

# A relevância dos algoritmos\*

## The relevance of algorithms

**Tarleton Gillespie**

Cornell University  
Ithaca, NY, Estados Unidos da America

\* Artigo publicado originalmente por Tarleton Gillespie sob o título "The relevance of algorithms", no livro *Media Technologies: Essays on Communication, Materiality, and Society* (MIT Press, 2014). Traduzido por Amanda Jurno mediante autorização do autor e da editora. Revisão: Carlos d'Andréa

## Resumo

Algoritmos (particularmente aqueles utilizados pelos mecanismos de busca, plataformas de mídia social, sistemas de recomendação e bases de dados) exercem uma função crescentemente importante em selecionar qual informação deve ser considerada mais relevante para nós, uma característica crucial da nossa participação na vida pública. À medida que assumimos as ferramentas computacionais como nossa forma primária de expressão, sujeitamos discurso e conhecimento humanos às lógicas de procedimento que sustentam a computação. Precisamos questionar os algoritmos como elementos chave de nosso sistema informacional e das formas culturais que emergem de suas sombras. Devemos ter especial atenção sobre onde e de que forma a introdução dos algoritmos nas práticas do conhecimento humano podem levar a ramificações políticas. Este artigo oferece um mapa conceitual para fazer isso. Ele propõe uma análise sociológica que não concebe os algoritmos como algo abstrato ou façanha técnica, e sim sugere como revelar as escolhas humanas e institucionais que estão por trás de suas elaborações. O objetivo é verificar como os algoritmos são convocados, recrutados e negociados como parte de esforços coletivos para conhecer e se tornar conhecido.

## Abstract

Algorithms (particularly those embedded in search engines, social media platforms, recommendation systems, and information databases) play an increasingly important role in selecting what information is considered most relevant to us, a crucial feature of our participation in public life. As we have embraced computational tools as our primary media of expression, we are subjecting human discourse and knowledge to the procedural logics that undergird computation. What we need is an interrogation of algorithms as a key feature of our information ecosystem, and of the cultural forms emerging in their shadows, with a close attention to where and in what ways the introduction of algorithms into human knowledge practices may have political ramifications. This essay is a conceptual map to do just that. It proposes a sociological analysis that does not conceive of algorithms as abstract, technical achievements, but suggests how to unpack the warm human and institutional choices that lie behind them, to see how algorithms are called into being by, enlisted as part of, and negotiated around collective efforts to know and be known.

## Palavras-chave

Algoritmos. Plataformas. Mecanismos de busca. Conhecimento. Participação.

## Keywords

Algorithms. Platforms. Search engines. Knowledge. Participation.

## Introdução

Os algoritmos desempenham um papel cada vez mais importante na seleção das informações consideradas de maior relevância para nós, um aspecto fundamental da nossa participação na vida pública. As ferramentas de busca nos ajudam a navegar em grandes bases de dados ou por toda a web. Os algoritmos de recomendação mapeiam nossas preferências em relação a outros usuários, trazendo ao nosso encontro sugestões de fragmentos novos ou esquecidos da cultura. Eles gerenciam as nossas interações em sites de redes sociais, destacando as novidades de um amigo enquanto excluem as novidades de outro.

Os algoritmos projetados para calcular o que “está em alta”, o que é “tendência” ou o que é “mais discutido” nos oferecem uma camada superficial das conversas aparentemente sem fim que estão disponíveis. Juntos, eles não só nos ajudam a encontrar informações, mas nos fornecem meios para saber o que há para ser conhecido e como fazê-lo; a participar dos discursos sociais e políticos e de nos familiarizarmos com os públicos dos quais participamos. Além disso, são hoje uma lógica central que controla os fluxos de informação dos quais dependemos, com o “poder de possibilitar e atribuir significados, gerenciando como a informação é percebida pelos usuários, a ‘distribuição do sensível’” (LANGLOIS, 2013).

Os algoritmos não são necessariamente *softwares*: em seu sentido mais amplo, são procedimentos codificados que, com base em cálculos específicos, transformam dados em resultados desejados. Os procedimentos dão nome tanto ao problema quanto aos passos pelos quais ele precisa passar para ser resolvido. Podemos considerar como algoritmos, por exemplo, instruções de navegação ou fórmulas matemáticas usadas para prever o movimento de um corpo celestial. “Algoritmos fazem coisas e sua sintaxe incorpora uma estrutura de comando para permitir que isso aconteça” (GO-

FFEY, 2008, p. 17). Podemos pensar, portanto, que os computadores são fundamentalmente máquinas algorítmicas - projetadas para armazenar e ler dados, aplicar procedimentos matemáticos de forma controlada e oferecer novas informações como resultado. Porém tratam-se de procedimentos que poderiam ser feitos manualmente - e, de fato, eram feitos (LIGHT, 1999).

Mas, à medida que adotamos ferramentas computacionais como nossos principais meios de expressão e passamos a fazer não só da matemática mas de toda a informação “digital”, passamos a sujeitar o discurso e o conhecimento humano a essas lógicas procedimentais que sustentam toda a computação. E há implicações específicas quando usamos algoritmos para selecionar o que é mais relevante a partir de um *corpus* de dados composto por rastros das nossas atividades, preferências e expressões.

Esses algoritmos, os quais chamaremos de *algoritmos de relevância pública*, estão - através dos mesmos procedimentos matemáticos - produzindo e certificando conhecimento. A avaliação algorítmica da informação, assim, representa uma *lógica de conhecimento* particular baseada em suposições específicas sobre o que é o conhecimento e como alguém deveria identificar seus componentes mais relevantes. O fato de estarmos recorrendo a algoritmos para identificar o que precisamos saber é tão marcante quanto termos recorrido aos especialistas credenciados, ao método científico, ao senso comum ou à palavra de Deus.

O que precisamos é um questionamento dos algoritmos enquanto elemento central do nosso *ecossistema informacional* (ANDERSON, 2011), e das formas culturais que emergem às suas sombras (STRIPHAS, 2010), prestando atenção especialmente em como e onde a introdução desses algoritmos nas práticas de conhecimento humano podem ter ramificações políticas. Este ensaio é um mapa conceitual exatamente pra isso. Aqui destaco seis dimensões dos algoritmos de relevância pública que têm valor político:

1. *Padrões de inclusão*: as escolhas por trás do que gera um índice, em primeiro lugar; o que é excluído; e como os dados são preparados para o algoritmo.
2. *Ciclos de antecipação*: as implicações das tentativas dos provedores dos algoritmos de conhecer a fundo e prever completamente os seus usuários; e como importam as conclusões às quais eles chegam.
3. *Avaliação de relevância*: os critérios pelos quais os algoritmos determinam o que é relevante; como esses critérios nos são ocultados; e como eles implementam escolhas políticas acerca de um conhecimento considerado apropriado e legítimo.
4. *A promessa da objetividade algorítmica*: a maneira como o caráter técnico do algoritmo é situada como garantia de imparcialidade; e como essa alegação é mantida diante de controvérsias.
5. *Entrelaçamento com a prática*: como os usuários reconfiguram suas práticas para se adequar aos algoritmos dos quais dependem; e como podem transformar algoritmos em espaços de disputa política, às vezes até mesmo para questionar as políticas do próprio algoritmo.
6. *A produção de públicos calculados*: como a apresentação algorítmica dos públicos, para eles mesmos, molda uma noção de si desse público; e quem está em melhor posição para se beneficiar desse conhecimento.

Considerando a rapidez com que mudam estas tecnologias e os usos que são feitos delas, esta lista deve ser vista como provisória e não como completa. Mas a meu ver, acredito que essas sejam as linhas de pesquisa mais importantes para compreender os algoritmos enquanto ferramentas emergentes para o conhecimento e discurso públicos.

Pode ser sedutoramente fácil interpretar tudo isso de forma errada. Na tentativa de dizer algo relevante sobre como os algoritmos transformam o discurso público, devemos resistir firmemente à tentação de tratar a tecnologia como condutora nas explicações. Ainda que recentes estudos sociológicos sobre a internet tenham se dedicado a desfazer um determinismo tecnológico simplista que contaminou trabalhos anteriores, este determinismo continua a ser uma instância analítica atraente. Uma análise sociológica não deve conceber os algoritmos como realizações técnicas abstratas, mas desvendar as escolhas humanas e institucionais que estão por trás desses mecanismos frios. Suspeito que uma abordagem mais frutífera seria nos voltarmos tanto para a sociologia do conhecimento, quanto para a sociologia da tecnologia - para observarmos como essas ferramentas são convocadas, alistadas como parte de, e negociadas em torno de esforços coletivos para conhecer e se tornar conhecido. Isso pode nos ajudar a revelar que algoritmos aparentemente sólidos são, de fato, realizações frágeis. Também deve nos lembrar que os algoritmos são hoje tecnologias de comunicação; assim como as tecnologias de transmissão e de publicação, eles são hoje "instrumentos científicos da sociedade em geral" (GITELMAN, 2006, p. 5) e estão envolvidos e influenciando as maneiras pelas quais ratificamos o conhecimento para a vida cívica, porém de maneiras mais "protocológicas" (GALLOWAY, 2004) - ou organizadas computacionalmente para dizer de outro jeito - do que qualquer outro meio anteriormente.

## Padrões de inclusão

Os algoritmos são máquinas inertes e sem sentido até serem combinados com bancos de dados para com eles funcionar. Uma pesquisa sociológica sobre um algoritmo deve sempre levar em consideração os bancos de dados aos quais ele está ligado; não fazê-lo seria o mesmo que estudar o que foi dito em um protesto público, sem

perceber que alguns dos protestantes foram barreados na entrada do parque.

Para os usuários, algoritmos e bancos de dados são conceitualmente conjugados: usuários geralmente os tratam como um único mecanismo em funcionamento. E, aos olhos do mercado, os criadores da base de dados e os provedores do algoritmo são geralmente os mesmos, ou trabalham em consonância econômica e, muitas vezes, ideológica. “Juntos, estruturas de dados e algoritmos são, para um computador, as duas metades da ontologia do mundo” (MANOVICH, 1999, p. 84). No entanto, podemos tratar os dois como analiticamente distintos: antes que os resultados possam ser fornecidos algoritmicamente, a informação deve ser coletada, preparada para o algoritmo e, às vezes, excluída ou rebaixada.

## Coleção

Vivemos em um momento histórico no qual, mais do que nunca, quase todas as atividades públicas incluem o armazenamento de extensos registros, catalogação e arquivamento de documentos - e fazemos isso ainda mais nas redes de comunicação projetadas de tal forma que cada entrada, cada página visualizada e cada clique deixe um rastro digital. Transformar tais rastros em bases de dados envolve um conjunto complexo de técnicas de informação (STALDER, MAYER, 2009): o Google, por exemplo, rastreia os sites de indexação da web e seus metadados. Ele digitaliza as informações do mundo real, de acervos de bibliotecas a imagens de satélite ou registros fotográficos de ruas da cidade; convida os usuários a fornecerem seus detalhes pessoais e sociais como parte de seu perfil no Google+; mantém registros detalhados de cada pesquisa realizada e cada resultado clicado; adiciona informações com base na localização de cada usuário; armazena os rastros das experiências de navegação na web reunidas via suas redes massivas de publicidade.

Compreender o que está incluído nesses bancos de dados requer atenção às políticas de coleta dos serviços de informação, mas também deveria se estender para além das práticas efetivamente envolvidas. Não só para detectar casos de má conduta, embora existam alguns, mas para entender como um provedor de informações pensa sobre a coleta de dados que empreende. A resistência política ao projeto *StreetView* do Google, na Alemanha e na Índia, lembra-nos que a resposta à pergunta “Como é esta esquina?” tem diferentes implicações para pessoas que querem ir até aquela esquina, para pessoas que moram lá e para pessoas que acreditam que essa informação não deveria ser disponibilizada publicamente. Essa resposta também nos revela o que o Google pensa sobre o que é “público”, uma interpretação que está sendo amplamente implantado através do serviço prestado pela empresa.

