	Aplicações de inteligência artificial	*Ai experts, ai lab, ai insight, minds ai, chatbots impact, killer robots.*
3	Veículos elétricos	*Electric truck, electric cars, model 3s.*

Elaboração dos autores.

QUADRO 7
Resultados para o primeiro semestre de 2018

Ranking	Grande tema	Termos encontrados
1	Mercado de *criptocurrency*	*Cryptocurrency, crypto, coincheck, cryptocurrency mining, ban cryptocurrency, cryptocurrency Exchange, cryptocurrency ads, crypto Exchange, bans cryptocurrency, cryptocurrencies matter, blockchains cryptocurrencies.*
2	Aplicações comerciais em *Blockchain*	*Business blockchain, brightest blockchain, blockchain innovators, blockchains cryptocurrencies.*
	Robôs e inteligência artificial	*Ai robots, ai impact, robotics ai, ai driving.*

Elaboração dos autores.

QUADRO 8
Resultados para o segundo semestre de 2018

Ranking	Grande tema	Termos encontrados
1	Computação quântica	*Powerful quantum, quantum computing, quantum computers, quantum version.*
2	Câmeras fotográficas sem espelho	*Mirror, mirrorless, mirrorless cameras, full-frame mirrorless.*
	Veículos elétricos	*E-tron, cars electric, electric vehicles, cars startup*.
3	Nova geração de tecnologias de comunicação	5g, *5g phones, ahead 5g.*

Elaboração dos autores.

QUADRO 9
Resultados para o primeiro semestre de 2019

Ranking	Grande tema	Termos encontrados
1	Aplicações da inteligência artificial	*Ai-powered, ai-powered surveillance, confuse ai-powered, algorithm shows, ai headed, figure ai, likes deepmind, intelligence help.*
2	Nova geração de tecnologias de comunicação	5g, *5g internet-connected, scared 5g, internet-connected everything.*
3	Captura de carbono	*Carbon-capture, carbono, carbon-capture machines.*
	Baixa tecnologia	*Low-tech, low-tech inventions, recent low-tech.*

Elaboração dos autores.

5 VALIDAÇÃO

Uma vez obtidos os resultados, foi necessário buscar validação em algum grau, mesmo que de forma indireta. É evidente, nesse sentido, que o exercício não permite uma análise de validação estatística tradicional, isto é, não se pode falar em *p* valor.

É possível, no entanto, testar o momento em que o tema foi identificado como emergente. Isto é, pode-se verificar se, dentro do histórico de interesse na internet pelo tema, o ITTE se aproxima de seu momento de maior popularidade.

Evidentemente, não se espera uma relação direta entre o interesse pelo tema na internet e o ITTE, até porque o indicador aqui proposto se baseia no comportamento frente ao último semestre. Porém, caso o interesse pelo tema na internet seja *posterior ou concomitante* ao momento identificado pelo ITTE, tem-se um forte argumento a favor da qualidade do indicador.

Se, por exemplo, o termo IoT (*internet of things*) aparece no *ranking* ITTE, então, em algum momento próximo da identificação (preferencialmente depois), esse tema deve ter recebido um interesse especial, não só de comunidades específicas como da comunidade geral. Nessa perspectiva, tem-se a validação da emergência do tema frente a seu próprio comportamento passado.

A maneira mais intuitiva de verificar tal ocorrência é através do histórico de pesquisas do maior buscador da internet, o Google Search. Por isso, empregamos a ferramenta *on-line* Google Trends, que apresenta os picos de frequência das buscas de termos de pesquisa específicos. Isto é, o Google Trends responde quando o termo foi mais buscado, dado todo o histórico de pesquisa. O interesse de pesquisa é medido em pontos que vão de 0 (nenhum interesse) a 100 (máximo interesse até o momento). Segundo o Google Trends:

> Os números representam o interesse de pesquisa relativo ao ponto mais alto no gráfico de uma determinada região em um dado período. Um valor de 100 representa o pico de popularidade de um termo. Um valor de 50 significa que o termo teve metade da popularidade. Uma pontuação de 0 significa que não havia dados suficientes sobre o termo.[12]

12. Disponível em: <https://trends.google.com.br/trends/?geo=BR>.

Assim como em nosso indicador, empregamos termos em inglês idênticos ou muito próximos (alguma adaptação foi feita, uma vez que se trata de ferramenta de pesquisa e não de comunicação, tal como um tuíte) sem definição de país específico. Além disso, para o teste, procuramos os termos que de forma mais específica ilustram um tema. Foram testados os primeiros temas do *ranking* ITTE. Os dados foram obtidos em 24 de outubro de 2019.

