REINVENTING DISCOVERY

The New Era of Networked Science

MICHAEL NIELSEN

REINVENTANDO A DESCOBERTA

REINVENTAR DESCOBERTA

A Nova Era da Ciência em Rede

Michael Nielsen

PRINCETON UNIVERSITY PRESS
PRINCETON AND OXFORD

Publicado pela Princeton University Press, 41 William Street, Princeton, Nova Jersey 08540 No Reino Unido: Princeton
University Press, 6 Oxford Street, Woodstock, Oxfordshire OX20 1TW press.princeton.edu

Dados de catalogação na publicação da Biblioteca do Congresso

Nielsen, Michael A., 1974 –
Reinventando a descoberta: a nova era da ciência em rede / Michael Nielsen.

p. cm.
Inclui referências bibliográficas e índice.
ISBN 978-0-691-14890-8 (capa dura)
1. Pesquisa-Inovações tecnológicas. Ciência. 3. 2. Descobertas em Internet.
4. Redes de informação. 5. Tecnologia da informação. I. Título.
Q180.55.M4N54 2011 509-
dc23 2011020248

Os dados de catalogação na publicação da Biblioteca Britânica estão disponíveis

Este livro foi composto em Minion
Impresso em papel sem ácido ÿ

Composto por SR Nova Pvt Ltd, Bangalore, Índia
Impresso nos Estados Unidos da América

10 9 8 7 6 5 4 3 2 1

Conteúdo

REINVENTANDO A DESCOBERTA

CAPÍTULO 1

Reinventando a descoberta

Tim Gowers não é um blogueiro típico. Matemático da Universidade de Cambridge, Gowers recebeu a mais alta honraria da matemática, a Medalha Fields, frequentemente chamada de Prêmio Nobel da matemática. Seu blog irradia ideias e insights matemáticos.

Em janeiro de 2009, Gowers decidiu usar seu blog para realizar um experimento social bastante incomum. Ele escolheu um problema matemático importante e difícil, ainda não resolvido, um problema que ele disse que "adoraria resolver". Mas, em vez de atacar o problema sozinho ou com alguns colegas próximos, decidiu atacá-lo completamente abertamente, usando seu blog para postar ideias e progressos parciais. Além disso, ele lançou um convite aberto pedindo ajuda a outras pessoas. Qualquer um poderia acompanhar e, se tivesse uma ideia, explicá-la na seção de comentários do blog. Gowers esperava que muitas mentes fossem mais poderosas do que uma, que elas se estimulassem mutuamente com diferentes conhecimentos e perspectivas e, coletivamente, facilitassem o trabalho de seu difícil problema matemático. Ele apelidou o experimento de Projeto Pol

O Projeto Polímata começou devagar. Sete horas depois de Gowers abrir seu blog para discussão matemática, ninguém havia comentado. Então, um matemático chamado Jozsef Solymosi, da Universidade da Colúmbia Britânica, postou um comentário sugerindo uma variação do problema de Gowers, uma variação que era mais fácil, mas que Solymosi achou que poderia lançar luz sobre o problema original. Quinze minutos depois, um professor de ensino médio do Arizona chamado Jason Dyer entrou na conversa com uma ideia própria. E apenas três minutos depois, o matemático da UCLA Terence Tao — assim como Gowers, medalhista Fields — adicionou um comentário. Os comentários explodiram: nos 37 dias seguintes, 27 pessoas escreveram 800 comentários matemáticos, contendo mais de 170.000 palavras. Lendo os comentários, você vê ideias propostas, refinadas e descartadas, tudo com uma velocidade incrível. Você vê os melhores matemáticos cometendo erros, seguindo caminhos errados, sujando as mãos ao seguir os detalhes mais banais, perseguindo implacavelme

Solução. E, apesar de todos os falsos começos e desvios, você vê um gradual despertar de insights. Gowers descreveu o processo Polymath como sendo "para a pesquisa normal o que dirigir é para empurrar um carro". Apenas 37 dias após o início do projeto, Gowers anunciou que estava confiante de que os polímatas haviam resolvido não apenas seu problema original, mas um problema mais difícil que incluía o original como um caso especial. Ele o descreveu como "uma das seis semanas mais emocionantes da minha vida matemática". Meses de trabalho de limpeza ainda precisavam ser feitos, mas o problema matemático central havia sido resolvido. (Se você quiser saber os detalhes do problema de Gowers, eles estão descritos no apêndice. Se você simplesmente quiser continuar lendo este livro, pode pular esses detalhes com segurança.)

Os polímatas não estão parados. Desde o projeto original de Gowers, quase uma dúzia de projetos Polímatas e similares foram lançados, alguns abordando problemas ainda mais ambiciosos do que o problema original de Gowers. Mais de 100 matemáticos e outros cientistas participaram; a colaboração em massa está começando a se espalhar pela matemática. Assim como o primeiro Projeto Polímata, vários desses projetos foram grandes sucessos, impulsionando de fato nossa compreensão da matemática. Outros foram sucessos mais limitados, ficando aquém de seus objetivos (às vezes extremamente ambiciosos).

De qualquer forma, a matemática massivamente colaborativa é uma nova maneira poderosa de atacar problemas matemáticos difíceis.

Por que a colaboração online em massa é útil na resolução de problemas matemáticos? Parte da resposta é que mesmo os melhores matemáticos podem aprender muito com pessoas com conhecimentos complementares e ser estimulados a considerar ideias em direções que não teriam considerado por conta própria. Ferramentas online criam um espaço compartilhado onde isso pode acontecer, uma memória de trabalho coletiva de curto prazo onde as ideias podem ser rapidamente aprimoradas por muitas mentes. Essas ferramentas nos permitem ampliar a conversa criativa, de modo que conexões que normalmente exigiriam serendipidade fortuita, em vez disso, aconteçam naturalmente. Isso acelera o processo de resolução de problemas e expande a gama de problemas que podem ser resolvidos pela mente humana.

O Projeto Polímata é uma pequena parte de uma história muito maior, uma história sobre como as ferramentas online estão transformando a maneira como os cientistas fazem descobertas. Essas ferramentas são ***ferramentas cognitivas,*** que amplificam ativamente nossa inteligência coletiva, tornando-nos mais inteligentes e, portanto, mais capazes de resolver os problemas científicos mais complexos. Para entender por que tudo isso importa, pense no século XVII e nos primórdios da ciência moderna, a época de grandes descobertas como a observação das luas por Galileu.

de Júpiter e a formulação de Newton das leis da gravitação. O maior legado de Galileu, Newton e seus contemporâneos não foram esses avanços pontuais. Foi o próprio método de descoberta científica, uma maneira de entender como a natureza funciona. No início do século XVII, era necessário um gênio extraordinário para fazer até mesmo o menor dos avanços científicos. Ao desenvolver o método de descoberta científica, os primeiros cientistas garantiram que, até o final do século XVII, tais avanços científicos fossem comuns, o resultado provável de qualquer investigação científica competente. O que antes exigia gênio tornou-se rotina, e a ciência explodiu.

Essas melhorias na forma como as descobertas são feitas são mais importantes do que qualquer descoberta isolada. Elas ampliam o alcance da mente humana para novos domínios da natureza. Hoje, as ferramentas online nos oferecem uma nova oportunidade de aprimorar a forma como as descobertas são feitas, uma oportunidade em uma escala nunca vista desde os primórdios da ciência moderna. Acredito que o processo da ciência — como as descobertas são feitas — mudará mais nos próximos vinte anos do que nos últimos 300 anos.

O Projeto Polímata ilustra apenas um aspecto dessa mudança, uma mudança na forma como os cientistas trabalham juntos para criar conhecimento. Um segundo aspecto dessa mudança é uma expansão drástica na capacidade dos cientistas de encontrar significado no conhecimento. Considere, por exemplo, os estudos que você frequentemente vê nas notícias dizendo que "tais genes causam tal e tal doença". O que torna esses estudos possíveis é um mapa genético de seres humanos que foi compilado ao longo dos últimos vinte anos. A parte mais conhecida desse mapa é o genoma humano, que os cientistas concluíram em 2003. Menos conhecido, mas talvez ainda mais importante, é o HapMap (abreviação de mapa de haplótipos), concluído em 2007, que mapeia como e onde diferentes seres humanos podem ***diferir*** em seu código genético. Essas variações genéticas determinam muito sobre nossas diferentes suscetibilidades a doenças, e o HapMap diz onde essas variações podem ocorrer — é um mapa genético não apenas de um único ser humano, mas de todo o ser humano.
corrida.