## Preparada para o algoritmo



“Dado bruto é um oxímoro” (GITELMAN, JACKSON, 2013). Os dados já vêm dissecados e persistentemente bagunçados. Contudo, existe uma ordem premeditada necessária para que os algoritmos possam funcionar com esses dados. Mais do que qualquer coisa, os algoritmos são projetados e apreciados para serem funcionalmente automáticos; para, quando acionados, agirem sem qualquer intervenção ou supervisão regular de humanos (WINNER, 1977). Isso significa que as informações dos banco de dados devem ser transformadas e institucionalizadas de tal forma que os algoritmos possam agir sobre elas automaticamente. Os dados devem ser “imaginados e enunciados em contraste com a continuidade dos fenômenos” (GITELMAN, JACKSON, 2013). Reconhecer quais são as formas pelas quais os dados são “limpos” é um importante contraponto para a aparente automaticidade dos algoritmos. Assim como é possível conhecer sobre esculturas, estudando seus moldes invertidos, podemos

entender mais sobre os algoritmos ao analisarmos de perto como as informações são orientadas para enfrentá-los, como elas são *preparadas para o algoritmo*.

Nas primeiras arquiteturas de banco de dados, a informação era organizada em hierarquias rigorosas e, como foi revelado, inflexíveis. Com o desenvolvimento das arquiteturas relacionais e orientadas aos objetos, as informações puderam ser organizadas de maneira mais flexíveis, onde é possível associar os *bits* de dados de várias formas uns com os outros, as categorias podem mudar ao longo do tempo e os dados podem ser explorados sem ser preciso navegar ou mesmo entender a estrutura hierárquica na qual eles estão arquivados. As implicações sociológicas dos designs dos bancos de dados têm sido amplamente ignoradas; mas é preciso ter em mente que os estilos das bases de dados criaram políticas, além de apenas criar ferramentas informacionais essenciais para o funcionamento dos algoritmos. Como Rieder (2012) observa, com a ampla aceitação dos bancos de dados relacionais surge uma “ontologia relacional” que entende os dados como atomizados, “objetos regulares, uniformes e apenas vagamente conectados que podem ser requisitados de maneiras potencialmente ilimitadas quando recuperadas”, deslocando assim o poder expressivo do design estrutural do banco de dados para a busca.

Porém, mesmo com esses bancos de dados mais flexíveis, a categorização continua a ser de vital importância para seu design e seu gerenciamento. O processo de categorização é uma poderosa intervenção semântica e política: quais são as categorias, o que pertence a elas e quem decide como implementá-las na prática, são todas asserções poderosas sobre como as coisas são e como elas devem ser (BOWKER; STAR, 2000). Uma vez instituída, uma categoria desenha uma demarcação que será tratada com reverência por um futuro algoritmo. Como exemplo, trazemos o incidente *#amazonfail*. Em 2009, mais de 157 mil livros *gay-frien-*

*dly*<sup>1</sup> desapareceram instantaneamente das listas de vendas do site da *Amazon* porque foram acidentalmente categorizados como “adultos”. Naturalmente, sistemas informacionais complexos como esses estão propensos a erros. Mas esse erro em particular também nos revelou que o algoritmo da *Amazon*, que calcula a “lista dos livros mais vendidos”, é instruído a ignorar livros categorizados como adultos. Ou seja, mesmo na ausência de erros, qualquer que seja o critério usado pela *Amazon* para determinar o que é considerado “adulto”, ou não, está sendo aplicado e reificado - e se torna aparente apenas na inexplicável ausência de alguns livros da lista e na presença de outros.

## Exclusão e rebaixamento

Embora os produtores de banco de dados compartilhem o apetite pela coleta de informações, eles se distinguem mais pelo que escolhem excluir.

O arquivo, ao lembrar tudo e apenas certo conjunto de fatos / descobertas / observações, consistente e ativamente se envolve no esquecimento de outros conjuntos. Assim, a força impositiva do arquivo opera através da exclusão invisível. A invisibilidade é uma característica importante nesse caso: o arquivo se apresenta como o conjunto de todas as declarações possíveis, ao invés da lei do que pode ser dito (BOWKER, 2006, p. 12-14).

Mesmo nas condições atuais de abundância digital (KEANE, 1999), em que é mais barato e mais fácil errar mantendo as informações, sempre há uma sobra.

Os sites podem, eles próprios, se recusarem a ser indexados pelos coletores de dados (como ferramentas de busca, por exemplo). Elmer (2009) revela que o *robot.txt*, um trecho de código que

1 Nota da tradutora: o termo “gay-friendly” não costuma ser traduzido para português e é usado em referência a lugares ou conteúdos que são receptivos e/ou voltados para os membros da comunidade LGBT.



previne uma página ou site de ser indexado por ferramentas de busca, apesar de ter sido projetado inicialmente como uma ferramenta para preservar a privacidade dos criadores individuais, tem sido usado por instituições governamentais como uma forma alternativa de “reter” documentos públicos da apreciação do público. Mas além da auto-exclusão, algumas informações coletadas inicialmente são subsequentemente removidas antes mesmo que um algoritmo chegue até elas. Apesar de os serviços de informação em larga-escala se gabarem por serem abrangentes, esses sites são, e sempre devem ser, censores de informações também. Dados indexados excluem *spams* e vírus; vigiam violação de direitos autorais e pornografia; e retiram o que é obsceno, condenável ou politicamente controverso das bases de dados (Gillespie no prelo).

Conteúdos ofensivos podem ser simplesmente removidos dos dados indexados, ou uma conta pode ser suspensa, antes mesmo de chegar ao conhecimento de outro usuário. Contudo, junto com o algoritmo, é possível lidar com conteúdos problemáticos de formas mais sutis. O *YouTube* “rebaixa algoritmicamente” vídeos provocantes para que não apareçam nas listas dos mais assistidos ou na página inicial gerada para novos usuários. O *Twitter* não censura conteúdos profanos de tuítes públicos, mas ele os remove da avaliação algorítmica que determina quais termos figuram nos “*Trending*”.

Os padrões específicos pelos quais as informações são excluídas ou incluídas nos bancos de dados, e depois gerenciadas de maneiras particulares, são reminiscências dos debates do século XX (TUSHNET, 2008) sobre as formas em que as decisões feitas pela mídia comercial sobre o que deve ser sistematicamente deixado de fora e quais categorias de fala simplesmente não se qualificam podem moldar a diversidade e o caráter do discurso público. Essas escolhas, sejam feitas por um editor de jornal ou pelos mecanismos de indexação de uma ferramenta de pesquisa, ajudam a estabelecer e a confirmar padrões

de debate viável, legitimidade e decoro. Mas neste caso, os algoritmos podem ser apresentados como automáticos, já que são os padrões de inclusão nos bancos de dados que pré-determinam o que vai ou não aparecer entre seus resultados.

## Ciclos de antecipação

Algoritmos de busca determinam o que oferecer com base nas informações do usuário. Mas a maioria das plataformas hoje faz do seu negócio saber mais, muito mais, sobre o usuário do que apenas a pesquisa que ele acabou de fazer. Os sites esperam ser capazes de antecipar o usuário quando o algoritmo é acionado, o que requer tanto o conhecimento coletado naquele momento, quanto o conhecimento já acumulado sobre o usuário e sobre outros usuários considerados parecidos com ele em termos estatísticos e demográficos (BEER, 2009) - unindo o que Stalder e Mayer (2009) chamam de “segundo índice”. Se podemos dizer que as emissoras de TV e Rádio forneciam não só conteúdo para as audiências, mas também audiências para os anunciantes (SMYTHE, 2001), podemos dizer também que os provedores digitais não apenas fornecem informações para os usuários, mas usuários para seus algoritmos que são feitos e refeitos a cada momento de uso, porque cada clique e cada busca incrementam a ferramenta.

Muitos dos estudos sobre coleta de dados e práticas de rastreamento dos provedores de informação contemporâneos têm focado em questões importantes sobre privacidade provocadas por eles. Zimmer (2008) argumenta que as ferramentas de busca não só pretendem incessantemente indexar a web mas também desenvolver uma “coleta perfeita” de tudo relativo aos seus usuários. Para fazer isso, os provedores de informação precisam não só rastreá-los, mas construir infraestruturas técnicas e modelos de negócio que conectem sites individuais em um pacote de serviços (como as várias ferramentas e serviços do *Google*) ou em um ecossiste-

ma ainda mais amplo (como com o “social graph” do Facebook e seus botões de “curtir” espalhados ao longo pela web) para então criarem incentivos para que os usuários permaneçam neles. Isto permite que o provedor seja “passivo-agressivo” (BERRY, 2012) na forma como agrega as informações coletadas em vários sites e as transforma em um perfil coerente e cada vez mais detalhado do usuário. Os provedores também se beneficiam do caráter cada vez mais participativo da web, em que usuários são encorajados a oferecer todo tipo de informações sobre si mesmos e a se sentirem empoderados ao fazê-lo. À medida que nossas micro-práticas migram cada vez mais para essas plataformas, torna-se sedutor para os provedores de informação (apesar de não ser compulsório) rastrear essas atividades e transformá-las em mercadoria de várias formas (GILLESPIE; POSTIGO, 2012). Inclusive, os usuários podem não estar cientes de que suas atividades estão sendo rastreadas pelos maiores anunciantes online, em toda a web, e têm pouco ou nenhum meio de contestar este acordo mesmo que o saibam (TUROW, 2012).

Contudo, privacidade não é a única questão politicamente relevante nesse caso. Nestes ciclos de antecipação, os bits de informação são mais fáceis de serem lidos pelo algoritmo, e, portanto, tendem a substituir os usuários. O Facebook sabe muito sobre seus usuários; mas, mesmo assim, só sabe o que é possível saber. As informações mais conhecidas (geolocalização, plataforma computacional, informações de perfil, amigos, atualizações de status, links seguidos no site, tempo no site, atividade em outros sites que hospedam o botão de “curtir” ou cookies) são uma interpretação do usuário, um “dossiê digital” (SOLOVE, 2004) ou uma “identidade algorítmica” (CHENEY-LIPPOLD, 2011) que é imperfeita, mas suficiente. O que é menos legível, ou que não pode ser conhecido sobre os usuários, é deixado de lado ou é aproximado de forma grosseira por esses algoritmos. Como Balka (2011) descreveu, os sistemas de informação produzem “corpos

de sombra” ao enfatizar alguns aspectos dos seus sujeitos e ignorar outros. Esses corpos de sombra persistem e se proliferam através dos sistemas de informação, e as diferenças que separam o usuário antecipado do próprio usuário que é representado podem ser politicamente problemáticas ou politicamente produtivas.

Mas os algoritmos não só fazem previsões exaustivas sobre os usuários; às vezes, eles fazem aproximações suficientes. Talvez tão importante quanto a constante vigilância dos usuários são as conclusões que os provedores querem elaborar com base em relativamente pouca informação sobre eles. *Hunch.com*, um serviço de recomendação de conteúdo, gabou-se por conseguir saber as preferências de um usuário com 80-85 por cento de precisão, com base nas respostas a apenas cinco perguntas. Apesar de reduzir radicalmente a complexidade de uma pessoa a cinco pontos em um gráfico, o que é importante é que esta é uma precisão suficiente para os seus propósitos<sup>2</sup>. Dado que tais sites se sentem confortáveis em oferecer caricaturas dos usuários, as questões que parecem nos classificar de forma suficiente para eles, particularmente em relação às nossas preferências enquanto consumidores, tendem a ganhar relevância como formas de medir o público. E, em certa medida, nós somos convidados a nos formalizar nessas categorias. Diante desses provedores, somos encorajados a escolher nos menus que eles oferecem uma opção para que sejamos corretamente antecipados pelo sistema e para que ele nos ofereça as informações certas, as recomendações certas, as pessoas certas.