TABELA 3
Validação segundo o Google Trends

Termos da busca	Semestre do TTE	Máxima pontuação no semestre do ITTE	Mês de pico no Google Trends	Validação frente ao período de pico do Google Trends
Tidalmusicservice	2015-1	100	Abril de 2015	ITTE no mesmo período
Internet of Things	2015-2	89	Outubro de 2016	ITTE antecipou
Chatbot	2016-1	30	Outubro de 2018	ITTE antecipou
Uber self-driving	2016-2	33	Março de 2018	ITTE antecipou
Ransomware	2017-1	100	Maio de 2017	ITTE no mesmo período
Criptocurrency	2017-2	93	Janeiro de 2018	ITTE antecipou
Coincheck	2018-1	100	Janeiro de 2018	ITTE no mesmo período
Quantum computing	2018-2	57	Fevereiro de 2004	ITTE em período posterior
Ai-powered	2019-1	100	Janeiro	ITTE no mesmo período

Elaboração dos autores.

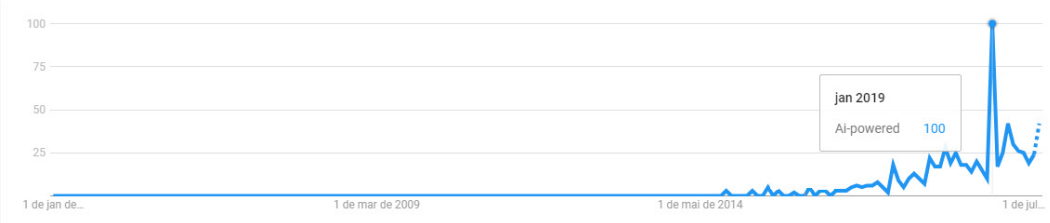
A tabela 3 sugere, então, a validade do ITTE como instrumento de acompanhamento de discussões emergentes. Com exceção do período 2018-2, em todos os períodos, o ITTE antecipou um grande interesse geral de pesquisa ou coincidiu com os picos de interesse. Ou seja, nossa metodologia, de fato, apresenta temas emergentes.

Além disso, a validação do período 2018-2 precisa ser relativizada, uma vez que a computação quântica é um tema antigo que ganhou relevância no período. Isto é, se, frente à história das buscas, esse período não é tão relevante, ele o é frente ao semestre anterior. Assim, o que o ITTE mostra nesse caso é a antecipação de uma retomada do interesse, haja vista o fato de que o indicador saltou de 57 pontos, em dezembro de 2018, para 73, em janeiro de 2019.

Adicionalmente, quando o segundo termo do *ranking*, "*mirrorless*" (referente ao tema de câmeras digitais sem espelho), é buscado na ferramenta, ele apresenta pico máximo no mesmo período do ITTE. Este fato corrobora a qualidade do indicador aqui proposto.

O gráfico 5 ilustra como a ferramenta Google Trends foi empregada na validação do indicador aqui proposto. Nele, é possível observar que o termo buscado (*ai-powered*, referente ao tema aplicações da inteligência artificial), retirado do ITTE, tem um crescimento exponencial justamente no mesmo período aqui identificado.

GRÁFICO 5
Interesse ao longo do tempo do termo *Ai-powered*, aqui identificado como emergente em 2019-1



Fonte: Google Trends. Disponível em: <https://bit.ly/3af4VWp>. Acesso em: 24 out. 2019.
Obs.: Gráfico cujos leiaute e textos não puderam ser padronizados e revisados em virtude das condições técnicas dos originais (nota do Editorial).

Portanto, os exercícios de validação conduzidos, apesar de não permitirem realizar nenhuma análise de validade estatística, permitem deduzir que o indicador aqui proposto tem forte aderência com a realidade empírica.

6 DISCUSSÃO

Especificamente sobre os resultados do primeiro semestre de 2019, é possível observar que os dois temas mais emergentes são a IA e a quinta geração das tecnologias de comunicação (5G). Não obstante, discussões sobre a captura de carbono e inovações baseadas em baixa tecnologia também foram identificadas como emergentes.