Este mapa genético humano foi o trabalho conjunto de muitos, muitos biólogos ao redor do mundo. Cada vez que obtinham um novo bloco de dados genéticos em seus laboratórios, eles os carregavam para serviços online centralizados, como o GenBank, o incrível repositório online de informações genéticas administrado pelo Centro Nacional de Informações sobre Biotecnologia dos EUA. O GenBank integra todas essas informações genéticas em um único banco de dados online de acesso público, uma compilação do trabalho de milhares de biólogos. São informações em uma escala quase infinita.

impossível de analisar manualmente. Felizmente, qualquer pessoa no mundo pode baixar gratuitamente o mapa genético e, em seguida, usar algoritmos de computador para analisá-lo, talvez descobrindo fatos até então insuspeitos sobre o genoma humano. Você pode, se quiser, acessar o site do GenBank agora mesmo e começar a navegar pelas informações genéticas. (Para links para o GenBank e outros recursos, consulte as "Notas sobre Fontes", a partir da página 347.) É isso, de fato, que torna possíveis os estudos que ligam genes a doenças: os cientistas que realizam os estudos começam encontrando um grande grupo de pessoas com a doença e também um grupo de controle de pessoas sem a doença. Eles então usam o mapa genético humano para encontrar correlações entre a incidência da doença e as diferenças genéticas entre os dois grupos.

Um padrão semelhante de descoberta está sendo usado em toda a ciência. Cientistas de muitas áreas estão colaborando online para criar enormes bancos de dados que mapeiam a estrutura do universo, o clima mundial, os oceanos do mundo, as línguas humanas e até mesmo todas as espécies de vida. Ao integrar o trabalho de centenas ou milhares de cientistas, estamos mapeando coletivamente o mundo inteiro. Com esses mapas integrados, qualquer pessoa pode usar algoritmos de computador para descobrir conexões que nunca foram suspeitadas antes. Mais adiante no livro, veremos exemplos que vão desde novas maneiras de rastrear surtos de gripe até a descoberta de pares de buracos negros supermassivos em órbita. Estamos, peça por peça, reunindo todo o conhecimento do mundo em um único edifício gigante. Esse edifício é vasto demais para ser compreendido por qualquer indivíduo trabalhando sozinho. Mas novas ferramentas computadorizadas podem nos ajudar a encontrar o significado oculto em todo esse conhecimento.

Se o Projeto Polymath ilustra uma mudança na forma como os cientistas colaboram para criar conhecimento, e o GenBank e os estudos genéticos ilustram uma mudança na forma como os cientistas encontram significado no conhecimento, uma terceira grande mudança é uma mudança na relação entre ciência e sociedade. Um exemplo dessa mudança é o site Galaxy Zoo, que recrutou mais de 200.000 voluntários online para ajudar astrônomos a classificar imagens de galáxias.
Esses voluntários recebem fotografias de galáxias e são solicitados a responder perguntas como "Esta é uma galáxia espiral ou elíptica?" e "Se for uma espiral, os braços giram no sentido horário ou anti-horário?". Trata-se de fotografias tiradas automaticamente por um telescópio robótico e nunca antes vistas por nenhum olho humano. Podemos pensar no Galaxy Zoo como um censo cosmológico, o maior já realizado, um censo que já produziu mais de 150 milhões de classificações de galáxias.

Os astrônomos voluntários que participam do Galaxy Zoo estão fazendo descobertas impressionantes. Eles, por exemplo, descobriram recentemente uma classe inteiramente nova de galáxias, as "galáxias ervilha-verde" — assim chamadas porque as galáxias realmente se parecem com pequenas ervilhas verdes — onde as estrelas estão se formando mais rápido do que em quase qualquer outro lugar do universo. Eles também descobriram o que se acredita ser o primeiro exemplo de um espelho de quasar, uma enorme nuvem de gás com dezenas de milhares de anos-luz de diâmetro, que brilha intensamente à medida que o gás é aquecido pela luz de um quasar próximo. Em apenas três anos, o trabalho dos voluntários do Galaxy Zoo resultou em 22 artigos científicos, e muitos outros estão e

O Galaxy Zoo é apenas um dos muitos projetos online de ciência cidadã que estão recrutando voluntários, a maioria sem formação científica, para ajudar a resolver problemas de pesquisa científica. Veremos exemplos que abrangem diversas áreas da ciência, desde voluntários que usam jogos de computador para prever o formato de moléculas de proteína até voluntários que ajudam a entender como os dinossauros evoluíram. Trata-se de projetos científicos sérios, projetos em que grandes grupos de voluntários com pouca formação científica conseguem abordar problemas científicos que estão além do alcance de pequenos grupos de profissionais. Não há como uma equipe de profissionais fazer o que o Galaxy Zoo faz — mesmo trabalhando em tempo integral, os profissionais não têm tempo para classificar centenas de milhares (ou mais) de galáxias. Você poderia imaginar que eles usariam computadores para classificar as imagens das galáxias, mas, na verdade, os voluntários humanos classificam as galáxias com mais precisão do que até mesmo os melhores programas de computador. Assim, os voluntários em projetos como o Galaxy Zoo estão expandindo os limites dos problemas científicos que podem ser resolvidos e, ao fazê-lo, mudando tanto quem pode ser um cientista quanto o que significa ser um cientista. Até que ponto a fronteira entre cientista profissional e amador pode ser turva? Será que um dia veremos Prêmios Nobel conquistados por grandes colaborações dominadas por amadores?

A ciência cidadã faz parte de uma mudança maior na relação entre ciência e sociedade. O Galaxy Zoo e projetos semelhantes são exemplos de instituições que estão conectando a comunidade científica e o restante da sociedade de novas maneiras. Veremos que ferramentas online possibilitam muitas outras novas instituições de conexão, incluindo a publicação em acesso aberto, que dá ao público acesso direto aos resultados da ciência, e os blogs científicos, que estão ajudando a criar uma comunidade científica mais aberta e transparente.
Que outras novas maneiras podemos encontrar para construir pontes entre a ciência e o resto da sociedade? E qual será o impacto a longo prazo dessas novas instituições de ponte?

A história até agora é uma história otimista de possibilidades, de novas ferramentas que estão mudando o mundo. Mas há um problema com essa história: alguns obstáculos importantes que impedem os cientistas de aproveitar ao máximo as ferramentas online. Para entender os obstáculos, considere os estudos que relacionam genes a doenças que discutimos anteriormente. Há uma parte crucial dessa história que ignorei, mas que na verdade é bastante intrigante: ***por que*** os biólogos compartilham dados genéticos no GenBank? Pensando bem, é uma escolha peculiar: se você é um biólogo profissional, é vantajoso manter os dados em segredo pelo maior tempo possível.

Por que compartilhar seus dados online antes de ter a chance de publicar um artigo ou registrar uma patente para seu trabalho? No mundo científico, são os artigos e, em algumas áreas, as patentes que são recompensados com empregos e promoções. Divulgar dados publicamente normalmente não contribui em nada para sua carreira, e pode até prejudicá-la, ao ajudar seus concorrentes científicos.

Em parte por essas razões, o GenBank decolou lentamente após seu lançamento em 1982. Embora muitos biólogos estivessem felizes em acessar dados de outros no GenBank, eles tinham pouco interesse em contribuir com seus próprios dados.

Mas isso mudou com o tempo. Parte do motivo da mudança foi uma conferência histórica realizada nas Bermudas em 1996, da qual participaram muitos dos principais biólogos do mundo, incluindo vários dos líderes do Projeto Genoma Humano, patrocinado pelo governo. Também estava presente Craig Venter, que mais tarde lideraria um esforço privado para sequenciar o genoma humano. Embora muitos participantes não estivessem dispostos a dar o primeiro passo unilateralmente para compartilhar todos os seus dados genéticos antes da publicação, todos podiam ver que a ciência como um todo se beneficiaria enormemente se o compartilhamento aberto de dados se tornasse uma prática comum. Então, eles se sentaram e discutiram a questão por dias, chegando finalmente a um acordo conjunto — agora conhecido como Acordo das Bermudas — de que todos os dados genéticos humanos deveriam ser imediatamente compartilhados online. O acordo não era apenas retórica vazia. Os biólogos presentes tinham influência suficiente para convencer diversas agências científicas importantes a tornar o compartilhamento imediato de dados um requisito obrigatório para o trabalho com o genoma humano. Cientistas que se recusassem a compartilhar dados não receberiam verbas para pesquisas. Isso mudou o jogo, e o compartilhamento imediato de dados genéticos humanos tornou-se a norma. O Acordo das Bermudas finalmente chegou aos mais altos níveis do governo: em 14 de março de 2000, o presidente dos EUA, Bill Clinton, e o primeiro-ministro do Reino Unido, Tony Blair, emitiram uma declaração conjunta elogiando os princípios descritos no Acordo das Bermudas e instando cientistas de todos os países a adotarem princípios semelhantes. É por causa da

Acordo das Bermudas e acordos subsequentes semelhantes de que o genoma humano e o HapMap estão disponíveis publicamente.