Além de conhecer detalhes pessoais e demográficos sobre cada usuário, os provedores de informação realizam uma grande pesquisa para tentar entender, e depois operacionalizar, como

2 Nota do autor: ZUCKERMAN, Ethan. 26 maio 2011. Eli Pariser talks about the filter bubble. In: The Boston Phoenix. Disponível em: <http://thePhoenix.com/Boston/arts/121405-eli-pariser-talks-about-the-filter-bubble/>. Acesso em: 22 abr. 2013.



os humanos normalmente procuram, se engajam e compreendem as informações. Mais notavelmente no estudo das interações humano-computador (IHC), o entendimento da psicologia e da percepção humana é aplicado ao design dos algoritmos e às formas pelas quais os seus resultados deveriam ser representados aos usuários. Os designers esperam ser capazes de prever as tendências e capacidades psicofisiológicas dos usuários, além de seus hábitos e preferências. Mas, também nessas antecipações, valores implícitos e, muitas vezes, políticos podem estar inscritos na tecnologia: os hábitos de percepção ou interpretação de alguns usuários são tomados como modelo universal, hábitos contemporâneos são imaginados como atemporais, objetivos computacionais particulares são tratados como auto-evidentes.

Também estamos presenciando um novo tipo de poder informacional, reunido nessas enormes bases de dados sobre as atividades e as preferências dos usuários, que está ele mesmo modificando a paisagem política. Além das suas técnicas, os provedores de informação que acumulam esses dados, as indústrias terceirizadas que reúnem e compram dados de usuários como mercadoria, e aqueles que transacionam dados sobre os usuários por quaisquer outros motivos (ou seja, empresas de cartão de crédito) têm por essa razão uma voz mais forte tanto no mercado, quanto nos corredores do poder legislativo, e estão cada vez mais se envolvendo nos debates políticos sobre a proteção do consumidor e direitos digitais. Estamos vendo a implantação da mineração de dados nas arenas da organização política (HOWARD, 2005), no jornalismo (ANDERSON, 2011) e na publicação de conteúdo (STRIPHAS, 2009), onde os segredos extraídos das quantidades massivas de dados dos usuários são tomados como diretrizes convincentes para a produção de conteúdo no futuro - seja a próxima campanha micro-segmentada ou o próximo fenômeno pop.

## A avaliação da relevância

Quando os usuários clicam em “Pesquisar”, carregam seu “Feed de Notícias” do Facebook, ou pedem recomendações do Netflix, os algoritmos precisam identificar instantânea e automaticamente quais, dos trilhões de bits de informação, melhor atendem os critérios em questão, e que vão melhor satisfazer um usuário específico e seus objetivos presumidos. Embora estes cálculos nunca tenham sido simples, eles se tornaram mais complexos à medida que o uso público desses serviços foi amadurecendo. Os algoritmos de pesquisa, por exemplo, eram antes baseados apenas em dizer qual a frequência com que determinados termos de pesquisa apareciam nas páginas indexadas da web. Agora, esses algoritmos incorporam informações contextuais sobre os sites e sobre onde eles estão hospedados; consideram a frequência e como o site é relacionado por outros; e convocam técnicas de processamento de linguagem natural para melhor “entender” tanto a consulta, quanto os recursos que o algoritmo pode oferecer como resposta. De acordo com o Google, seu algoritmo de pesquisa examina mais de duzentos indicadores para cada consulta feita no site<sup>3</sup>.

Esses indicadores são os meios pelos quais o algoritmo estima a “relevância” de um conteúdo. É nessa questão que os sociólogos dos algoritmos devem fincar os pés: “relevante” é um julgamento fluido e carregado de significados, tão aberto à interpretação quanto alguns dos termos avaliativos

3 Nota do autor: GOOGLE. 2013. Facts about Google and competition. In: Google. Disponível em: <http://www.google.com/competition/howgooglesearchworks.html>. Acesso em: 22 abr. 2013.

O Google e o Bing já se envolveram em uma guerra de competição por “indicadores”, primeiro quando o Bing anunciou que usava 1.000 indicadores e o Google seguindo com a informação que seus 200 indicadores tinham até 50 variações, levando seu total a quase 10.000. Veja: SULLIVAN, Danny. 11 nov. 2010. Dear Bing we have 10,000 ranking signals to your 1,000. Love, Google. In: Search Engine Land. Disponível em: <https://searchengineland.com/bing-10000-ranking-signals-google-55473>. Acesso em 22 abr. 2013.

sobre os quais estudiosos da mídia já se debruçaram, como “notável” ou “popular”. Como não há uma métrica independente capaz de dizer quais seriam *realmente* os resultados de pesquisa mais relevantes para qualquer busca, os engenheiros precisam decidir quais resultados parecem “corretos” e ajustar seu algoritmo para atingir esse resultado, ou fazer alterações com base em evidências dos usuários, considerando cliques rápidos e pesquisas sem continuação como uma aproximação, não de relevância exatamente, mas de satisfação com o conteúdo oferecido. Acusar um algoritmo de ser enviesado implica considerar que exista um julgamento imparcial de relevância, o qual a ferramenta estaria falhando em alcançar. Uma vez que tal medida não está disponível, as disputas sobre as avaliações algorítmicas não têm um terreno sólido para se basear.

## Critérios

Para ser possível dizer que um determinado algoritmo faz suposições avaliativas, do tipo que tem consequências para o esforço do conhecimento humano, seria preciso fazer uma análise crítica do algoritmo a fim de questionar seus critérios subjacentes. Porém, em quase todos os casos, esses critérios de avaliação estão ocultos e devem permanecer assim. O algoritmo de “*Trendings*” do Twitter, que informa ao usuário sobre os termos mais discutidos naquele momento em uma determinada área geográfica, deixa inclusive a definição do que seria “mais discutido” sem especificação. Os critérios que eles usam para medir o que é “mais discutido” são descritos apenas em termos gerais: a velocidade do surgimento de um determinado termo; se ele já apareceu na lista “*Trendings*” do Twitter antes; e se ele circula dentro de um determinado grupo ou se expande através de vários grupos de usuários. O que não é declarado pela empresa é como esses critérios são medidos; como são avaliados em relação uns

aos outros; quais outros critérios também foram incorporados à avaliação; e se algum desses critérios foi substituído. Isso faz com que os algoritmos estejam eternamente abertos à suspeita dos usuários de que seus critérios seriam enviesados para favorecer os provedores comerciais ou politicamente, ou que incorporariam premissas embutidas e não-verificáveis que agem de forma incompreensível, até mesmo para os designers (GILLESPIE, 2012).

Um provedor de informações como o Twitter não pode ser muito mais explícito ou preciso quanto ao funcionamento de seus algoritmos. Fazê-lo seria dar meios para que seus concorrentes duplicassem e superassem seu serviço facilmente, além de exigir uma explicação mais técnica do que a maioria dos usuários está preparada para receber. Isso prejudicaria a capacidade de mudança dos critérios conforme necessário, mas acima de tudo, forneceria um roteiro para quem quisesse “burlar o sistema” e colocar seus sites no topo dos resultados de pesquisa ou fazer suas *hashtags* aparecerem na lista de “*Trendings*”. Embora alguns sites de recomendações colaborativas, como o Reddit, tenham tornado públicos seus algoritmos de classificação das histórias e comentários dos usuários, eles precisam estar constantemente procurando e corrigindo casos de “desvalorização” organizada pelos usuários, e essas táticas não podem ser tornadas públicas. Com poucas exceções, a tendência é mesmo que os provedores sejam pouco claros<sup>4</sup>.



## Objetivos comerciais

Uma segunda abordagem poderia implicar na consideração cuidadosa dos contextos econômicos e culturais dos quais o algoritmo veio.

4 Nota do autor: De acordo com a já mencionada possibilidade de um algoritmo ser perfeitamente transparente, existe uma gama de escolhas abertas para o desenvolvedor sobre o quão evidente ser. Pode ser simples como decidir ser mais franco na caracterização da ferramenta; fornecer uma explicação para a razão pela qual certos anúncios são oferecidos junto a uma página; ou fornecendo documentações mais criteriosas do site.

Qualquer sistema de conhecimento emerge em meio aos objetivos econômicos e políticos da disponibilização de informações, e será moldado pelos objetivos e estratégias das poderosas instituições que buscam capitalizá-lo (HESMON-DHALGH, 2006). As pressões enfrentadas pelas ferramentas de busca, plataformas de conteúdo e provedores de informações podem moldar sutilmente o design do próprio algoritmo e a apresentação de seus resultados (VAINDHYANATHAN, 2011). À medida que passa a se apresentar como uma lógica de conhecimento legítima, novos esforços comerciais são feitos sob medida para o algoritmo (por exemplo, a otimização dos sites para aparecerem nas ferramentas de busca - em inglês, SEO, *search engine optimization*), reificando escolhas feitas e forçando outras adicionais.

Por exemplo, as primeiras críticas se preocupavam se as ferramentas de busca iriam oferecer propagandas como *links* ou conteúdo em destaque, apresentadas como resultado de cálculos algorítmicos. A rápida e clara rejeição pública a essa manobra demonstrou quão forte é nossa crença nesses algoritmos: os usuários não queriam que o conteúdo que os provedores pagassem para serem exibidos fosse misturado com o conteúdo selecionado algorítmicamente. Mas a preocupação agora é multidimensional: a paisagem do “Feed de Notícias” do Facebook, por exemplo, não pode mais ser descrita como dois territórios distintos, um social e um comercial. Em vez disso, entrelaça os resultados dos cálculos algorítmicos (quais atualizações de status e outras atividades dos seus amigos devem ser listadas no Feed, quais *links* serão recomendados para esse usuário, quais amigos estão ativos no site naquele momento), elementos estruturais (ferramentas para contribuir com uma atualização de status ou para comentar em um elemento informativo, *links* para grupos e páginas) e elementos colocados ali com base em relações de patrocínio (*banners* publicitários, aplicativos de sites de terceiros). Para mapear este terreno complexo, é necessário compreender

a fundo as relações econômicas e os pressupostos sociais que ele representa.

## Premissas epistemológicas

Por fim, devemos considerar se os critérios avaliativos do algoritmo são estruturados por princípios políticos ou organizacionais específicos que têm eles mesmos ramificações políticas. Não é apenas se um algoritmo pode ser parcial para este ou aquele fornecedor, ou pode favorecer seus próprios interesses comerciais sobre outros. É uma questão sobre a importância das premissas filosóficas sobre o conhecimento relevante nas quais o algoritmo se baseia. Alguns estudos que analisaram vieses na seleção das ferramentas de busca (em ordem de publicação, INTRONA; NISSENBAUM, 2000, HALAVAIS, 2008, ROGERS, 2009, GRANKA, 2010) observaram algumas tendências estruturais em retornar como resultado o que já era popular, sites de língua inglesa e fornecedores de informações comerciais. Ao debater sobre o que significaria exigir neutralidade nos resultados de pesquisa, os estudiosos jurídicos (GRIMMELMANN, 2010, PASQUALE; BRACHA, 2008) se referiam a muito mais do que oferecer dicas sobre parceiros comerciais nos resultados. Os critérios que os algoritmos de informação pública levam em consideração são miríades; cada um está equipado com um patamar que define o que será empurrado para os resultados, como um resultado será posicionado acima de outro e assim por diante.

Dessa forma, as avaliações realizadas pelos algoritmos dependem sempre de pressupostos inscritos nos seus códigos que versam sobre o que e como importa ser identificado por eles. Quando uma simples ferramenta de pesquisa conta o número de vezes que um termo pesquisado aparece nas páginas da web indexadas, ela está reificando uma lógica particular que assume que as páginas que incluem o termo pesquisado provavelmente são relevantes para alguém interessado neste termo. Quando o Google desenvolveu o

PageRank, que determina o valor de uma determinada página com base no número de *links* em referência a essa página, ele construiu uma lógica diferente: uma página linkada em vários lugares na *web*, em sites considerados de alta qualidade, é vista como sendo “ratificada” por outros usuários e, portanto, têm maior probabilidade de ser relevante para este usuário também. Para Finkelstein, ao preferir *links* advindos de sites vistos como sendo de alta qualidade, o Google mudou de uma abordagem mais populista para uma “democracia de acionistas”.