No sentido de ilustrar as possibilidades analíticas da metodologia aqui proposta, optou-se por detalhar apenas as discussões sobre IA e 5G. O detalhamento dos outros temas seguiria, então, a mesma lógica.

Ao analisar todos os termos que compõem o tema da aplicação de IA, observamos a repetição de termos ligados a aparelhos ou sistemas que funcionam com IA. Nesse mesmo sentido, ao se elaborar uma nuvem de palavras a partir dos tuítes que continham os termos selecionados com o tema denominado de "aplicações da inteligência artificial" –

retirando os termos específicos e deixando apenas as palavras associadas –, observa-se que a aplicação na área de saúde (detecção precoce de câncer de pulmão) também possui destaque, assim como questões ligadas a ganhos de velocidade e confiabilidade no uso da IA (figura 1).

FIGURA 1
Nuvem de palavras dos tuítes ligados ao tema "aplicações de inteligência artificial" (2019-1)[1]

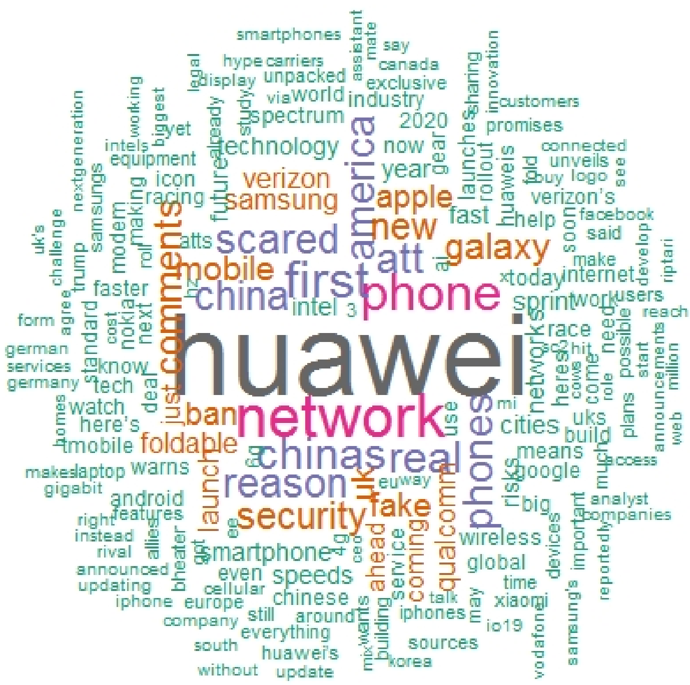
Elaboração dos autores.
Nota: [1] Para este exercício, mantivemos os termos *inteligence* e *help*, pois eles foram usados em outros tuítes e surgiram em função de apenas um único tuíte.
Obs.: Apenas termos associados.

No Brasil, a discussão desse tema tem implicações diretas na aplicação da Estratégia Digital Brasileira e na própria definição das agendas de pesquisa prioritárias. De fato, o ITTE corrobora a necessidade de aprofundar tais discussões que envolvem, inclusive, questões de segurança nacional e defesa.

Destacam-se ainda discussões relacionadas ao emprego da IA na detecção precoce de câncer de pulmão através de exames de imagens. Um aprofundamento dessa discussão pode ser valioso para estratégias do Sistema Único de Saúde brasileiro (SUS), principalmente para estratégias de prevenção.

O tema “nova geração de tecnologias de comunicação”, que está associado à repetição de termos ligados à 5G, ficou em segundo lugar no semestre 2019-1 e, ao se realizar o mesmo exercício de nuvens de palavras feito para o tema IA, percebe-se que a discussão sobre a liderança e protagonismo da empresa chinesa Huawei foi destaque (figura 2).

FIGURA 2
Nuvem de palavras dos tuítes ligados ao tema 5G (2019-1)



Elaboração dos autores.
Obs.: Apenas termos associados.

Em realidade, tal discussão está associada a questões de segurança nacional e dependência tecnológica, uma vez que a tecnologia encontra-se numa fase de definições de padrões que podem causar forte inércia em decisões futuras. Ou seja, o tema trata de quem dominará essa tecnologia habilitadora no futuro próximo.