Esta é uma história feliz, mas tem um desfecho infeliz. O Acordo das Bermudas originalmente se aplicava apenas a dados genéticos humanos. Desde então, houve muitas tentativas de estender o espírito do acordo, para que mais dados genéticos sejam compartilhados. Mas, apesar dessas tentativas, ainda existem muitas formas de vida para as quais os dados genéticos permanecem secretos. Por exemplo, em 2010, não havia um acordo mundial para compartilhar dados sobre o vírus da gripe. Os passos em direção a tal acordo permanecem atolados em disputas entre as principais partes. Para dar uma ideia de como muitos cientistas pensam sobre compartilhar dados genéticos não humanos, um cientista me disse recentemente que estava "sentado sobre um genoma" de uma espécie inteira (!) por mais de um ano. Sem qualquer incentivo para compartilhar, e com muitas razões para não fazê-lo, os cientistas acumulam seus dados. Como resultado, há uma lacuna de dados emergente entre nossa compreensão de formas de vida como os seres humanos, onde quase todos os dados genéticos estão disponíveis online, e formas de vida como a gripe, onde dados importantes permanecem bloqueados.

Essa história dá a impressão de que os cientistas envolvidos são gananciosos e destrutivos. Afinal, essa pesquisa normalmente é paga com recursos públicos. Os cientistas não deveriam disponibilizar seus resultados o mais rápido possível? Há verdade nessas ideias, mas a situação é complexa. Para entender o que está acontecendo, é preciso entender as incríveis pressões competitivas sobre jovens cientistas ambiciosos. Nas raras ocasiões em que surge uma boa vaga de emprego de longo prazo em uma grande universidade, frequentemente há centenas de candidatos extremamente qualificados. A competição por empregos é tão acirrada que semanas de trabalho de mais de oitenta horas são comuns entre jovens cientistas. A maior parte desse tempo possível é dedicada à única coisa que garantirá esse emprego: acumular um histórico impressionante de artigos científicos. Esses artigos renderão as bolsas de pesquisa e cartas de recomendação necessárias para encontrar um emprego de longo prazo. O ritmo diminui após a estabilidade, mas o apoio contínuo às bolsas ainda exige uma forte ética de trabalho. O resultado é que, embora muitos cientistas concordem, em princípio, que adorariam compartilhar seus dados antes da publicação, eles temem que isso dê aos seus concorrentes uma vantagem injusta. Esses concorrentes poderiam explorar esse conhecimento para apressar a publicação de seus resultados ou, pior, até mesmo roubar os dados diretamente e apresentá-los como se fossem seus. Compartilhar dados só é prático se todos estiverem protegidos por um acordo coletivo, como o Acordo das Bermudas.

Um padrão semelhante tem levado cientistas a resistir a contribuir para muitos outros projetos online. Veja a Wikipédia, a enciclopédia online. Wikipédia

tem uma declaração de visão para aquecer o coração de um cientista: ***"Imagine um mundo em que cada ser humano possa compartilhar livremente a soma de todo o conhecimento. Esse é o nosso compromisso."*** Você pode pensar que a Wikipédia foi fundada por cientistas ávidos por compartilhar todo o conhecimento do mundo, mas estaria enganado. Na verdade, ela foi fundada por Jimmy "Jimbo" Wales, que na época era cofundador de uma empresa online especializada principalmente em conteúdo adulto, e Larry Sanger, um filósofo que deixou a academia para trabalhar com Wales em enciclopédias online. Nos primeiros dias da Wikipédia, houve pouco envolvimento de cientistas. Isso apesar do fato de que qualquer pessoa no mundo pode editar a Wikipédia e, de fato, ela é escrita inteiramente por seus usuários. Então, aqui está este projeto incrivelmente empolgante, no qual qualquer pessoa pode se envolver, que está decolando rapidamente e que expressa valores científicos fundamentais. Por que os cientistas não estavam se alinhando para se envolver? O problema é o mesmo dos dados genéticos: por que os cientistas perderiam tempo contribuindo para a Wikipédia quando poderiam estar fazendo algo mais respeitável entre seus pares, como escrever um artigo? Esse é o tipo de atividade que leva a empregos, bolsas e promoções. Não importa que contribuir para a Wikipédia possa ser intrinsecamente mais valioso. No início, o trabalho na Wikipédia era visto pelos cientistas como frívolo, uma perda de tempo, como se não fosse ciência séria. Fico feliz em dizer que isso mudou ao longo dos anos, e hoje o sucesso da Wikipédia legitimou, em certa medida, o trabalho científico sobre ela. Mas não é estranho que a Biblioteca de Alexandria moderna tenh

Há um enigma aqui. Cientistas ajudaram a criar a internet e a World Wide Web. Eles adotaram com entusiasmo ferramentas online como e-mail e foram pioneiros em projetos impressionantes como o Projeto Polymath e o Galaxy Zoo. Por que adotaram ferramentas como o GenBank e a Wikipédia com relutância? A razão é que, apesar de sua aparência radical, o Projeto Polymath, o Galaxy Zoo e empreendimentos semelhantes têm um conservadorismo inerente: são, em última análise, projetos a serviço do objetivo convencional de escrever artigos científicos. Esse conservadorismo os ajuda a atrair colaboradores dispostos a usar meios não convencionais, como blogs, para atingir um objetivo convencional (escrever um artigo científico) de forma mais eficaz. Mas quando o objetivo não é simplesmente produzir um artigo científico — como acontece com o GenBank, a Wikipédia e muitas outras ferramentas — não há motivação direta para os cientistas contribuírem. E isso é um problema, porque algumas das melhores ideias para melhorar a forma como os cientistas trabalham envolvem uma ruptura com o artigo científico como objetivo final da pesquisa científica. Há oportunidades perdidas que ofuscam o GenBank e a Wikipédia em seus

Impacto potencial. Neste livro, vamos nos aprofundar na história e na cultura da ciência e ver como surgiu essa situação, na qual os cientistas frequentemente relutam em compartilhar suas ideias e dados de maneiras que acelerem o avanço da ciência. A boa notícia é que encontraremos pontos de alavancagem onde pequenas mudanças hoje levarão a um futuro em que os cientistas aproveitarão ao máximo as ferramentas online, aumentando significativamente nossa capacidade de descoberta científica.

Revoluções são às vezes marcadas por um único evento espetacular: a tomada da Bastilha durante a Revolução Francesa ou a assinatura da Declaração de Independência dos Estados Unidos. Mas muitas vezes as revoluções mais importantes não são anunciadas com o toque de trombetas. Elas ocorrem silenciosamente, devagar demais para virar notícia, mas rápido o suficiente para que, se você não estiver alerta, a revolução termine antes que você perceba que está acontecendo. A mudança descrita neste livro é assim. Não é um evento único, nem é uma mudança que está acontecendo rapidamente. É uma revolução lenta que vem ganhando força silenciosamente há anos. De fato, é uma mudança que muitos cientistas perderam ou subestimaram, estando tão focados em sua própria especialidade que não percebem o quão abrangente é o impacto das novas ferramentas online. Eles são como surfistas na praia que estão tão concentrados em assistir as ondas quebrando e recuando que estão perdendo a subida da maré. Mas você não deve se deixar enganar pela natureza lenta e silenciosa das mudanças atuais na forma como a ciência é feita. Estamos no meio de uma grande mudança na forma como o conhecimento é construído. Imagine que você estivesse vivo no século XVII, no alvorecer da ciência moderna. A maioria das pessoas vivas naquela época não tinha ideia da grande transformação que estava acontecendo, uma transformação na forma como conhecemos. Mesmo que você não fosse um cientista, não gostaria de pelo menos estar ciente da notável transformação que estava ocorrendo na forma como entendíamos o mundo? Uma mudança de magnitude semelhante está acontecendo hoje: estamos reinventando a descoberta.

Escrevi este livro porque acredito que a reinvenção da descoberta é uma das grandes mudanças do nosso tempo. Para os historiadores que olham para trás daqui a cem anos, haverá duas eras da ciência: a ciência pré-rede e a ciência em rede. Vivemos na época da transição para a segunda era da ciência. Mas será uma transição acidentada, e existe a possibilidade de que fracasse ou fique aquém do seu potencial. E, por isso, também escrevi o livro para ajudar a criar uma compreensão pública amplamente compartilhada da oportunidade que agora temos diante de nós, uma compreensão de que uma abordagem mais aberta à ciência não é apenas uma boa ideia, mas que deve ser exigida de nossos cientistas e instituições científicas.