Um *link* não é um voto, mas tem influência proporcional ao poder relativo do eleitor (em termos de popularidade). Como os blocos de interesses comuns, ou facções sociais, podem afetar os resultados de uma pesquisa em um grau que depende do seu peso relativo na rede, os resultados do cálculo algorítmico das ferramentas de busca refletem lutas políticas da sociedade (FINKELSTEIN, 2008, p.107).

Quando um site de discussão de notícias decide qual a proporção de queixas negativas para o número de visualizações é suficiente para justificar e esconder automaticamente um tópico de comentários, essa proporção representa a análise do *site* sobre a volatilidade do discurso público, ou pelo menos a volatilidade que o *site* prefere para a comunidade de usuários que eles acham que têm (BRAUN, 2011). Muita experiência e julgamentos podem ser incorporados a esses artefatos cognitivos (HUTCHINS 1995, LATOUR, 1987), mas são julgamentos que são logo submergidos e automatizados.

A maioria dos usuários não se debruça sobre os critérios algorítmicos e tende a tratá-los como ferramentas não problemáticas a serviço de uma atividade maior: encontrar uma resposta, resolver um problema, entreter. No entanto, embora a tecnologia possa ser igualmente “encaixapretada” (LATOUR 1987, PINCH; BIJKER, 1984) por designers e usuários, isso não deveria nos levar a crer que ela permaneça estável. Na

verdade, os algoritmos podem ser facilmente, instantaneamente, radicalmente e invisivelmente alterados. Embora grandes atualizações possam acontecer apenas ocasionalmente, os algoritmos estão regularmente sendo “ajustados”. As mudanças podem ocorrer sem que a interface com o algoritmo mude minimamente: o “Feed de Notícias” e a barra de pesquisa podem parecer os mesmos de ontem, ainda que as avaliações que acontecem por baixo deles tenham sido completamente refeitas. A metáfora da caixa-preta não nos ajuda nesse caso, já que o funcionamento do algoritmo é obscuro e maleável, “provavelmente tão dinâmico que uma fotografia instantânea nos daria poucas chances de acessar seus vieses” (PASQUALE, 2009). Na verdade, o que nós nos referimos enquanto um algoritmo muitas vezes não é apenas um, mas são muitos algoritmos. Ferramentas de busca como o Google participam regularmente de testes “A/B”<sup>5</sup>, apresentando *rankings* diferentes para diferentes subconjuntos de usuários para obter dados sobre a velocidade e satisfação dos clientes em tempo real. As preferências encontradas nos resultados do teste são, então, incorporadas aos ajustes da ferramenta em uma atualização posterior.

Cada algoritmo tem como premissa tanto um pressuposto sobre a análise adequada de relevância, quanto em uma representação desse pressuposto em uma técnica de avaliação (computacional). Podem existir premissas implícitas incorporadas na ideia de relevância de um site, podem existir atalhos incorporados em sua representação técnica dessa idéia, e podem existir fricções entre elas duas.

## A promessa da objetividade algorítmica

Mais do que meras ferramentas, os algoritmos também são estabilizadores da confiança, garantias práticas e simbólicas de que suas avaliações são justas e precisas, livres de subjetividade, erro

<sup>5</sup> Nota do autor: CHRISTIAN, Brian. 25 abr. 2012. The A/B Test: Inside the technology that's changing the rules of business. In: Wired.com. Disponível em: [https://www.wired.com/2012/04/ff\\_abtesting/](https://www.wired.com/2012/04/ff_abtesting/). Acesso em: 22 abr. 2013.

ou tentativas de influência. Mas, embora os algoritmos possam parecer automáticos e intocáveis pelas intervenções de seus provedores, esta é uma ideia cuidadosamente elaborada. “As ferramentas de busca se gabam por serem automatizadas, exceto quando não o são” (GRIMMELMANN, 2008, p. 950). Na verdade, nenhum serviço de informações pode ser completamente isento de interferência humana ao entregar informações: embora um algoritmo possa avaliar qualquer site como o mais relevante para sua busca, esse resultado não aparecerá se for pornografia infantil, não aparecerá na China se for um discurso político dissidente, e não aparecerá na França se promover o nazismo. No entanto, é muito importante para os provedores desses algoritmos que eles pareçam ser isentos de interferência humana. A legitimidade desses mecanismos deve ser performada junto à disponibilização da própria informação.

As articulações oferecidas pelo provedor do algoritmo junto à ferramenta servem para proporcionar o que Pfaffenberger (1992) chama de “controle logonômico”; ou seja, para definir a ferramenta entre as práticas dos usuários e concedê-la uma legitimidade que é levada às informações fornecidas e, de quebra, ao provedor. A cuidadosa articulação de um algoritmo como imparcial (mesmo quando essa caracterização é mais ofusca do que a explicativa) o certifica como um ator sociotécnico confiável concede relevância e credibilidade aos seus resultados e mantém a aparente neutralidade do provedor em face as milhões de avaliações que faz. Esta articulação do algoritmo é tão crucial para a sua manutenção social, quanto seu design material e suas obrigações econômicas.

Em grande parte, cabe ao provedor descrever seu algoritmo como tendo uma forma particular; tendo, portanto, um certo conjunto de valores; e, assim, conferindo-lhe um certo tipo de legitimidade. Isso inclui caracterizar cuidadosamente a ferramenta e o seu valor para várias audiências, às vezes de várias formas diferentes: um algoritmo pode ser

defendido como uma ferramenta de avaliação imparcial para os críticos dos seus resultados e, ao mesmo tempo, ser prometido como uma ferramenta para a promoção seletiva e para anunciantes em potencial (GILLESPIE, 2010). Como observa Mackenzie (2005), esse processo requer mais que uma descrição única e enfática: depende da “repetição e citação” (*idem*, 2005, p. 81) e, ao mesmo tempo, requer que se encubra o “conjunto autoritário de práticas” (*idem*, 2005, p. 82) que dá força ao algoritmo. Quando um provedor de informações se vê criticado pelos resultados que fornece, a legitimidade do seu algoritmo deve ser reparada tanto discursiva, quanto tecnicamente. E os usuários são cúmplices nisso:

uma sociedade obcecada pelos principais resultados do Google Notícias fez com que esses resultados se tornassem importantes, e seria insensato supor o contrário (que as pessoas estão obcecadas com os resultados porque eles são importantes) sem uma descrição do porquê o algoritmo é superior a, por exemplo, ‘o julgamento sobre notícias’ dos editores da mídia tradicional. (PASQUALE, 2009).

Esta articulação se dá primeiro na apresentação da ferramenta, na sua implantação dentro de um serviço de informação mais amplo. Chamá-los de “resultados”, “melhores”, “principais histórias” ou “tendências” diz não só sobre o que o algoritmo está medindo, mas também sobre o que deve ser entendido como medida. Uma parte igualmente importante deste trabalho discursivo está na descrição de como o algoritmo funciona. O que pode parecer uma explicação clara sobre que acontece por trás dos bastidores pode não ser realmente uma visada do verdadeiro processo que ali se dá, mas de “bastidores performados” (HILGARTNER, 2000) e cuidadosamente elaborados a fim de legitimar ainda mais esse processo e seus resultados. A descrição do sistema PageRank do Google, o elemento mais antigo do seu complexo algoritmo de busca, foi publicada primeiro em um artigo científico (já sendo essa



uma entrega elaborada sobre seus funcionamentos matemáticos), mas foi mitificada em seguida - como a característica definidora da ferramenta, como o elemento central que fez o *Google* se tornar destaque frente a seus concorrentes e como uma lógica computacional fundamentalmente democrática - mesmo quando o algoritmo estava sendo redesenhado para considerar centenas de outros critérios no processo de busca.

Acima de qualquer coisa, os provedores dos algoritmos de informações devem assegurar que seus algoritmos são imparciais. A performance da objetividade algorítmica tornou-se fundamental para a manutenção dessas ferramentas como mediadoras legítimas do conhecimento relevante. Nenhum provedor tem sido mais inflexível sobre a neutralidade do seu algoritmo quanto o *Google*, que responde regularmente às solicitações de alteração nos resultados de busca reivindicando que o algoritmo não deve ser adulterado. A empresa celebroumente se retirou completamente do mercado chinês, em 2010, ao invés de censurar seus resultados, embora tenha consentido com as regras da China antes, e pode ter se retirado para não ter de admitir que estava perdendo para as concorrentes chinesas. Mesmo apesar dessa postura, o *Google* alterou seus resultados quando surgiram reclamações sobre uma imagem de Michelle Obama, editada de forma racista, aparecendo no topo dos resultados de busca de Imagens. A empresa também fornece um mecanismo de “busca segura” para não exibir palavrões e imagens sexuais para menores, e se recusa a auto-completar pesquisas que indiquem serviços de troca de arquivos em formato torrent. No entanto, o *Google* afirma regularmente que não altera seus índices, nem manipula seus resultados. Morozov (2011) acredita que esta é uma maneira de desviar a responsabilidade:

a suspensão espiritual do *Google* para a ‘neutralidade algorítmica’ trai o crescente desconforto da empresa em ser o gatekeeper de informações

mais importante do mundo. Seus fundadores preferem tratar a tecnologia como uma força autônoma e totalmente objetiva ao invés de passar noites sem dormir, preocupando-se com vieses inerentes em como seus sistemas operam - sistemas estes que cresceram de forma tão complexa que nenhum engenheiro no *Google* os entende completamente. (MOROZOV, 2011)

Essa afirmação da objetividade algorítmica desempenha, em muitos aspectos, um papel equivalente à norma de objetividade no jornalismo ocidental. Assim como os motores de busca, os jornalistas desenvolveram táticas para determinar o que é mais relevante, como reportá-lo e como garantir sua relevância - um conjunto de práticas que são relativamente invisíveis para sua audiência; uma meta que eles admitem ser mais difícil de seguir do que parece; e um princípio que ajuda a deixar julgamentos de valor e políticas pessoais de lado, mas não os elimina. Essas práticas institucionalizadas são movidas por uma promessa conceitual que, no discurso do jornalismo, é regularmente expressa (ou exagerada) como uma espécie de totem. Os jornalistas usam a norma de objetividade como um “ritual estratégico” (TUCHMAN, 1972), para dar legitimidade pública às táticas de produção do conhecimento que são inerentemente precárias. “Estabelecer jurisdição sobre a capacidade de analisar objetivamente a realidade é uma reivindicação de um tipo especial de autoridade” (SCHUDSON; ANDERSON, 2009, p. 96).

Objetividades jornalísticas e algorítmicas não são a mesma, de forma alguma. A objetividade jornalística depende de uma promessa institucional de adequada diligência, que é incorporada e transmitida através de um conjunto de normas que os jornalistas aprendem em treinamento e no trabalho. As escolhas feitas pelos jornalistas representam uma competência apoiada em um compromisso profundamente infundido, filosófico e profissional para deixar de lado seus próprios preconceitos e convicções políticas. Já a promessa do algoritmo



se baseia muito menos em normas institucionais e competências adquiridas, e mais em uma promessa tecnicamente influenciada de neutralidade mecânica. Quaisquer que sejam as escolhas feitas, elas são apresentadas como sendo tanto livres de intervenção das mãos humanas, quanto submersas no frio funcionamento da máquina.

Mas em ambas, a legitimidade depende da acumulação de diretrizes para a *proceduralização* da seleção de informações. Os discursos e práticas de objetividade passaram a servir como uma regra constitutiva do jornalismo (RYFE, 2006). A objetividade faz parte do entendimento dos jornalistas sobre si próprios e sobre o que significa ser um jornalista. Participa da avaliação do seu trabalho, por editores, colegas e leitores, e é um sinal determinante do que os próprios jornalistas reconhecem como conteúdo jornalístico. A promessa da objetividade algorítmica também foi nitidamente incorporada nas práticas de trabalho dos provedores de algoritmos, definindo o que se constitui enquanto função e propósito de um serviço de informações. Quando o Google inclui no seu manifesto de “Dez coisas que sabemos ser verdade” que “Nossos usuários confiam em nossa objetividade e nenhum ganho de curto prazo poderia justificar a violação dessa confiança”, não se trata nem de artimanha, nem de um “Maria-vai-com-as-outras”<sup>6</sup> corporativo. É uma compreensão profundamente enraizada do caráter público do serviço de informação do Google que influencia e legitima muitos dos seus projetos técnicos e comerciais, e ajuda a obscurecer a confusa realidade sobre o serviço que a empresa presta.