É evidente, contudo, que as dificuldades (e mesmo banimento) da empresa em operar no mercado norte-americano impulsionaram as discussões (ver, por exemplo, Technology..., 2019). De fato, durante vários dias, a relação do governo norte-americano com a empresa chinesa foi tema de diversas matérias na imprensa internacional.

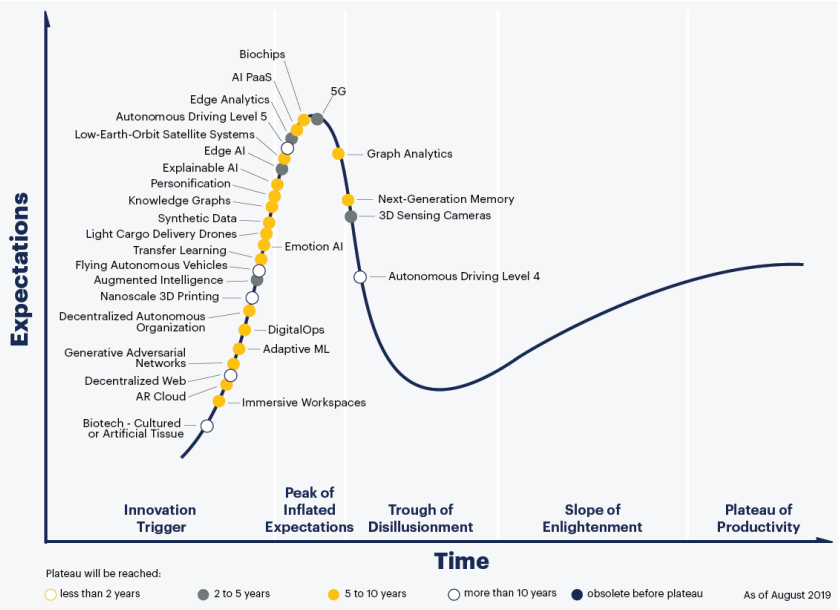
Para o Brasil, esse tema é central, uma vez que o país se prepara para adotar a tecnologia e ainda precisa fazer inúmeras escolhas sobre padrões e protocolos, bem como sobre a infraestrutura necessária para fazer funcionar tal tecnologia. Acompanhar essa discussão, portanto, pode fornecer respostas relevantes ao caso nacional.

Uma maneira alternativa de qualificar a ferramenta aqui proposta é através da comparação com o ciclo *hype* de tecnologias emergentes. Esta, apesar de todas as ressalvas já feitas, é uma ferramenta de monitoramento tecnológico interessante quando o objetivo é classificar uma tecnologia emergente quanto às expectativas dentro de um dado ano.

Como mostra o gráfico 6, tanto as questões que envolvem a IA quanto as tecnologias de comunicação de quinta geração estão no auge das expectativas. Contudo, tais tecnologias parecem ter temas distintos de discussão.

Enquanto, no caso da IA, as discussões já se encontram em temas relacionados a aplicação real, os temas da 5G (inédita nesta fase do ciclo *hype*, quando são considerados anos anteriores) não apresentam qualquer especificidade. De fato, a IA já tinha aparecido, de uma forma ou de outra, nesta mesma fase em anos anteriores.

GRÁFICO 6
Ciclo *hype* de tecnologias emergentes para 2019



Fonte: Gartner *apud* Panetta (2019).
Obs.: Gráfico cujos leiaute e textos não puderam ser padronizados e revisados em virtude das condições técnicas dos originais (nota do Editorial).

Evidentemente, o resultado do ciclo *hype* de 2019 ajuda não só a qualificar a discussão aqui realizada, mas também a validar a metodologia proposta. Pode-se afirmar isso visto que os temas aqui identificados têm relação direta com as tecnologias de maior expectativa no ano de 2019, independentemente do que possa acontecer com tais tecnologias em momentos posteriores.

7 CONSIDERAÇÕES FINAIS

Este texto para discussão propôs um indicador de temas tecnológicos emergentes cujo objetivo é o de ser utilizado para reduzir o vasto ruído de informações atualmente produzido.

A aplicação dessa ferramenta no primeiro semestre de 2019 permitiu observar que os temas tecnológicos emergentes dizem respeito a desafios da aplicação de IA, bem como ao domínio inicial dos protocolos e trajetórias da quinta geração de tecnologias de comunicação.