Essa mudança é importante. Aprimorar a forma como a ciência é feita significa acelerar o ritmo de todas as descobertas científicas. Significa acelerar coisas como a cura do câncer, a solução do problema das mudanças climáticas, o lançamento permanente da humanidade ao espaço. Significa insights fundamentais sobre a condição humana, sobre como o universo funciona e do que ele é feito. Significa descobertas com as quais ainda não sonhamos. Nos próximos anos, temos uma oportunidade surpreendente de mudar e aprimorar a forma como a ciência é feita. Este livro é a história dessa mudança, o que ela significa para nós e o que precisamos fazer para que aconteça.

PARTE 1

Amplificando a Inteligência Coletiva

CAPÍTULO 2

Ferramentas online nos tornam mais inteligentes

Em 1999, o campeão mundial de xadrez Garry Kasparov jogou uma partida de xadrez contra "o Mundo". Neste evento, organizado pela Microsoft, a ideia era que qualquer pessoa no mundo pudesse acessar o site do jogo e votar no próximo movimento. Em um movimento típico, mais de 5.000 pessoas votaram e, ao longo de toda a partida, 50.000 pessoas de 75 países votaram. A Equipe Mundial decidia um novo movimento a cada 24 horas e, em cada turno, o movimento realizado era o que recebesse mais votos. O jogo foi anunciado como "Kasparov contra o Mundo".

A partida superou todas as expectativas. Após 62 lances de xadrez inovador, nos quais o equilíbrio do jogo mudou diversas vezes, a Equipe Mundial finalmente desistiu. Kasparov a chamou de "a maior partida da história do xadrez" e revelou que, durante a partida, muitas vezes não conseguia distinguir quem estava ganhando e quem estava perdendo; foi somente no 51º lance que o equilíbrio pendeu decisivamente a seu favor. Após a partida, Kasparov escreveu um livro sobre o assunto, no qual comentou que despendeu mais energia nesta partida do que em qualquer outra em sua carreira, incluindo partidas de campeonatos mundiais.

Embora a Equipe Mundial tenha contado com a contribuição de alguns jogadores fortes, nenhum foi tão forte quanto o próprio Kasparov, e a qualidade média dos jogadores estava muito abaixo da de Kasparov. No entanto, coletivamente, a Equipe Mundial jogou uma partida muito mais forte do que qualquer um dos participantes jogaria normalmente — de fato, uma das partidas mais fortes da história do xadrez. Eles não apenas jogaram com Kasparov em sua melhor forma, como grande parte de suas deliberações sobre estratégia e táticas foram realizadas em público, uma vantagem que Kasparov utilizou extensivamente. Imagine não apenas jogar com Garry Kasparov, mas também ter que explicar a ele o raciocínio por trás dos seus movimentos!

Como isso foi possível? Como milhares de enxadristas, a maioria amadores, puderam competir em uma partida de xadrez com Kasparov no auge?
A Equipe Mundial era composta por pessoas de todos os níveis de habilidade no xadrez, desde iniciantes até grandes mestres. Movimentos considerados pelos especialistas como obviamente equivocados às vezes obtinham até 10% dos votos, sugerindo

que muitos iniciantes estavam participando. Em um lance, 2,4% dos votos foram para lances que não eram apenas ruins, mas que na verdade violavam as regras do xadrez!

A Equipe Mundial coordenou seu jogo de várias maneiras. A Microsoft criou um fórum onde as pessoas podiam discutir o jogo e também nomeou quatro conselheiros oficiais para a Equipe Mundial. Eram excelentes jogadores de xadrez adolescentes, entre os melhores de sua idade no mundo, embora nenhum estivesse na classe de Kasparov. A cada lance, os conselheiros publicavam suas recomendações no site da Microsoft e, se quisessem, um comentário explicando a recomendação. Isso era feito bem antes do encerramento da votação da Equipe Mundial, para que as recomendações pudessem influenciar a votação. À medida que o jogo progredia, vários outros jogadores de xadrez fortes também ofereciam conselhos. Particularmente influente, embora nem sempre acatado, era a Escola de GM, um clube de xadrez russo que contava com vários grandes mestres.

A maioria dos conselheiros e outros jogadores experientes ignorou a discussão no fórum do jogo, sem fazer qualquer tentativa de se envolver com a maioria das pessoas que compunham a Equipe Mundial, distanciando-se, assim, das pessoas cujos votos decidiam os movimentos da Equipe Mundial. Mas um dos conselheiros interagiu ativamente com a Equipe Mundial. Tratava-se de uma jovem e extraordinária jogadora de xadrez chamada Irina Krush. Com quinze anos, Krush havia se tornado recentemente campeã de xadrez feminino dos Estados Unidos. Embora não fosse tão bem avaliada quanto dois dos outros conselheiros da Equipe Mundial, Krush certamente estava na elite internacional de enxadristas juniores.

Ao contrário de seus colegas especialistas, Krush dedicou tempo e atenção consideráveis ao fórum de jogo da Equipe Mundial. Ignorando ofensas e insultos, ela extraiu muitas das melhores ideias e análises do fórum, escreveu comentários extensos descrevendo o raciocínio por trás de seus movimentos recomendados e, gradualmente, construiu uma rede de correspondentes enxadristas experientes, incluindo alguns dos grandes mestres que ofereciam conselhos.

Simultaneamente, Krush e sua equipe de gestão, uma empresa chamada Smart Chess, criaram uma árvore de análise do jogo acessível ao público, mostrando possíveis movimentos e contra-ataques, e contendo os melhores argumentos a favor e contra diferentes linhas de jogo. Esses argumentos foram extraídos não apenas de sua própria análise, mas também do fórum do jogo e de suas correspondências com outras pessoas, incluindo a Escola de Mestres. Essa árvore de análise ajudou a Equipe Mundial a coordenar seus esforços, evitou a duplicação de esforços e serviu como ponto de referência para a Equipe Mundial durante as discussões e votações.

À medida que o jogo avançava, o papel de Krush na Seleção Mundial se tornou fundamental. Parte do motivo foi a qualidade de seu jogo. No lance 10, Krush sugeriu um lance que Kasparov chamou de "um grande lance, uma contribuição importante para o xadrez", abrindo o jogo e levando-o para um território enxadrístico inexplorado. Esse movimento elevou sua posição na Equipe Mundial e a ajudou a assumir um papel de coordenação. Entre os lances 10 e 50, o lance recomendado por Krush sempre foi jogado pela Equipe Mundial, mesmo quando discordava das recomendações dos outros três conselheiros da Equipe Mundial ou de comentaristas influentes, como a Escola de Mestres.

Como resultado, algumas pessoas dizem que o jogo foi, na verdade, Kasparov versus Krush, apesar de Kasparov normalmente ter vencido Krush facilmente. O próprio Kasparov disse acreditar que estava, na verdade, jogando contra a Smart Chess, a equipe de gerenciamento de Krush. Krush rejeitou ambos os pontos de vista. Em uma série de ensaios escritos após o jogo, ela explicou o raciocínio por trás dos movimentos recomendados e como se baseou em ideias de diversas fontes, desde usuários anônimos no fórum do jogo até grandes mestres. Ela explica repetidamente como mudou e, em alguns casos, abandonou suas próprias ideias, convencida pela análise superior de outra pessoa. Assim, Krush não estava jogando sozinha nem como parte de uma pequena equipe, mas sim no centro do esforço de coordenação de toda a Equipe Mundial. Como resultado, ela teve a melhor compreensão de todas as sugestões feitas pelos membros da Equipe Mundial. Outros jogadores mais fortes não entendiam tão bem os diferentes pontos de vista e, portanto, não tomavam decisões tão acertadas sobre qual movimento fazer em seguida, nem tinham a posição na Equipe Mundial para influenciar a votação tão fortemente quanto Krush. O papel de coordenação de Krush, portanto, reunia as melhores ideias de todos os participantes em um todo coerente. O resultado foi que a Equipe Mundial emergiu muito mais forte do que qualquer jogador individual da equipe e, sem dúvida, mais forte do que qualquer jogador na história, exceto Kasparov em seu auge.