Ainda assim, essas reivindicações devem competir com outras articulações no diálogo público que podem ser favoráveis, ou não, aos objetivos econômicos e ideológicos das partes interessadas. Bijker

(1997) chama de “quadros tecnológicos” concorrentes essas caracterizações discursivas de uma tecnologia feitas por grupos de atores que também têm interesse na operação, no significado e no valor social dessa tecnologia. O que os usuários de um algoritmo de informação o consideram, sejam eles astutos ou ignorantes, é importante. A forma como a imprensa retrata tais ferramentas vai fortalecer ou prejudicar os cuidadosos esforços discursivos dos seus fornecedores. Isso significa que, embora o algoritmo pareça possuir em si uma aura de neutralidade tecnológica, ou incorporar ideais populistas e meritocráticos, a sua aparência não depende apenas do seu *design*, mas também das realidades mundanas dos ciclos de notícias, assessorias de imprensa, *blogs* de tecnologia, discussões de fãs, revoltas de usuários e maquinações dos concorrentes do seu provedor.

Existe um paradoxo fundamental na articulação dos algoritmos. A objetividade algorítmica é uma declaração importante para um provedor, particularmente de algoritmos que fornecem informações vitais e voláteis para o consumo público. Articular o algoritmo como uma intervenção claramente técnica ajuda um provedor de informações a responder acusações de viés, erro e manipulação dos resultados. Ao mesmo tempo, como pode ser visto com o *PageRank* do Google, existe um valor sociopolítico no destaque do populismo dos critérios que o algoritmo usa. Afirmar que um algoritmo de um determinado site é uma representação democrática da opinião coletiva em toda a web lhe confere autoridade. E há um valor comercial na alegação de que o algoritmo retorna resultados “melhores” do que os dos seus concorrentes, o que coloca a satisfação do consumidor acima da noção de exatidão (VAN COUVERING, 2007). Ao examinar a articulação de um algoritmo, devemos prestar atenção especial à forma como esta tensão, entre neutralidade tecnicamente assegurada e o sabor social da avaliação que está sendo feita, é gerenciada - e, às vezes, como ela se rompe.

<sup>6</sup> Nota da tradutora: a expressão usada pelo autor é “corporate Kool-Aid”. Em inglês, a expressão “Beber o Kool-Aid” é usada em referência a alguém que segue uma ideia condenada ou perigosa por pressão dos colegas. Assim, escolhemos traduzi-la como “Maria-vai-com-as-outras” que tem um significado parecido em português.

## Entrelaçamento com a prática

Apesar de poderem ser estudados como ferramentas computacionais abstratas, os algoritmos são construídos para serem incorporados às práticas do mundo, que produzem as informações que eles processam, e ao mundo vivido pelos seus usuários (COULDRY, 2012). Isso se dá principalmente quando o algoritmo é o instrumento de um negócio que tem como produto a informação que entrega (ou as propagandas a que se associa). Se os usuários falharem ou se recusarem a se encaixar nas suas práticas e a conceder significado a elas, o algoritmo irá falhar. Isso significa que não devemos considerar seus “efeitos” nas pessoas, mas um “entrelaçamento” multidimensional entre algoritmos postos em prática e as táticas dos usuários que fazem face a eles. Claro, essa relação é um alvo em movimento, porque os algoritmos mudam, assim como as populações de usuários e as atividades com as quais se deparam. Ainda assim, isso não deveria significar que não exista relação entre eles. À medida que esses algoritmos se abrigam na vida das pessoas e nas suas rotinas informacionais, os usuários moldam e rearticulam os algoritmos com os quais se deparam. Os algoritmos também afetam a maneira que as pessoas procuram informações, como elas percebem e pensam sobre os horizontes de conhecimento, e como elas se compreendem no e pelo discurso público.

É importante conceber este entrelaçamento não como uma influência unidirecional, mas como um ciclo recursivo entre os cálculos do algoritmo e os “cálculos” das pessoas. O algoritmo que ajuda os usuários a navegarem pelas fotos do *Flickr* é construído no arquivo de fotos postadas, o que significa que ele é projetado para compreender e refletir as escolhas feitas pelos fotógrafos. O que as pessoas fazem e não fotografam já é um tipo de cálculo avaliativo, embora seja histórico, multivalente, contingente e sociologicamente fundamentado. Mas estes não eram os únicos impulsionadores do de-

sign do *Flickr*; a sensibilidade para as práticas fotográficas teve que competir com custos, eficiência técnica, obrigações legais e imperativos comerciais. Além disso, a população dos usuários do *Flickr* e os tipos de fotos que publicava mudaram à medida que o site cresceu em popularidade; quando foi forçado a competir com o *Facebook*; quando introduziu preços diferenciados; quando foi comprado pelo *Yahoo*, e assim por diante.

Vários usuários do *Flickr* postam fotos com o propósito claro de que elas sejam vistas: alguns são fotógrafos profissionais procurando emprego, alguns são pessoas procurando comunidades de amadores semelhantes a elas, outros estão simplesmente orgulhosos do seu trabalho. Assim, tal como o algoritmo precisa ser sensível aos fotógrafos, eles têm interesse em serem sensíveis para o algoritmo, sabendo que aparecer como resultado das buscas certas pode colocar suas fotografias diante das pessoas certas. Assim como a ênfase de *Hollywood* em gêneros específicos convida os roteiristas a escrever de maneira genérica<sup>7</sup>, o algoritmo do *Flickr* pode induzir reorientações sutis nas práticas dos fotógrafos em direção à sua lógica construída, ou seja, induzir os fotógrafos a buscar fotografar de modo a aderir a certas categorias emergentes ou a orientar sua escolha de tema e composição em direção a coisas que o algoritmo parece privilegiar. “Deixamos rastros não sobre a forma como fomos, mas sobre uma negociação tácita entre nós e os nossos auditores imaginados” (BOWKER, 2006, p. 6-7).

## Algoritmicamente identificável

Essa negociação tácita consiste, em primeiro lugar, na reorientação cotidiana e estratégica das práticas que muitos usuários realizam, perante uma ferramenta que eles sabem que pode ampliar seus esforços. Há, para os produtores de informações, um estímulo poderoso e compreensível para

<sup>7</sup> Nota do autor: Christian Sandvig, comunicação pessoal.

tornar seu conteúdo, e eles próprios, identificáveis para um algoritmo. Toda uma indústria, de otimização para mecanismos de busca (SEO - *Search Engine Optimization*, em inglês), promete impulsionar os sites para o topo dos resultados de pesquisa. Mas podemos pensar na otimização (deliberada, profissional) como apenas a linha de frente de um processo muito mais variado, orgânico e complexo pelo qual produtores de conteúdo de todos os tipos se orientam perante os algoritmos. Quando usamos *hashtags* em nossos *tuítes* - uma inovação criada pelos usuários e adotada posteriormente pelo *Twitter* - não estamos apenas nos juntando a uma conversa ou esperando ser lidos por outros, estamos redesenhando nossa manifestação para ser melhor reconhecida e distribuída pelo algoritmo de busca do *Twitter*. Alguns usuários podem trabalhar para serem percebidos pelo algoritmo: sabe-se que adolescentes marcam suas atualizações de *status* com nomes de marcas conhecidas, não relacionadas ao assunto, esperando que o *Facebook* privilegie essas atualizações nos "Feeds de Notícias" dos seus amigos<sup>8</sup>. Outros podem trabalhar para fugir de um algoritmo: sabemos que usuários de *Napster* e *P2P*, que compartilham músicas protegidas por direitos autorais, escrevem os nomes dos artistas com pequenos erros ortográficos para que os usuários consigam encontrar as músicas da "Britny Speers", mas o *software* da indústria musical não<sup>9</sup>.

Isso é burlar o sistema? Ou é uma maneira fundamental pela qual, até certo ponto, orientamo-nos perante os meios de distribuição através dos quais esperamos falar? Com base nos critérios do algoritmo em questão (ou de acordo com nossa melhor estimativa do seu funcionamento), nós já nos tornamos *algoritmicamente identificáveis* de várias formas. Isso não é tão diferente do que assessores de imprensa que orientam seus esforços para encaixar

seus clientes nas rotinas da indústria de notícias: programar um comunicado de imprensa para sair na transmissão da noite, ou fornecer gravações de vídeo para uma emissora ávida por imagens emocionantes, são técnicas para lidar com uma mídia que pode ampliar seus esforços. Hoje, para todos nós, as redes sociais e a web oferecem uma espécie análoga de "visibilidade mediada" (THOMPSON, 2005, p. 49), e nós nos beneficiamos de forma semelhante quando lidamos com esses algoritmos.

## Backstage Access - acesso aos bastidores

Mas quem está em melhor posição de compreender e operar os algoritmos públicos que tanto importam para a circulação pública de conhecimento? A visão sobre como funcionam os algoritmos de informação é uma forma de poder: vital para a participação no discurso público, essencial para ganhar visibilidade online, constituidor de credibilidade e das oportunidades que resultam dela. Como mencionamos anteriormente, os critérios e os códigos dos algoritmos geralmente são obscuros - mas não da mesma forma para todos. Para a maioria dos usuários, o conhecimento sobre os algoritmos é vago, simplista e, às vezes, equivocado; eles podem tentar empurrar o algoritmo tanto de maneiras que simplesmente consideram melhores (*hashtags*, *metadados*) ou que fundamentalmente equivocadas em relação aos critérios do algoritmo (como retuitar várias vezes a mesma mensagem na esperança de figurar entre os "Trendings" do *Twitter*). Os profissionais de SEO e os *spammers* também têm pouco acesso, mas desenvolveram habilidades técnicas para deduzir os critérios do algoritmo a partir de testes e engenharia reversa. Comunidades de entusiastas em tecnologia e críticos se engajam em tentativas semelhantes de descobrir o funcionamento desses sistemas, seja por diversão, conhecimento, vantagem pessoal ou determinadas interrupções. Os legisladores, que estão apenas começando a questionar as implicações dos algoritmos para um

8 Nota do autor: danah boyd, comunicação pessoal.

9 Nota do autor: ABC News. 5 mar. 2001. *Napster faced with big list, trick names*. In: ABC News. Disponível em: <http://abcnews.go.com/Entertainment/story?id=108389>. Acesso em 22 abr. 2013.

comércio justo ou para o discurso político, receberam até agora apenas as explicações mais gerais possíveis: os provedores de informação afirmam que seus algoritmos são segredos comerciais que não devem ser divulgados em espaços públicos.

Além disso, de fato algumas das partes interessadas recebem acesso garantido ao algoritmo, ainda que sob condições controladas. Os anunciantes recebem um tipo de acesso aos bastidores de funcionamento para saber como fazer o melhor lance<sup>10</sup>. Os provedores de informação que oferecem Interfaces de Programação de Aplicativos (APIs - *Application Programming Interfaces*, em inglês), para seus parceiros comerciais e desenvolvedores, oferecem um vislumbre sob o capô, mas os vincula com contratos e acordos de não-divulgação no mesmo instante. O acesso, a compreensão e os direitos sobre os algoritmos que desempenham um papel crucial no discurso e no conhecimento públicos tendem a mudar entre as diferentes partes interessadas e em circunstâncias específicas - mudando também o poder disponível para essas partes, e para aqueles que elas representam, de construir para os algoritmos, navegá-los e regulá-los.