A metodologia aqui proposta tem a vantagem de ser rápida, barata e de fácil entendimento. Portanto, pode ser aplicada semestralmente por qualquer instituição ou indivíduo e empregada para lançar algum foco num universo vasto e crescente de informações.

Finalmente, é preciso destacar uma importante limitação do exercício aqui realizado. Toda a análise foi baseada em contas estrangeiras de países líderes tecnológicos de língua inglesa; portanto, discussões de países como a China não são captadas. Além disso, pode-se questionar se as discussões realizadas por estes países são diretamente válidas para a realidade periférica brasileira.

De qualquer forma, a metodologia aqui proposta tem o mérito de apontar as discussões tecnológicas emergentes de boa parte do mundo industrializado, e isso, por si só, tem grande relevância na qualificação do debate no Brasil.

Aponta-se, por fim, que futuros possíveis desdobramentos incluem: *i*) o uso de IA na escolha das contas e na aglutinação de termos; e *ii*) introdução de ponderação por número de contas diferentes em que aparece o termo.

REFERÊNCIAS

ALLEN, D. *et al*. Blockchain and the evolution of institutional technologies: implications for innovation policy. **Research Policy**, v. 49, n. 1, 24 p., 2020.

BAUER, D.; WUTZKE, R.; BAUERNHANSL, T. Wear@ Work–A new approach for data acquisition using wearables. **Procedia CIRP**, v. 50, p. 529-534, 2016.

BJÖRKLUND, F. E-government and moral citizenship: the case of stonia. **Citizenship studies**, v. 20, n. 6-7, p. 914-931, 2016.

BRYNJOLFSSON, E.; MCAFEE, A. **The second machine age**: work, progress, and prosperity in a time of brilliant technologies. New York: WW Norton & Company, 2014.

_____. **Machine, platform, crowd**: harnessing our digital future. New York: WW Norton & Company, 2017.

DANIEL, M.; NEVES, R. F.; HORTA, N. Company event popularity for financial markets using Twitter and sentiment analysis. **Expert Systems with Applications**, v. 71, p. 111-124, 2017.

DAVIDSON, S.; FILIPPI, P.; POTTS, J. Blockchains and the economic institutions of capitalism. **Journal of Institutional Economics**, v. 14, n. 4, p. 639-658, 2018.

DESJARDINS, J. What happens in an internet minute in 2019? **Visual Capitalist**, Mar. 13, 2019. Disponível em: <https://bit.ly/3fMMpb7>.

KELLY, K. **The inevitable**: understanding the 12 technological forces that will shape our future. [s.l.]: Penguin, 2017.

LI, X. *et al*. Identifying and monitoring the development trends of emerging technologies using patent analysis and Twitter data mining: the case of perovskite solar cell technology. **Technological Forecasting and Social Change**, v. 146, p. 687-705, 2019.

LINDEN, A.; FENN, J. Understanding Gartner's hype cycles. **Strategic Analysis Report**, n. R-20-1971. Gartner, 2003.

MULLANY, M. **8 lessons from 20 years of hype cycles**. Linked-in, 2016. Disponível em: <https://bit.ly/3curfNI>. Acesso em: 4 nov. 2019.

NAAMAN, M.; BECKER, H.; GRAVANO, L. Hip and trendy: characterizing emerging trends on Twitter. **Journal of the American Society for Information Science and Technology**, v. 62, n. 5, p. 902-918, 2011.

OLIVEIRA, N.; CORTEZ, P.; AREAL, N. The impact of microblogging data for stock market prediction: using twitter to predict returns, volatility, trading volume and survey sentiment indices. **Expert Systems with Applications**, v. 73, p. 125-144, 2017.

O'NEIL, C. **Weapons of math destruction: how big data increases inequality and threatens democracy**. New York: Broadway Books, 2016.

PANETTA, K.**5 trends appear on the Gartner hype cycle for emerging technologies, 2019**. Gartner, 2019. Disponível em: <https://gtnr.it/2TlUI50>. Acesso em: 14 jan. 2020.

RAMANATHAN, V.; MEYYAPPAN, T. Twitter text mining for sentiment analysis on people's Feedback about Oman Tourism. *In*: MEC INTERNATIONAL CONFERENCE ON BIG DATA AND SMART CITY, 4., 2019, Muscat. **Anais**... Muscat: IEEE, 2019. p. 1-5.