Kasparov versus o Mundo não foi a primeira partida a colocar um grande mestre do xadrez contra o Mundo. Três anos antes, em 1996, o ex-campeão mundial de xadrez Anatoly Karpov também jogou uma partida semelhante.
“Karpov Contra o Mundo” utilizou um sistema online diferente para decidir os movimentos, sem fórum de jogo ou conselheiros oficiais, e dando aos membros da Equipe Mundial apenas dez minutos para votar em seu movimento preferido. Sem os meios para coordenar suas ações, a Equipe Mundial jogou mal, e Karpov os esmagou em apenas 32 movimentos. Talvez influenciado pelo sucesso de Karpov, Kasparov admitiu que antes de sua partida ele “não estava

antecipando quaisquer dificuldades específicas", e estava confiante de que "seria capaz de terminar as coisas em menos de 40 movimentos". Como ele deve ter ficado surpreso.

Amplificando a Inteligência Coletiva

Exemplos como Kasparov versus o Mundo e o Projeto Polímata mostram que grupos podem usar ferramentas online para se tornarem coletivamente mais inteligentes. Ou seja, essas ferramentas podem ser usadas para ampliar nossa inteligência coletiva, da mesma forma que ferramentas manuais têm sido usadas há milênios para ampliar nossa força física. Como essas novas ferramentas alcançam esse feito incrível? Seria apenas um acaso? Ou as ferramentas online podem ser usadas de forma mais geral para resolver problemas criativos que desafiam a engenhosidade até mesmo dos indivíduos mais inteligentes? Existem princípios gerais de design que podem ser usados para ampliar a inteligência coletiva, uma espécie de ciência do design da colaboração?

Uma abordagem comum a essas questões é sugerir que as ferramentas online possibilitam algum tipo de cérebro coletivo, com as pessoas do grupo desempenhando o papel de neurônios. Uma inteligência maior, então, de alguma forma, emerge das conexões entre esses neurônios humanos. Embora essa metáfora seja estimulante, ela apresenta muitos problemas. A origem e o hardware do cérebro são completamente diferentes daqueles da internet, e não há nenhuma razão convincente para supor que o cérebro seja um modelo preciso de como a inteligência coletiva funciona, ou de como ela pode ser melhor amplificada.
Seja lá o que for que nosso cérebro coletivo esteja fazendo, parece provável que funcione de acordo com princípios muito diferentes dos do cérebro dentro de nossas cabeças. Além disso, ainda não temos uma boa compreensão de como o cérebro humano funciona, então a metáfora é, em qualquer caso, de uso limitado, na melhor das hipóteses. Se quisermos entender como ampliar a inteligência coletiva, precisamos olhar além da metáfora do cérebro coletivo.

Muitos livros e artigos de revistas já foram escritos sobre inteligência coletiva. Talvez o exemplo mais conhecido desse trabalho seja o livro de James Surowiecki, " ***A Sabedoria das Multidões", de 2004,*** que explica como grandes grupos de pessoas podem, às vezes, ter um desempenho surpreendentemente bom na resolução de problemas. Surowiecki abre seu livro com uma história marcante sobre o cientista Francis Galton. Em 1906, Galton frequentava uma escola de inglês.

feira rural, e entre as atrações da feira estava uma competição de julgamento de peso, onde as pessoas competiam para adivinhar o peso de um boi. Galton esperava que a maioria dos competidores errasse bastante em suas estimativas e ficou surpreso ao descobrir que a média de todos os palpites dos competidores (1.197 libras) estava apenas uma libra abaixo do peso correto de 1.198 libras. Em outras palavras, coletivamente, se alguém fizesse a média dos palpites, o público na feira chutaria o peso quase perfeitamente. O livro de Surowiecki continua discutindo muitas outras maneiras pelas quais podemos combinar nossa sabedoria coletiva com resultados surpreendentemente bons.

Este livro vai além de ***"A Sabedoria das Multidões"*** e obras semelhantes em dois aspectos. Primeiro, nosso objetivo é entender como as ferramentas online podem ***amplificar*** ativamente a inteligência coletiva. Ou seja, não estamos interessados apenas na inteligência coletiva em si, mas em como projetar ferramentas que aumentem drasticamente a inteligência coletiva. Segundo, não estamos discutindo apenas problemas cotidianos, como estimar o peso de um boi.
Em vez disso, nosso foco está em problemas que estão no limite da capacidade humana de resolução de problemas, problemas como competir com Garry Kasparov no auge de sua habilidade no xadrez, ou resolver problemas matemáticos que desafiam os melhores matemáticos do mundo. Nosso principal interesse será a resolução de problemas científicos e, claro, são os problemas que estão no limite da capacidade humana de resolução de problemas que os cientistas mais desejam resolver, e cuja solução trará o maior benefício.

Superficialmente, a ideia de que ferramentas online podem nos tornar coletivamente mais inteligentes contradiz a ideia, atualmente em voga em alguns círculos, de que a internet está reduzindo nossa inteligência. Por exemplo, em 2010, o autor Nicholas Carr publicou um livro intitulado ***The Shallows: What the Internet Is Doing to Our Brains (O Superficial: O que a Internet Está Fazendo com Nossos Cérebros),*** argumentando que a internet está reduzindo nossa capacidade de concentração e contemplação. O livro de Carr e outras obras semelhantes apresentam muitos pontos positivos e têm sido amplamente discutidos. Mas as novas tecnologias raramente têm apenas um único impacto, e não há contradição em acreditar que ferramentas online podem tanto aumentar quanto reduzir a inteligência. Você pode usar um martelo para construir uma casa; você também pode usá-lo para quebrar o polegar. Tecnologias complexas, especialmente, muitas vezes exigem habilidade considerável para serem bem utilizadas. Automóveis são ferramentas incríveis, mas todos sabemos como motoristas iniciantes podem aterrorizar as ruas. Olhar para a internet e concluir que o principal impacto é nos tornar estúpidos é como olhar para o automóvel e concluir que ele é uma ferramenta para motoristas iniciantes atropelarem pedestres aterrorizados. Online, ainda somos todos aprendizes de direção, e não é de se surpreender que as ferramentas online sejam às vezes mal utilizadas, amplificando nossa estupidez individual e coletiva. Mas, como já vimos,

Como visto, também há exemplos que mostram que ferramentas online podem ser usadas para aumentar nossa inteligência coletiva. Nossa preocupação, portanto, será entender como essas ferramentas podem ser usadas para nos tornar coletivamente mais inteligentes e o que essa mudança significará.

Ainda estamos engatinhando na compreensão de como ampliar a inteligência coletiva. É revelador que muitas das melhores ferramentas que temos — ferramentas como blogs, wikis e fóruns online — não tenham sido inventadas por pessoas que poderíamos supor serem especialistas em comportamento e inteligência de grupo, especialistas de áreas como psicologia de grupo, sociologia e economia. Em vez disso, foram inventadas por amadores, pessoas como Matt Mullenweg, que era um estudante de 19 anos quando criou o Wordpress, um dos softwares de blog mais populares, e Linus Torvalds, que era um estudante de 21 anos quando criou o sistema operacional Linux de código aberto. Isso nos diz que devemos ser cautelosos com a teoria atual: embora possamos aprender muito com os estudos acadêmicos existentes, o quadro de inteligência coletiva que emerge também é incompleto. Por esse motivo, basearemos nossa discussão em exemplos concretos nos moldes do Projeto Polímata e de Kasparov versus o Mundo. Na parte 1 deste livro, usaremos esses exemplos concretos para destilar um conjunto de princípios que explicam como as ferramentas online podem ampliar a inteligência coletiva.

Concentrei deliberadamente a discussão da Parte 1 em um número relativamente pequeno de exemplos, com a ideia de que, à medida que desenvolvemos uma estrutura conceitual para a compreensão da inteligência coletiva, revisitaremos cada um desses exemplos diversas vezes e os compreenderemos mais profundamente. Além disso, os exemplos não vêm apenas da ciência, mas também de áreas como xadrez e programação de computadores.
A razão é que alguns dos exemplos mais marcantes de amplificação da inteligência coletiva — exemplos como Kasparov versus o Mundo — vêm de fora da ciência, e podemos aprender muito estudando-os.

À medida que nossa compreensão se aprofunda, veremos que os problemas científicos são especialmente adequados para o ataque da inteligência coletiva e, na parte 2, estreitaremos nosso foco para como a inteligência coletiva está mudando a ciência.

CAPÍTULO 3

Atenção de Especialistas em Reestruturação

Em 2003, uma jovem chamada Nita Umashankar, de Tucson, Arizona, foi morar por um ano na Índia, onde trabalhou com uma organização sem fins lucrativos para ajudar profissionais do sexo a escapar do comércio sexual. O que ela encontrou na Índia a frustrou. Muitas das profissionais do sexo tinham tão poucas habilidades que era quase impossível ajudá-las a encontrar empregos fora da prostituição. Ao retornar aos Estados Unidos, Umashankar decidiu fundar uma organização sem fins lucrativos que abordasse o problema central, treinando meninas indianas em situação de risco em tecnologia e, em seguida, ajudando-as a encontrar empregos em empresas de tecnologia.