## Domesticação

Por mais que essas ferramentas nos incitem a nos tornarmos legíveis para elas, nós também as inserimos nas nossas práticas, alterando seu sentido e até mesmo seu design, algumas vezes. Silverstone (1994) sugere que à medida que as tecnologias são oferecidas ao público, elas passam por um processo de “domesticação”: essas tecnologias entram em nossas casas literalmente, mas também de forma figurada - os usuários as transformam em propriedade, incorporando-as em suas rotinas e imbuindo-as com significados adicionais que os provedores não puderam antecipar. Algoritmos de informação

10 Nota da tradutora: o sistema de anúncios pagos nas plataformas segue o mesmo modelo de um leilão - há um preço inicial, mas esse valor vai sendo ajustado de acordo com os outros “competidores” pelo mesmo público.

pública certamente são importantes para o modo como os usuários encontram informações, comunicam-se uns com os outros e percebem o mundo ao seu redor. Mas, mais do que isso, os usuários expressam preferências por suas ferramentas de busca favoritas, opinam sobre as recomendações de um site como sendo incômodas, intuitivas ou precisas. Alguns usuários confiam em uma determinada ferramenta, enquanto outros desconfiam dela, usando-a com cuidado ou nem a utilizando de todo. Os usuários do *iPhone* da *Apple* trocam dicas entre si sobre como fazer a *Siri*, a assistente virtual da empresa, a falar seu repertório de réplicas<sup>11</sup> e depois compartilham a indignação com as respostas dadas a questões políticas polêmicas<sup>12</sup>. Usuários satisfeitos com o *Facebook* tornam-se críticos amanhã quando o algoritmo do “Feed de Notícias” é alterado, aparentemente por motivações econômicas - enquanto continuam a postar atualizações de status durante e depois da revolta. Os usuários, perante as assimetrias de poder da coleta de dados e vigilância online, desenvolveram uma série de táticas de “ofuscação” para escapar ou contaminar as tentativas dos algoritmos de conhecê-los (BRUNTON; NISSENBAUM, 2011). Embora seja crucial considerarmos as formas pelas quais as ferramentas algorítmicas moldam nosso contato com as informações, não devemos insinuar que os usuários estejam sob a influência dessas ferramentas. A realidade é mais complicada e mais particular.

Os usuários também podem recorrer a esses algoritmos, refletirem sobre si com base nos seus dados; muitos sites permitem nos apresentarmos para os outros e para nós mesmos, incluindo o nosso perfil público, o desempenho das nossas amizades, a expressão das nossas preferências ou um re-

11 Nota do autor: <http://siri-sayings.tumblr.com/>. Acesso em: 22 abr. 2013.

12 Nota do autor: WORTHAM, Jenna. 30 nov. 2011. Apple says Siri's abortion answers are a glitch. In: The New York Times. Disponível em: <http://bits.blogs.nytimes.com/2011/11/30/apple-says-siris-abortion-answers-are-a-glitch/>. Acesso em: 22 abr. 2013.

gistro das nossas atividades recentes. O recurso da linha do tempo do *Facebook* organiza as atividades dos usuários como lembranças cronológicas sobre si mesmos; o prazer de ver o que é selecionado algoritmicamente oferece uma espécie de encantamento que vai além da composição das fotos e das postagens em primeiro lugar. Mas os algoritmos também podem funcionar como uma “tecnologia do *self*” (FOUCAULT, 1988), particularmente convincente, quando parecem ratificar, independentemente, a visibilidade pública de uma determinada pessoa. É comum, hoje, procurar a si mesmo no *Google*: ver-me aparecer como o melhor resultado da busca pelo meu nome, oferece uma espécie de garantia da minha tênue existência pública. Existe uma sensação de validação quando um tópico sobre seu animal de estimação aparece no “*Trendings*” e no *Twitter*; quando a *Amazon* lhe recomenda um livro que você ama ou quando a função “*Genius*” do *iTunes* da *Apple* compõe uma lista de reprodução interessante a partir da sua biblioteca de músicas. Seja quando nós efetivamente adaptamos nossas compras da *Amazon* para parecermos versados (assim como as famílias pesquisadas pela *Nielsen* exageravam ao informar quanto assistiam a *PBS* e *C-Span*)<sup>13</sup> ou se simplesmente gostamos quando o algoritmo confirma nosso senso de identidade, os algoritmos são um convite poderoso para nos entendermos através de lentes independentes que eles prometem fornecer.

Os algoritmos não são apenas o que seus designers fazem deles, ou o que eles fazem a partir da informação que processam. São também o que nós fazemos deles dia após dia - mas com uma ressalva: como a lógica, a manutenção e o *redesign* desses algoritmos permanecem nas mãos dos provedores de informação, eles claramente estão numa posição privilegiada para reescrever nosso entendimento

<sup>13</sup> Nota da tradutora: o autor fala sobre quando os usuários ajustam seus algoritmos para que eles pareçam “mais inteligentes” a partir das recomendações oferecidas. O exemplo refere-se às famílias pesquisadas para avaliação da *Nielsen* que afirmavam assistir a canais considerados “mais cultos” ou “mais sérios”.

sobre eles, ou para gerar uma longa incerteza sobre os critérios que dificultam tratarmos os algoritmos como sendo verdadeiramente nossos.

## Lógicas de conhecimento

Teorizar é fácil, mas é substancialmente mais difícil documentar como os usuários podem mudar suas visões de mundo para acomodar as lógicas subjacentes e os pressupostos implícitos dos algoritmos que usam regularmente. Há uma razão para que as lógicas de funcionamento desses algoritmos não apenas moldem as práticas dos usuários, mas também os levem a internalizar suas normas e prioridades: Bucher (2012) argumenta que o algoritmo *EdgeRank*, usado pelo *Facebook* para determinar quais atualizações de *status* são prioritariamente exibidas no “*Feed de Notícias*” de um usuário, encoraja uma “subjetividade participativa” nos usuários, que reconhecem que gestos de afinidade (como comentar a foto de um amigo) são um critério central para esse algoritmo. Longford (2005) argumenta que o código da plataforma comercial nos “habituata”, através de incessantes solicitações e configurações padrão cuidadosamente projetadas, a dar mais informações pessoais sobre nós mesmos. Mager (2012) e Van Couvering (2010) propõem que os princípios do capitalismo estão incorporados no funcionamento das ferramentas de busca.

Mas, nós não precisamos recorrer a tais teorias de dominação ideológica para sugerir que os algoritmos, projetados para oferecer conhecimentos relevantes, também oferecem formas de conhecer - e que, à medida que se tornam mais comuns e confiáveis, suas lógicas se tornam auto-afirmativas. A ferramenta de pesquisa do *Google*, entre suas 200 indicações, presume que um conhecimento relevante é assegurado em grande parte pela ratificação pública e está ajustada para dar grande peso às opiniões dos que já são autenticados publicamente. Esta mistura da sabedoria das massas e das autoridades coletivamente certificadas é a solução encontrada



pelo *Google* para a antiga tensão entre o conhecimento e o senso-comum, no permanente problema de “como conhecer”. Não é sem precedentes, e não é uma maneira fundamentalmente errada de conhecer, mas é específica, com suas próprias ênfases e miopias. Hoje, a solução do *Google* é operacionalizada em uma ferramenta que bilhões de pessoas usam todos os dias, a maioria das quais a experimenta simplesmente como algo que “funciona” de forma simples e sem problemas. Até certo ponto, o *Google* e seu algoritmo ajudam a afirmar e a normalizar essa lógica de conhecimento como “correta”, tão correta quanto seus resultados parecem ser.

## A produção de públicos calculados

Ito, Boyd e outros autores recentemente introduziram o termo “públicos em rede” (BOYD, 2010; ITO, 2008; VARNELIS, 2008) para destacar as comunidades de usuários que podem se unir através das mídias sociais e a forma como as tecnologias estruturam como esses públicos podem se formar, interagir e, às vezes, se desfazer. “Enquanto os públicos em rede têm muito em comum com outros tipos de públicos, a forma como a tecnologia os estrutura introduz affordances distintas que moldam como as pessoas se engajam com esses ambientes” (BOYD, 2010, p.39). Visto que os algoritmos são um componente tecnológico central nesses ambientes mediados, eles também ajudam a estruturar os públicos que podem emergir usando a tecnologia digital.

Algumas questões têm sido levantadas sobre as maneiras que o funcionamento dos algoritmos de informação, e os modos pelos quais escolhemos navegar por eles, poderiam minar nossos esforços de sermos cidadãos envolvidos. A capacidade de personalizar resultados de pesquisa e notícias online foi a primeira, e talvez melhor articulada, dessas preocupações. Com as ferramentas de busca contemporâneas, os resultados que dois usuários recebem para a mesma busca podem ser bem diferentes; em um serviço

de notícias ou em uma rede social, as informações oferecidas podem ser adaptadas especificamente às preferências do usuário (pelo usuário ou pelo provedor), de modo que, na prática, as histórias apresentadas, como as mais importantes, podem ser tão distintas de usuário para usuário que não exista sequer um objeto comum de diálogo entre elas. Sunstein (2001) e, mais recentemente, Pariser (2011) argumentam que, quando os serviços de informação algorítmica podem ser personalizados a esse ponto, a diversidade do conhecimento público e do diálogo político pode ser minada. Somos levados - por algoritmos e por nossa própria preferência pelos que pensam de forma semelhante - para dentro de “filtros bolha” (*ibid.*), onde encontramos apenas as notícias que esperamos encontrar e as perspectivas políticas que já nos são caras.

Mas os algoritmos não estruturam apenas nossas interações com os outros enquanto membros de públicos em rede. Eles também trafegam por *públicos calculados* que eles mesmos produzem. Quando a *Amazon* recomenda um livro que “clientes como você” compraram, está invocando e afirmando conhecer um público com o qual somos convidados a sentir afinidade - embora a população em que essas recomendações se baseiam não seja transparente e certamente não coincida com toda sua base de consumidores. Quando o *Facebook* oferece como configuração de privacidade que as informações de um usuário sejam vistas por “amigos e amigos de amigos”, ele transforma um conjunto distinto de usuários em uma audiência - trata-se de um grupo que não existia até aquele momento, e que só o *Facebook* sabe sua composição precisa. Esses grupos gerados por algoritmos podem se sobrepor, podem ser uma aproximação imprecisa ou podem não ter nada a ver com os públicos que o usuário procurou.

Alguns algoritmos vão além, fazendo afirmações sobre os públicos que pretendem conhecer



e o lugar dos usuários no meio deles. Em outro trabalho, argumento que o algoritmo do “*Trendings*” do *Twitter* promete aos usuários um vislumbre do que um público particular (nacional ou regional) está falando em um determinado momento, mas se trata de um público construído, moldado pelos critérios específicos do *Twitter* e que não são especificados em grande parte (GILLESPIE, 2012). O *Klout*, um serviço online que rastreia as atividades e a reputação dos usuários no *Facebook*, no *Twitter* e em outros lugares, promete calcular a influência dos usuários através dessas várias plataformas de redes sociais. Suas medidas são intuitivas em sua definição, mas completamente opacas em seus mecanismos. A disputa entre os “públicos em rede” forjados pelos usuários e os “públicos calculados” oferecidos pelos algoritmos complica ainda mais a dinâmica da sociabilidade em rede.

Junto a outras medidas de opinião pública, como sondagens ou levantamentos, o problema central é a extrapolação, na qual se presume que um subconjunto represente uma população inteira. Com os algoritmos, a população pode ser a base de usuários inteira, às vezes, composta por centenas de milhões de pessoas (mas apenas a base de usuários a qual o provedor do algoritmo tem acesso). Em vez disso, o problema central aqui é que a intenção por trás dessas representações calculadas do público não é atuarial, de modo algum. Algoritmos, que alegam identificar o que está “em voga”, dedicam-se a estimar uma aproximação calculada de um público a partir dos rastros das atividades de seus participantes, e em seguida lhes apresentam um relatório sobre o que eles mais falaram. Mas, por trás disso, podemos nos perguntar: qual é a vantagem de fazer tais caracterizações para esses provedores e como isso molda o que eles estão medindo? Quem está sendo escolhido para ser medido a fim de produzir essa representação e quem é deixado de fora desse cálculo? E, talvez o mais importante, como estas tecnologias, agora não apenas tecnologias

de avaliação, mas de representação, ajudam a constituir e codificar os públicos que eles afirmam medir, públicos que de outra forma não existiriam, exceto quando o algoritmo os convoca?