RASOOL, A. *et al*. Twitter sentiment analysis: a case study for apparel brands. **Journal of Physics Conference Series**. IOP Publishing, 2019. p. 022015.

ROGERS, P.; PURYEAR, R.; ROOT, J. Infobesity: the enemy of good decisions. **Insights: BainBrief**, v. 11, 2013.

SCHUMPETER, J. **A teoria do desenvolvimento econômico**. São Paulo: Abril Cultural, 1982.

SHEN, D.; URQUHART, A.; WANG, P. Does Twitter predict Bitcoin? **Economics Letters**, v. 174, p. 118-122, 2019.

SKORIC, M. *et al*. Tweets and votes: a study of the 2011 Singapore general election. In: HAWAII INTERNATIONAL CONFERENCE ON SYSTEM SCIENCES, 45., 2012, Hawaii. **Anais**... Washington: IEEE, p. 2583-2591, 2012.

SONG, M.; KIM, M. C. RT^ 2M: real-time Twitter trend mining system. *In*: 2013 INTERNATIONAL CONFERENCE ON SOCIAL INTELLIGENCE AND TECHNOLOGY, 2013, State College, Pennsylvania. **Anais**... State College: IEEE, p. 64-71, 2013.

TECHNOLOGY and society: pessimism v progress. **The Economist**,18 Dec. 2019. Disponível em: <https://econ.st/2IjLO1n>. Acesso em: 14 jan. 2020.

WURMSER, Y. **Wearables 2019**: advanced wearables pick up pace as fitness trackers slow. eMarketer, 2019. Disponível em: <https://bit.ly/38jcnhJ>. Acesso em 4 nov. 2019.

BIBLIOGRAFIA COMPLEMENTAR

AGARWAL, A. *et al*. Sentiment analysis of twitter data. *In*: NAGARAJAN, M.; GAMON, M. (Eds.). Proceedings of the workshop on language in social media (LSM 2011). Portland: Association for Computational Linguistics, 2011. p. 30-38.

PAK, A.; PAROUBEK, P. **Twitter as a corpus for sentiment analysis and opinion mining**. LREc, 2010, p. 1320-1326.

WILSON, D. W. Monitoring technology trends with podcasts, RSS and Twitter. **Library Hi Tech News**, v. 25, n. 10, p. 8-12, 2008.

Ipea – Instituto de Pesquisa Econômica Aplicada

EDITORIAL

Coordenação
Reginaldo da Silva Domingos

Assistente de Coordenação
Rafael Augusto Ferreira Cardoso

Supervisão
Camilla de Miranda Mariath Gomes
Everson da Silva Moura

Revisão
Ana Clara Escórcio Xavier
Clícia Silveira Rodrigues
Idalina Barbara de Castro
Olavo Mesquita de Carvalho
Regina Marta de Aguiar
Amanda Ramos Marques (estagiária)
Hellen Pereira de Oliveira Fonseca (estagiária)
Ingrid Verena Sampaio Cerqueira Sodré (estagiária)
Isabella Silva Queiroz da Cunha (estagiária)

Editoração
Aeromilson Trajano de Mesquita
Cristiano Ferreira de Araújo
Danilo Leite de Macedo Tavares
Herllyson da Silva Souza
Jeovah Herculano Szervinsk Junior
Leonardo Hideki Higa

Capa
Danielle de Oliveira Ayres
Flaviane Dias de Sant'ana

Projeto Gráfico
Renato Rodrigues Bueno

The manuscripts in languages other than Portuguese published herein have not been proofread.

Livraria Ipea
SBS – Quadra 1 – Bloco J – Ed. BNDES, Térreo
70076-900 – Brasília – DF
Tel.: (61) 2026-5336
Correio eletrônico: livraria@ipea.gov.br

Composto em adobe garamond pro 12/16 (texto)
Frutiger 67 bold condensed (títulos, gráficos e tabelas)
Brasília-DF

Missão do Ipea
Aprimorar as políticas públicas essenciais ao desenvolvimento brasileiro por meio da produção e disseminação de conhecimentos e da assessoria ao Estado nas suas decisões estratégicas.

ISSN 1415-4765

ipea Instituto de Pesquisa Econômica Aplicada

MINISTÉRIO DA
ECONOMIA

PÁTRIA AMADA BRASIL
GOVERNO FEDERAL