Oito anos depois, a organização sem fins lucrativos que ela fundou, a ASSET Índia, abriu centros de treinamento em tecnologia em cinco cidades indianas. Eles ajudaram centenas de jovens a escapar do comércio sexual e têm planos de expansão. Infelizmente, muitas das cidades menores para as quais gostariam de se expandir não têm a eletricidade confiável necessária para alimentar tecnologias cruciais, como os roteadores sem fio usados para acessar a internet.

A ASSET fez experiências com o uso de roteadores sem fio alimentados por energia solar, mas descobriu que os dispositivos já existentes no mercado não funcionam de forma confiável durante as longas horas em que seus centros de treinamento ficam abertos.

Para resolver o problema com roteadores sem fio, a ASSET tentou algo não convencional: buscou ajuda em um mercado online para problemas científicos chamado InnoCentive. O InnoCentive é como o eBay ou o Craigslist, mas voltado para problemas científicos. A ideia é que as organizações participantes possam publicar "Desafios" online — problemas científicos que desejam resolver — com prêmios para quem os solucionar, geralmente de dezenas de milhares de dólares.

Qualquer pessoa no mundo pode baixar uma descrição detalhada de um Desafio, tentar resolver o problema e ganhar o prêmio.

Usando US$ 20.000 em prêmios oferecidos pela Fundação Rockefeller, a ASSET lançou um Desafio InnoCentive para projetar um roteador sem fio confiável alimentado por energia solar, utilizando hardware e software de baixo custo e facilmente disponíveis. Nos dois meses em que o Desafio foi publicado na InnoCentive, ele foi baixado 400 vezes e 27 soluções foram enviadas.

Um prêmio de US$ 20.000 foi concedido a um engenheiro de software texano de 31 anos chamado Zacary Brown, e um protótipo está sendo construído por estudantes de engenharia da Universidade do Arizona.

Zacary Brown não era um engenheiro de software qualquer. Um entusiasmado operador de rádio sem fio amador, ele trabalhava com o objetivo de estabelecer contato via rádio com todos os países do mundo. Enquanto crescia, ele se encantava com a explicação de seus pais sobre como os painéis solares que Jimmy Carter instalou na Casa Branca geravam eletricidade a partir da luz solar e, já adulto, estava experimentando o uso de painéis solares para alimentar seu equipamento de rádio sem fio. A longo prazo, ele esperava abastecer todo o seu escritório em casa com energia solar. Em suma, Zacary Brown era exatamente a pessoa certa para a ASSET conversar. A InnoCentive simplesmente forneceu uma maneira de fazer a conexão.

A InnoCentive parte da premissa de que existe um enorme potencial inexplorado para descobertas científicas no mundo, potencial que pode ser liberado conectando as pessoas certas. Essa premissa foi confirmada com mais de 160.000 inscritos de 175 países no InnoCentive e prêmios concedidos em mais de 200 Desafios. Os Desafios abrangem diversas áreas da ciência e tecnologia.

Exemplos incluem encontrar métodos mais econômicos para fabricar medicamentos para tuberculose, desenvolver um repelente de mosquitos movido a energia solar (não estou inventando!) para combater a malária e encontrar melhores maneiras de identificar pessoas em risco de desenvolver doenças do neurônio motor. Muitos dos solucionadores bem-sucedidos relatam, como Zacary Brown, que os desafios que resolvem correspondem perfeitamente às suas habilidades e interesses. Além disso, como na história da ASSET, conexões geralmente são feitas entre partes que, de outra forma, só teriam se conhecido acidentalmente. A InnoCentive faz essas conexões sistematicamente, não como eventos pontuais e de sorte, mas em larga escala.

A razão pela qual as conexões feitas pela InnoCentive são tão valiosas é, obviamente, a grande lacuna entre as habilidades das pessoas que propõem os Desafios e as que os resolvem. Enquanto projetar um roteador sem fio alimentado por energia solar pode levar apenas alguns dias para um especialista como Zacary Brown, levaria meses ou anos para o pessoal da ASSET Índia.
Eles simplesmente não têm a expertise necessária. É porque Zacary Brown tem uma vantagem comparativa tão grande que ele e a ASSET podem trabalhar juntos em benefício mútuo. De forma mais geral, a atenção do especialista certo no momento certo costuma ser o recurso mais valioso que se pode ter na resolução criativa de problemas. A atenção especializada é para a resolução criativa de problemas o que a água é para a vida no deserto: é o recurso escasso fundamental. A InnoCentive cria valor ***reestruturando a atenção especializada,*** portanto

que pessoas como Zacary Brown podem usar sua expertise de maneiras altamente alavancadas: a InnoCentive ajudou Zacary Brown a concentrar sua expertise no problema da ASSET, em vez de trabalhar em casa em seus hobbies.

Neste capítulo, veremos que é essa capacidade de reestruturar a atenção dos especialistas que está no cerne de como as ferramentas online amplificam a inteligência coletiva. O que exemplos como InnoCentive, o Projeto Polymath e Kasparov versus o Mundo compartilham é a capacidade de atrair a atenção do especialista certo para o problema certo, no momento certo. Na primeira metade do capítulo, examinaremos esses exemplos com mais detalhes e desenvolveremos uma ampla estrutura conceitual que explica como eles reestruturam a atenção dos especialistas. Na segunda metade do capítulo, aplicaremos essa estrutura para entender como as colaborações online podem funcionar em conjunto de maneiras essencialmente diferentes das colaborações offline.

Aproveitando a microespecialização latente

Embora a história da ASSET-InnoCentive seja marcante, Kasparov versus o Mundo é um exemplo ainda mais impressionante de inteligência coletiva. Assim como na história da ASSET-InnoCentive, Kasparov versus o Mundo contou com uma reestruturação da atenção especializada. Para entender como isso funcionou, vamos retornar ao lance decisivo sugerido por Irina Krush, o lance número 10, o lance que Kasparov elogiou como "um grande lance, uma importante contribuição para o xadrez". A sugestão de Krush não surgiu do nada. Ela teve a ideia para o lance 10 um mês antes do início de Kasparov versus o Mundo, durante uma sessão de estudos no torneio de xadrez World Open na Filadélfia. Na época, ela fez uma breve análise e discutiu o lance com seus treinadores, os grandes mestres Giorgi Kacheishvili e Ron Henley, antes de deixar a ideia de lado. Foi uma sorte que Kasparov versus o Mundo tenha aberto de uma forma que permitiu a Krush usar o lance que ela vinha considerando na Filadélfia. Certamente não era algo que ela pudesse controlar completamente, pois Kasparov estava jogando com as peças brancas e, portanto, jogando primeiro, o que lhe permitia ditar a direção inicial do jogo. Ainda assim, uma semana antes do lance 10, Krush e seus treinadores estavam cientes da possibilidade de que o jogo pudesse tomar essa direção e começaram a analisar os prós e os contras da ideia de Krush com mais atenção.

É importante reconhecer que, em quase todos os aspectos, Kasparov era de longe superior a Krush como jogador de xadrez. Podemos expressar a diferença entre eles com bastante precisão, visto que existe um sistema de classificação numérica usado para classificar jogadores de xadrez. Nesse sistema de classificação, um bom jogador de clube terá uma classificação na faixa de 1.800 a 2.000. Uma mestre internacional como Irina Krush terá uma classificação em torno de 2.400. Em 1999, na época de sua partida contra o Mundo, a classificação de Kasparov atingiu o pico de 2.851 — não apenas a classificação mais alta da história do xadrez, mas consideravelmente mais alta do que a classificação de qualquer outro jogador antes ou depois. A diferença de 450 pontos de classificação entre Kasparov e Krush era praticamente a mesma que a diferença entre Krush e um bom jogador de clube. Isso significava que Krush só teria chance de vencer uma partida contra Kasparov se cometesse um erro grave. Isso não quer dizer que Krush fosse uma jogadora fraca — lembre-se, ela era campeã feminina dos EUA — mas na época do jogo Kasparov estava em outra classe.

Dada a grande diferença de habilidade entre Kasparov e Krush, parece muito positivo que a partida tenha se desenrolado de forma a dar a Krush a chance de explorar sua expertise extremamente especializada na abertura que levou ao lance 10. Nessa pequena fatia do xadrez, ela era superior a Kasparov e poderia dar vantagem à Equipe Mundial. Em outras palavras, embora Krush fosse inferior a Kasparov em quase todas as áreas do xadrez, nessa área específica de ***microespecialização*** ela superou até mesmo Kasparov.