Essas questões são muito importantes, e vão importar mais ainda, na medida em que as representações do público produzidas pelos algoritmos de informação sejam tomadas como legítimas, pelos usuários ou pelas autoridades, e incorporadas ao projeto modernista mais amplo de reflexividade (GIDDENS, 1990). “A sociedade está envolvida no monitoramento de si mesma, examinando-se, retratando-se de várias formas e usando os entendimentos resultantes para organizar suas atividades” (BOYER; HANNERZ, 2006, p. 9). O que o *Twitter* afirma importar para os “americanos” ou o que a *Amazon* diz que os adolescentes leem são formas de conhecimento autoritário que podem e serão invocadas por instituições cujo objetivo é regular essas populações.

A crença de que tais algoritmos, combinados com quantidades massivas de dados de usuários, são melhores para nos dizer coisas sobre a natureza do público ou sobre a constituição da sociedade, também provou ser sedutora para os estudiosos. A ciência social voltou-se avidamente para as técnicas computacionais ou para o estudo da socialidade humana através do “*big data*” (LAZER et al, 2009); para uma visão crítica, ver (BOYD; CRAWFORD, 2012), na esperança de desfrutar do tipo de *insights* que as ciências biológicas alcançaram, buscando algoritmicamente agulhas nesses palheiros digitais com todos esses dados. A abordagem é sedutora: ter milhões de dados gera maior legitimidade e é empolgante ver como os algoritmos parecem detectar padrões que os pesquisadores não puderam enxergar de outras formas. “Para um certo tipo de cientista social, os padrões de tráfego de milhões de e-mails parecem ter caído do céu” (NATURE, 2007). Mas essa abordagem metodológica deve considerar as complexidades descritas até aqui,

particularmente quando os dados de um pesquisador foram gerados por algoritmos comerciais. As técnicas de pesquisa computacional não são barômetros do social. Elas produzem hieróglifos: moldados pela ferramenta pela qual são esculpidos, exigindo uma interpretação sacerdotal, eles contam histórias poderosas, mas muitas vezes mitológicas - geralmente a serviço dos deuses.

Finalmente, quando nós somos os dados, o que devemos fazer com as associações que os algoritmos afirmam identificar sobre nós enquanto sociedade - o que não sabemos ou que talvez não queremos saber? No inquietante exemplo de Ananny (2011), ele percebeu que o mercado de aplicativos do *Android* recomendava um aplicativo sobre a localização de "predadores sexuais"<sup>14</sup> para os usuários que baixaram o *Grindr*, uma ferramenta de rede social baseada em localização para homens gays. Ele especula sobre como os algoritmos do mercado do *Android* podem ter feito essa associação - até mesmo as operadoras do *Android* não conseguiram explicar com facilidade. O algoritmo cometeu um erro? O algoritmo fez uma associação direta, simplesmente combinando aplicativos que continham "sexo" na descrição? Ou o mecanismo de recomendação do *Android* realmente identificou uma associação sutil que, embora não desejemos, é recorrentemente feita na nossa cultura, entre homossexualidade e comportamento predatório? Zimmer (2007) observa um caso semelhante: uma busca pela frase "ela inventou" retornava a pergunta "você quis dizer 'ele inventou'"? Ou seja, isso aconteceu até que *Google* alterou os resultados. Embora preocupante em sua política de gênero, a resposta do *Google* foi completamente "correta", explicada pelo triste fato de que, em todo o *corpus* da *web*, a palavra "inventado" é precedida muito mais vezes por "ele" do que por "ela". O algoritmo reconheceu isso - e presumiu erroneamente que a pesquisa "ela inventou" fosse apenas

um erro tipográfico. Aqui, o *Google* se prova muito menos sexista do que nós somos. Em resposta ao exemplo de Ananny, Gray sugere que, assim como devemos examinar os algoritmos que fazem esse tipo de associações, também devemos investigar os "algoritmos culturais" que essas associações representam (isto é, associando sistematicamente a homossexualidade com a predação sexual) em um conjunto massivo e distribuído de "pontos de dados" - nós mesmos.

### Considerações finais

Compreender os algoritmos e seu impacto no discurso público, então, requer pensar não apenas sobre como eles funcionam, onde são implantados ou o que os movimenta financeiramente. Este não é simplesmente um chamado para desvendar seu funcionamento interno e destacar seus critérios implícitos. É uma investigação sociológica que não interessa aos provedores desses algoritmos, que nem sempre estão na melhor posição para sequer perguntar. Requer analisar por que os algoritmos estão sendo vistos como uma lógica de conhecimento confiável; como eles desmoronam e são reparados ao entrar em contato com o vai e vem do discurso público; e onde os pressupostos políticos podem estar gravados não só em seu design, mas também serem constitutivos da sua utilização e sua legitimidade generalizada.

Vejo a emergência do algoritmo enquanto uma ferramenta de informação confiável, como a mais recente resposta à fundamental tensão do discurso público. Os meios pelos quais produzimos, circulamos e consumimos informações em uma sociedade complexa devem ser necessariamente operados através da divisão do trabalho: alguns produzem e selecionam informações, e o resto de nós, pelo menos naquele momento, somente pode conceder aquilo que valha a pena e só pode aceitá-las, se serve de consolo. Toda mídia pública anterior a essa também enfrentou esse desafio, desde

<sup>14</sup> Nota da tradutora: o termo usado pelo autor em inglês é "sex-offender".

os pregoeiros, até os jornais e as transmissões via televisão. Em cada uma delas, quando entregamos o fornecimento de conhecimento para outros, ficamos vulneráveis a suas escolhas, métodos e subjetividades. Às vezes, é um fornecimento positivo de competência, perspicácia editorial, gosto refinado. Mas também somos cautelosos à intervenção, às falhas humanas e aos interesses particulares, e nos encontramos com mecanismos apenas secundários de confiança social para confirmar o que é verdadeiro e relevante (SHAPIN, 1995). Seus procedimentos estão em grande parte indisponíveis para nós e são inevitavelmente seletivos, enfatizando algumas informações e descartando outras, e suas escolhas podem trazer consequências. Existe a nítida possibilidade de erro, viés, manipulação, comodismo, influência comercial ou política, ou falhas sistêmicas. O processo de seleção sempre pode ser uma oportunidade de organizar as informações por razões outras que não a relevância: por propriedade, por interesses próprios comercial ou institucional, ou por ganho político. Juntas, essas questões representam uma vulnerabilidade fundamental que nunca poderemos resolver completamente. Nós podemos apenas criar garantias da melhor maneira possível.

A partir dessa perspectiva, devemos ver os algoritmos não apenas como códigos com consequências, mas sim como o mais recente mecanismo construído socialmente e institucionalmente gerenciado para convencer o julgamento público: uma nova lógica de conhecimento. Podemos considerar a lógica algorítmica como oposta, e até talvez suplantar, à editorial enquanto *lógica concorrente*. A lógica editorial depende das escolhas subjetivas de especialistas, que são eles próprios feitos e autorizados como tal através de processos institucionais de treinamento e certificação, ou validados pelo público através dos mecanismos do mercado. A lógica algorítmica, em contrapartida, depende das escolhas procedimentalizadas de uma máquina, projetadas por operadores humanos para automatizar alguma representação do julgamento humano ou desenterrar padrões atra-

vés de traços sociais coletados. Ambas lutam com, e afirmam resolver, o problema fundamental do conhecimento humano: como identificar informações relevantes cruciais para o público através de meios inevitavelmente humanos, mas de formas livres de erros humanos, vieses ou manipulações. Ambas abordagens algorítmicas e editoriais do conhecimento são profundamente importantes e profundamente problemáticas; grande parte dos estudos em comunicação, mídia, tecnologia e público lida com uma ou ambas técnicas e suas armadilhas.

Uma investigação sociológica sobre os algoritmos deve aspirar revelar o complexo funcionamento desta máquina de conhecimento, tanto o processo pelo qual ela escolhe as informações para os usuários, quanto o processo social pelo qual ela se torna um sistema legítimo. Mas pode haver algo, no final, impenetrável em relação aos algoritmos. Eles são projetados para trabalhar sem intervenção humana, são deliberadamente ofuscados, e trabalham com informações em uma escala que é difícil de compreender (pelo menos sem outras ferramentas algorítmicas). E talvez mais do que isso, nós queremos nos livrar do dever de sermos céticos sobre informações que nunca poderemos garantir com certeza. Através desses mecanismos, chegamos a um acordo sobre esse problema (se não o resolvemos) e, assim, eles são soluções com as quais não apenas podemos contar, mas em que devemos acreditar. Mas esse tipo de fé (VAIDHYANATHAN, 2011) torna difícil reconhecer sobriamente suas falhas e fragilidades.

Assim, de várias formas, os algoritmos continuam fora do nosso alcance e eles são projetados para continuar mesmo. Isso não quer dizer que não devemos aspirar a iluminar seu funcionamento e a seu impacto. Nós deveríamos. Mas talvez nós também precisemos nos preparar para nos depararmos, mais e mais, com associações inesperadas e indescritíveis que eles vão desenhar para nós, às vezes; a incerteza fundamental sobre com quem estamos falando ou quem estamos ouvindo; e as

implicações palpáveis, porém opacas, que se movem silenciosamente por baixo do conhecimento quando ele é gerenciado por algoritmos.

## Agradecimentos

Eu gostaria de agradecer aos meus colegas do *Culture Digitally* pela ajuda e conselhos nesse artigo e pelo generoso apoio do *Collegium de Lyon* e do *The European Institutes for Advanced Study* (EU-RIAS) *Fellowship Programme*.

## Refêrencias

ANANNY, Mike. The curious connection between apps for gay men and sex offenders. *The Atlantic*, 2011, 14 v.

ANDERSON, Chris W. Deliberative, agonistic, and algorithmic audiences: Journalism's vision of its public in an age of audience transparency. *International Journal of Communication*, p. 529-547, 2011, 5 v.

BALKA, Ellen. *Mapping the body across diverse information systems: Shadow bodies and how they make us human*. Trabalho apresentado no Annual Meeting for the Society for Social Studies of Science, Cleveland, OH. 2011

BEER, David. Power through the algorithm? Participatory web cultures and the technological unconscious. *New Media & Society*, n. 6, p. 985-1002, 2009, 11 v.

BERRY, David Michael. Introduction. In: BERRY, David Michael (Ed.). *Life in Code and Software: Mediated Life in a Complex Computational Ecology*. London: Open Humanities Press, 2012.

BLIJKER, Wiebe E.. *Of Bicycles, Bakelites, and Bulbs: Toward a Theory of Sociotechnical Change*. Cambridge, MA: MIT Press, 1997.

BOWKER, Geoffrey C. *Memory practices in the sciences*. Cambridge, MA: Mit Press, 2006.

BOWKER, Geoffrey C.; STAR, Susan Leigh. *Sorting things out: Classification and its consequences*. Cambridge, MA: MIT Press, 2000.

BOYD, Danah; CRAWFORD, Kate. Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon. *Information, communication & society*, n. 5, p. 662-679, 2012, 15 v.

BOYD, danah. Social network sites as networked publics: Affordances, dynamics, and implications. PAPACHARISSI, Zizi (Ed.). *A networked self: Identity, community, and culture on social network sites*. Routledge, 2010.

BOYER, Dominic; HANNERZ, Ulf. Introduction: Worlds of journalism. *Ethnography* 7, p. 5-17, 2006, 1 v.

BRACHA, Oren; PASQUALE, Frank. Federal Search Commission-Access, Fairness, and Accountability in the Law of Search. *Cornell Law Review*, n. 6 p. 1149-1210, 2008, 93 v.