Mas, embora tenha sido sorte que a microespecialização específica de Krush tenha ajudado a Equipe Mundial a obter vantagem, isso não significa que tenha sido simplesmente sorte que permitiu à Equipe Mundial jogar tão bem. O jogo foi amplamente divulgado na comunidade enxadrística, e centenas de enxadristas experientes o acompanhavam. O xadrez é tão rico em possíveis variações que muitos desses jogadores tinham suas próprias áreas individuais de microespecialização, nas quais também se igualavam ou até superavam Kasparov. A chave para o jogo da Equipe Mundial era garantir que toda essa microespecialização normalmente latente fosse descoberta e aplicada em resposta às contingências da partida. Portanto, embora tenha sido uma sorte que Krush, ***em particular,*** tenha sido a pessoa cuja microespecialização foi decisiva no lance 10, dado o número de enxadristas experientes envolvidos, era altamente provável que a microespecialização latente desses jogadores viesse à tona em momentos críticos da partida, ajudando a Equipe Mundial a se igualar a Kasparov.

Foi exatamente isso que aconteceu. Por exemplo, após o término do jogo, Krush destacou o lance número 26 como um dos seus três lances favoritos do Time Mundial. O lance número 26 não foi ideia de Krush, nem de

um dos especialistas em xadrez consagrados que acompanharam a partida. Em vez disso, o lance 26 foi proposto por um dos participantes do fórum do jogo, usando o nome Yasha, que mais tarde se revelou ser Yaaqov Vaingorten, um jogador júnior razoavelmente sério, mas não de elite. Isso fazia parte de um padrão, já que durante a partida Krush se baseou extensivamente no pensamento de muitos contribuidores desconhecidos ou mesmo anônimos do fórum do jogo, pessoas que usavam pseudônimos como Agente Scully, Solnushka e Alekhine via Ouji. Ao mesmo tempo, ela também consultou jogadores de xadrez consagrados, como os mestres internacionais Ken Regan e Antti Pihlajasalo, e o grande mestre Alexander Khalifman, da Escola de Mestres de Jogo. A Equipe Mundial não teve sorte alguma. Em vez disso, a Equipe Mundial tinha um conjunto tão diverso de talentos disponíveis que, cada vez que surgia um problema, um membro da equipe se destacava; alguém com a microespecialização certa surgia para preencher a lacuna.

Serendipidade projetada

Vimos como projetos colaborativos como Kasparov versus the World e InnoCentive utilizam microespecialização latente para superar desafios que frustrariam a maioria dos membros da colaboração. Nas colaborações online mais bem-sucedidas, esse uso de microespecialização se aproxima de um ideal no qual a colaboração ***rotineiramente*** localiza pessoas como Yasha e Zacary Brown, pessoas com a microespecialização certa para a ocasião. Em particular, à medida que a colaboração criativa se expande, os problemas podem ser expostos a pessoas com uma gama cada vez maior de expertise, aumentando significativamente a chance de alguém ver o que parece ser um problema difícil para a maioria dos participantes e pensar: "Ei, isso é fácil de resolver". Em vez de ser uma coincidência fortuita ocasional, a serendipidade se torna comum. A colaboração alcança um tipo de ***serendipidade projetada,*** um termo que adaptei do autor Jon Udell.

Para compreender o valor dessa serendipidade no trabalho criativo, é útil ter um exemplo histórico concreto. Vejamos o trabalho de Einstein sobre sua maior contribuição para a ciência, sua teoria da gravidade, frequentemente chamada de

Teoria da Relatividade Geral. Ele trabalhou intermitentemente no desenvolvimento da relatividade geral entre 1907 e 1915, frequentemente enfrentando grandes dificuldades. Em 1912, seu trabalho o levou à surpreendente conclusão de que nossa concepção comum da geometria do espaço, na qual os ângulos de um triângulo somam 180 graus, é apenas aproximadamente correta, e um novo tipo de geometria é necessário para descrever o espaço e o tempo. Agora, caso você esteja se perguntando o que a geometria do espaço e do tempo tem a ver com a gravidade, você está em boa companhia: foi uma surpresa para Einstein também.
Ao se preparar para entender a gravidade, Einstein não imaginava que acabaria pensando nela como um problema geométrico. No entanto, lá estava ele em 1912 com a ideia de que a gravidade estava de alguma forma conectada a um tipo não padrão de geometria. E ele estava travado, porque tais ideias geométricas estavam fora de sua especialidade. Ele conversou sobre seus problemas com um amigo matemático de longa data, Marcel Grossmann, dizendo-lhe: "Grossmann, você precisa me ajudar, senão eu vou enlouquecer!" Felizmente, para Einstein, Grossmann era a pessoa certa para conversar. Ele disse a Einstein que as ideias geométricas de que Einstein precisava já haviam sido totalmente desenvolvidas, décadas antes, pelo matemático Bernhard Riemann. Einstein rapidamente mergulhou na geometria riemanniana e percebeu que Grossmann estava certo. A geometria riemanniana se tornou a linguagem matemática da relatividade geral.

Conexões fortuitas como essa são cruciais no trabalho criativo. Na ciência, especialmente, todo cientista ativo carrega na cabeça uma série de problemas não resolvidos. Alguns desses problemas são grandes ("Descubra como o universo começou"), alguns são pequenos ("Onde aquele maldito sinal de menos desapareceu no meu cálculo?"), mas todos eles são matéria-prima para o progresso futuro. Se você é um cientista, depende principalmente de você resolver esses problemas sozinho. Se tiver sorte, poderá ter alguns colegas que o apoiam e que podem ajudá-lo. Às vezes, porém, você resolverá um problema de uma maneira completamente diferente. Você estará conversando com um conhecido quando um dos seus problemas surgir. Você está conversando quando, BANG, de repente você percebe que essa é exatamente a pessoa certa para conversar. Às vezes, ela pode simplesmente resolver o seu problema. Ou às vezes, ela lhe dá algum insight ou ideia crucial que fornece o impulso necessário para vencê-lo. Esse tipo de conexão fortuita é um dos momentos mais emocionantes e importantes da ciência. O problema é que essas conexões casuais ocorrem muito raramente. A razão pela qual a serendipidade planejada é importante é porque, no trabalho criativo, a maioria de nós — até mesmo Einstein! — passa grande parte do nosso tempo bloqueada por problemas que seriam rotineiros, se

Só nós poderíamos encontrar o especialista certo para nos ajudar. Há apenas 20 anos, encontrar o especialista certo provavelmente seria difícil. Mas, como mostram exemplos como InnoCentive e Kasparov versus o Mundo, agora podemos projetar sistemas que tornem isso uma rotina. A serendipidade projetada nos permite resolver de forma rápida e rotineira muitos desses problemas antes insolúveis, expandindo assim o alcance da nossa capacidade de resolução de problemas.

Massa Crítica Conversacional

É desafiador transmitir a experiência da serendipidade projetada. Uma coisa é descrever exemplos, mas outra coisa é fazer parte de uma colaboração onde a serendipidade projetada está realmente acontecendo. De repente, você sente como se sua mente tivesse ganhado asas. Você se liberta de grande parte do fardo de problemas incômodos, problemas que seriam rotineiros se você tivesse acesso a um especialista com as habilidades certas. É profundamente prazeroso, em vez disso, dedicar seu tempo concentrando-se nos problemas em que você tem uma visão e uma vantagem especiais. A serendipidade projetada é algo que precisa ser vivenciado para ser totalmente compreendido.

Dito isso, existe um modelo simples que pode ajudar a explicar por que a serendipidade planejada é importante e como ela pode mudar qualitativamente a natureza da colaboração. Esse modelo é uma reação em cadeia nuclear. Ao nos lembrarmos do que acontece durante uma reação em cadeia, entenderemos por que a serendipidade planejada é importante.