BRAUN, Josh. *Electronic components and human interventions: Distributing television news online*. Tese de Doutorado. Cornell University, 2011.

BRUNTON, Finn; NISSENBAUM, Helen. Vernacular resistance to data collection and analysis: A political theory of obfuscation. *First Monday*, n. 5, 2011, 16 v.

BUCHER, Taina. Want to be on the top? Algorithmic power and the threat of invisibility on Facebook. *New Media & Society*, n. 7, p. 1164-1180, 2012, 14 v.

CHENEY-LIPPOLD, John. A new algorithmic identity: Soft biopolitics and the modulation of control. *Theory, Culture & Society*, n. 6, p. 164-181, 2011, 28 v.

COULDRY, Nick. *Media, Society, World: Social Theory and Digital Media Practice*. Cambridge, UK: Polity Press, 2012.

ELMER, Greg. Exclusionary rules? The politics of protocols. In: CHADWICK, Andrew; HOWARD, Philip

- N. (Ed.). *Routledge handbook of internet politics*, p. 376-283, 2008.
- FINDELSTEIN, Seth. Google, links, and popularity versus authority. In: TUROW, Joseph; TSUI, Lokman (Ed.). *The Hyperlinked Society: Questioning Connections in the Digital Age*. Ann Arbor: University of Michigan Press, p. 104-120, 2008.
- FOUCAULT, Michel. Technologies of the self. In: MARTIN, Luther; GUTMAN, Huck; HUTTON, Patrick (Ed.). *Technologies of the Self: A Seminar with Michel Foucault*. London: Tavistock, p. 16-49, 1988.
- GALLOWAY, Alexander R. *Protocol: How control exists after decentralization*. Cambridge: MA, MIT press, 2004.
- GIDDENS, Anthony. *The Consequences of Modernity*. Stanford, CA: Stanford University Press, 1990.
- GILLESPIE, Tarleton; POSTIGO, Hector Postigo. 4 maio 2012. "Five more points." In: *Culture Digitally*. Disponível em: <<http://culturedigitally.org/2012/05/five-more-points>>. Acesso em: 9 mar. 2018.
- GILLESPIE, Tarleton. 2012. "Can an algorithm be wrong?." In: *Limn 2*. Disponível em: <<http://limn.it/can-an-algorithm-be-wrong/>>. Acesso em: 22 abr. 2013.
- GILLESPIE, Tarleton. *Speaking from Platforms*. New Haven, CT: Yale University Press. Forthcoming, 2012.
- GILLESPIE, Tarleton. The politics of "platforms." *New Media & Society*, n. 3, p. 1-18, 2010, 12 v.
- GITELMAN, Lisa; JACKSON, Virginia. Introduction. In: GITELMAN, Lisa (Ed.). *Raw data is an oxymoron*. Cambridge: MA, MIT Press, 2013. p. 1-4.
- GITELMAN, Lisa. *Always already new. Media, history, and the data of culture*. Cambridge: MA, MIT Press, 2006.
- GOFFEY, Andrew. Algorithm. In: FULLER, Matthew (Ed.). *Software Studies: A Lexicom*. Cambridge, MA: MIT Press, p. 15-20, 2008.
- GRANKA, Laura A. The politics of search: A decade retrospective. *The Information Society*, n. 5, p. 364-374, 2010, 26 v.
- GRIMMELMAN, James. *The Google dilemma*. New York Law School Law Review, p. 939-950., 2008, 53 v.
- GRIMMELMANN, James. Some skepticism about search neutrality. In: SZOKA, Berin; MARCUS, Adam (Ed.). *The Next Digital Decade: essays on the future of the internet*. Washington, DC: TechFreedom, 2010.
- HALAVAIS, Alexander. *Search Engine Society*. Cambridge, UK: Polity Press, 2008.
- HESMONDHALGH, David. Bourdieu, the media and cultural production. *Media, culture & society*, n. 2, p. 211-231, 2006, 28 v.
- HILGARTNER, Stephen. *Science on Stage: Expert Advice as Public Drama*. Stanford, CA: Stanford University Press, 2000.
- HOWARD, Philip N. *New Media Campaigns and the Managed Citizen*. Cambridge, UK: Cambridge University Press, 2005.
- HUTCHINS, Edwin. *Cognition in the Wild*. Cambridge, MA: MIT Press, 1995.
- INTRONA, Lucas D.; NISSENBAUM, Helen. Shaping the Web: Why the politics of search engines matters. *The information society*, n. 3, p. 169-185, 2000, 16 v.
- ITO, Mizuko. Introduction. In: VARNELIS, Kazys (Ed.). *Networked Publics*. Cambridge, MA: MIT Press, p. 1-14, 2008.
- KEANE, John. Public life in the era of communicative abundance. *Canadian Journal of Communication*, n. 2, p. 165-178, 1999, 24 v.
- LANGLOIS, Ganaele. Participatory culture and the new governance of communication: The paradox of participatory media. *Television & New Media*, n. 2, p. 91-105, 2013, 14 v.



- LATOUR, Bruno. *Aramis, or the Love of Technology*, trans. Catherine Porter. Cambridge, MA: Harvard University Press, 1996.
- LATOUR, Bruno. *Science in Action: How to Follow Scientists and Engineers through Society*. Cambridge, MA: Harvard University Press, 1987.
- LAZER, David, et. al. Computational social science. *Science*, p. 721-723, 2009, 323 v.
- LIGHT, Jennifer S. When computers were women. *Technology and culture*, n. 3, p. 455-483, 1999, 40 v.
- LONGFORD, Graham. Pedagogies of digital citizenship and the politics of code. *Techné: Research in Philosophy and Technology*, v. 9, n. 1, p. 68-96, 2005, 9 v.
- MACKENZIE, Adrian. The performativity of code: Software and cultures of circulation. *Theory, Culture & Society*, n. 1, p. 71-92, 2005, 22 v.
- MAGER, Astrid. Algorithmic ideology: How capitalist society shapes search engines. *Information, Communication and Society*, n. 5, p. 769-787, 2012, 15 v.
- MANOVICH, Lev. Database as symbolic form. *Convergence*, n. 2, p. 80-99, 1999, 5 v.
- MOROZOV, Evgeny. Don't be evil. *The New Republic*, p. 13-07, 2011.
- NATURE. "A matter of trust." *Nature*, p. 637-638, 2007, 449 v.
- PARISER, Eli. *The Filter Bubble: What the Internet Is Hiding from You*. New York: Penguin Press, 2011.
- PASQUALE, Frank. 18. nov. 2009. Assessing algorithmic authority. In: Madisonian.net. Disponível em: <<http://madisonian.net/2009/11/18/assessing-algorithmic-authority/>>. Acesso em: 22 abr. 2013.
- PFAFFENBERGER, Brian. Technological dramas. *Science, Technology & Human Values*, n. 3, p. 282-312, 1992, 17 v.
- PINCH, Trevor; BJIKER, Wiebe. The social construction of facts and artefacts: Or how the sociology of science and the sociology of technology might benefit each other. *Social Studies of Science*, n. 3, p.399-411, 1984, 14 v.
- RIEDER, Bernhard. "ORDER BY column\_name: The relational database as pervasive cultural form". Trabalho apresentado na *Conferência The Lived Logics of Database Machinery*, Londres, 2012.
- ROGERS, Richard. *The Googlization question, and the inculpable engine*. In: BECKER, Konrad; 2009.
- STALDER, Felix (Ed.). *Deep Search: The Politics of Search Engines*. Edison, NJ: Transaction Publishers, p. 173-184, 2009.
- RYFE, David. The nature of news rules. *Political Communication*, n.2, p. 203-214, 2006, 23 v.
- SCHUDSON, Michael; ANDERSON, Chris Anderson. Objectivity, professionalism, and truth seeking in journalism. WAHL-JORGENSEN, Karin; HANITZSCH, Thomas (Ed.). *The handbook of journalism studies*. Routledge, 2009.
- SHAPIN, Steven. Trust, honesty, and the authority of science. In: BULGER, Ruth Ellen; BOBBY, Elizabeth Meyer; FINEBERG, Harvey (Ed.) *Society's Choices: Social and Ethical Decision Making in Biomedicine*. Washington, DC: National Academies Press, p. 388-408, 1995.
- SILVERSTONE, Roger. *Television and Everyday Life*. London: Routledge, 1994.
- SMYTHE, Dallas W. On the audience commodity and its work. In: DURHAM, Meenakshi;
- KELLNER, Douglas (Ed.) *Media and Cultural Studies: KeyWorks*. Malden, MA: Blackwell Publishing, 1981, p. 230-56.
- SOLOVE, Daniel J. *The digital person: Technology and privacy in the information age*. New York: New



York University Press, 2004.

STALDER, Felix; MAYER, Christine. The Second Index: Search Engines, Personalization and Surveillance. In: BECKER, Konrad; STALDER, Felix (Ed.) *Deep Search: The Politics of Search beyond Google*. London: Transaction Publishers, 2009, p. 98-115.

STERNE, Jonathan. *The audible past: Cultural origins of sound reproduction*. Durham NC: Duke University Press, 2003.

STRIPHAS, Theodore G. How to have culture in an algorithmic age. In: *Differences and Repetitions*. 13 jul.2010. Disponível em: <<http://striphas.blogspot.com.br/2010/07/how-to-have-culture-in-algorithmic-age.html>>. Acesso em: 9 mar. 2018.

STRIPHAS, Theodore G. *The Late Age of Print: Everyday Book Culture from Consumerism to Control*. New York: Columbia University Press, 2009.

SUNSTEIN, Cass R. *Republic.com 2.0*. Princeton, NJ: Princeton University Press, 2001.

THOMPSON, John. The new visibility. *Theory, Culture & Society*, n. 6, p. 31-51, 2005, 22 v.

TUCHMAN, Gaye. Objectivity as strategic ritual: An examination of newsmen's notions of objectivity. *American Journal of Sociology*, n. 4, p. 660-679, 1972, 77 v.

TUROW, Joseph. *The daily you: How the new advertising industry is defining your identity and your worth*. New Haven, CT: Yale University Press, 2012.

TUSHNET, Rebecca. Power without responsibility: Intermediaries and the First Amendment. *George Washington Law Review*, n.4, p. 986-1016, 2007, 76 v.

VAIDHYANATHAN, Siva. *The Googlization of everything: (and why we should worry)*. Berkeley: University of California Press, 2011.

VAN COUVERING, Elizabeth. Is relevance relevant? Market, science, and war: Discourses of search engine quality. *Journal of Computer - Mediated Communication*, n. 3, p. 866-887, 2007, 12 v.

VAN COUVERING, Elizabeth. *Search engine bias: the structuration of traffic on the World-Wide Web*. 2010. Tese de Doutorado. The London School of Economics and Political Science (LSE).

VARNELIS, Kazys (Ed.). *Networked Publics*. Cambridge, MA: MIT Press, 2008

WINNER, Langdon. *Autonomous technology: Technics-out-of-control as a theme in political thought*. Cambridge, MA: MIT Press, 1977.

ZIMMER, Michael. 9 maio 2007. Google: "Did you mean: 'He invented'?" In: Michael Zimmer. Disponível em: <<http://www.michaelzimmer.org/2007/05/09/google-did-you-mean-he-invented/>>. Acesso em: 9 mar. 2018.

ZIMMER, Michael. The externalities of search 2.0: The emerging privacy threats when the drive for the perfect search engine meets Web 2.0. *First Monday*, n. 3, 2008, 13 v.

**Tarleton Gillespie** – Pesquisador-chefe da Microsoft Research New England, Professor Associado adjunto na Cornell University (EUA), membro do Social Media Collective ([socialmediacollective.org](http://socialmediacollective.org)) e um dos fundadores do blog Culture Digitally ([culturedigitally.org](http://culturedigitally.org)). **E-mail:** [tlg28@cornell.edu](mailto:tlg28@cornell.edu)