O funcionamento de uma reação em cadeia é simples. Imagine que você, de alguma forma, tenha obtido um pequeno pedaço de urânio — urânio-235, o tipo de urânio usado em bombas nucleares. (Existem vários tipos de urânio, mas nem todos sofrem reações nucleares em cadeia. De agora em diante, quando digo "urânio", quero dizer urânio-235.) Os átomos de urânio, ao que parece, não são muito estáveis. De vez em quando, o núcleo de um átomo de urânio se desintegra, expelindo um ou mais nêutrons. Esse nêutron então voa através do pedaço de urânio. O urânio, como todos os sólidos, só parece sólido ao olho humano. Na verdade, no nível atômico, é principalmente espaço vazio, e o nêutron pode viajar uma longa distância antes de encontrar o núcleo de outro átomo de urânio. Em um pequeno pedaço de urânio - digamos, meio quilo (cerca de uma libra) - as chances são muito boas de que o nêutron nunca encontre outro núcleo e, em vez disso, voe por todo o

para fora do pedaço de urânio e simplesmente seguir em frente. Mas se o pedaço de urânio for um pouco maior — digamos, um quilograma — as chances são bem maiores de que o nêutron colida com o núcleo de outro átomo de urânio. Esse núcleo então se desintegra e, ao que parece, libera mais três nêutrons. Agora, há quatro nêutrons passando pelo urânio — são quatro porque precisamos incluir em nossa contagem o nêutron original que iniciou o processo, que continua a se mover, mesmo depois de colidir com o núcleo. Cada um desses nêutrons, por sua vez, provavelmente colidirá com outros quatro núcleos, resultando em 16 nêutrons soltos. Eles provavelmente colidirão com ainda mais núcleos, e as coisas rapidamente saem do controle: após 40 colisões como essa, temos um trilhão de trilhões de nêutrons voando. É por causa dessa taxa de crescimento incrivelmente rápida que o processo é chamado de reação em cadeia. Abaixo de uma certa massa, chamada massa crítica, um pedaço de urânio é simplesmente um pedaço inerte de rocha. Os átomos em seu interior ocasionalmente decaem e liberam nêutrons, mas para cada um desses nêutrons, o número médio dos chamados nêutrons-filhos causados por colisões posteriores é menor que um, e qualquer possível reação em cadeia se extingue rapidamente. Mas com um pedaço de urânio apenas um pouco maior, maior que a massa crítica, o número médio de nêutrons-filhos é ligeiramente maior que um. E se o número médio de nêutrons-filhos for mesmo um pouquinho maior que um, a reação em cadeia decolará e se espalhará descontroladamente. Se o número médio de nêutrons-filhos for 1,1, então, após apenas 200 colisões, o urânio terá mais de 100 milhões de nêutrons voando em seu interior, causando ainda mais colisões. É por isso que dois pedaços de urânio aparentemente semelhantes se comportarão de maneiras completamente diferentes. Um permanecerá inerte, enquanto outro pedaço um pouco maior explodirá com a força de milhares de toneladas de dinamite. Um pequeno aumento no tamanho pode causar uma mudança qualitativa completa no comportamento.

Algo semelhante acontece em uma boa colaboração criativa. Quando tentamos resolver um problema criativo difícil sozinhos, a maioria das nossas ideias não leva a lugar nenhum. Mas em uma boa colaboração criativa, algumas das nossas ideias — ideias que não poderíamos ter levado adiante sozinhos — estimulam outras pessoas a criarem suas próprias ideias filhas. Essas, por sua vez, estimulam outras pessoas a criarem ainda mais ideias. E assim por diante.

Idealmente, alcançamos uma espécie de massa crítica conversacional, onde a colaboração se torna autoestimulante e obtemos o benefício mútuo da conexão serendipitária repetidas vezes. É essa transição que é possibilitada pela serendipidade projetada, e é por isso que a experiência da serendipidade projetada é tão diferente da colaboração comum.

ocorre quando a colaboração é ampliada, aumentando o número e a diversidade de participantes e, assim, aumentando a chance de uma ideia estimular outra nova. No Projeto Polymath, por exemplo, Tim Gowers comentou que o principal fator que acelerou o processo foi que ele e outros participantes frequentemente "se viam tendo pensamentos que não teriam tido sem algum comentário casual de outro colaborador". Em Kasparov versus o Mundo, a mesma coisa aconteceu, com uma ideia de um membro da equipe frequentemente gerando ideias de outros, permitindo que a Equipe Mundial explorasse muitas direções diferentes.

É claro que o modelo de reação em cadeia não deve ser levado tão literalmente como um modelo de colaboração. Ideias não são nêutrons, e o objetivo da colaboração não é simplesmente atingir o ponto crítico, produzindo um número crescente de ideias. Precisamos, pelo menos ocasionalmente, ter as ideias certas, ideias que realmente nos aproximem de uma solução para o nosso problema. É possível que em algum ponto do problema que está sendo abordado haja um gargalo, exigindo algum insight-chave que ninguém na colaboração está pronto para ter. Ainda assim, o modelo de reação em cadeia transmite bem a mudança qualitativa que ocorre quando uma colaboração "atinge o ponto crítico", quando a serendipidade projetada faz com que o número de ideias geradas em uma colaboração salte tanto que o processo se torna autossustentável. Esse salto muda qualitativamente a forma como resolvemos problemas, levando-nos a um nível novo e superior.

Amplificando a Inteligência Coletiva

Vamos fazer um balanço do panorama da inteligência coletiva que estamos desenvolvendo. Ele parte da ideia de que, em grandes grupos, pode haver uma quantidade enorme de expertise, muito maior do que a disponível em qualquer indivíduo do grupo. Idealmente, esses grupos são extremamente diversos cognitivamente — o que significa que possuem uma ampla gama de expertises sem sobreposição —, mas seus membros têm o suficiente em comum para que possam se comunicar de forma eficaz.

Normalmente, a maior parte dessa expertise é latente. Um bom, mas não excelente, jogador de xadrez pode ter áreas individuais de microexpertise nas quais se iguala ou supera os melhores enxadristas do mundo, mas em uma partida de xadrez comum isso não é suficiente para compensar as muitas áreas em que ele é inferior.

Mas se o grupo for grande o suficiente e cognitivamente diverso o suficiente, as ferramentas certas podem permitir que o grupo aproveite essa microespecialização quando necessário, superando em muito o talento de qualquer indivíduo. A serendipidade planejada pode se consolidar, resultando em uma massa crítica conversacional que explora rapidamente um espaço de ideias muito maior do que qualquer indivíduo conseguiria sozinho.

Subjacente a esse panorama amplo está o fato de que, coletivamente, sabemos muito mais do que até mesmo os indivíduos mais brilhantes. Séculos atrás, talvez fosse possível para um único indivíduo brilhante — um Aristóteles, Hipátia ou Leonardo — superar todos os outros em muitas áreas do conhecimento. Hoje, o conhecimento humano se expandiu tanto que isso não é mais possível. O conhecimento foi descentralizado e agora está distribuído entre muitas mentes. Mesmo as pessoas mais brilhantes, como os matemáticos Tim Gowers e Terence Tao e o jogador de xadrez Garry Kasparov, têm um domínio insuperável de apenas uma pequena fração do nosso conhecimento. Mesmo dentro de suas áreas de especialização, elas são frequentemente superadas em aspectos especializados por outras pessoas, pessoas com áreas específicas de microespecialização. Ao reestruturar a atenção especializada, as ferramentas online podem permitir que essa microespecialização seja aplicada quando e onde for mais nec

Com esse ponto de vista em mente, vemos que o problema de amplificar a inteligência coletiva é direcionar a microespecialização para onde ela será mais útil. O objetivo das ferramentas online é ajudar as pessoas a descobrir para onde devem direcionar sua atenção. Quanto melhor as ferramentas direcionarem a atenção das pessoas, mais bem-sucedida será a colaboração. Em outras palavras, as ferramentas online criam uma ***arquitetura de atenção*** cujo propósito é ajudar os participantes a encontrar tarefas nas quais tenham a maior vantagem comparativa. Idealmente, essa arquitetura de atenção direcionará a atenção do especialista certo para os problemas certos. Quanto mais eficazmente a atenção do especialista for alocada dessa maneira, mais eficazmente os problemas poderão ser resolvidos. (Veja as notas finais para a discussão da ideia relacionada da arquitetura da ***participação,*** sugerida pelo especialista em tecnologia Tim O'Reilly.) Essa visão da inteligência coletiva está resumida no quadro Resumo e Prévia, que também apresenta uma prévia de muitas das ideias sobre amplificação da inteligência coletiva desenvolvidas no restante da Parte 1.

Resumo e Prévia: Como Amplificar a Inteligência Coletiva

Para ampliar a coletividade, devemos ampliar as colaborações, aumentando ao máximo a diversidade cognitiva e a gama de conhecimentos disponíveis. Isso amplia a gama de problemas que podem ser facilmente resolvidos. O desafio de ampliar a colaboração é que cada participante

tem apenas uma quantidade limitada de atenção para dedicar à colaboração.
Isso limita o volume de contribuições à colaboração às quais qualquer participante pode prestar atenção. Para ampliar a colaboração, respeitando essa limitação, as ferramentas online devem estabelecer uma arquitetura de atenção que direcione a atenção de cada participante para onde ela é mais adequada — ou seja, onde eles têm a máxima vantagem comparativa.
Idealmente, a colaboração alcançará a serendipidade planejada, de modo que um problema que parece difícil para quem o propõe encontre o caminho para alguém com a microespecialização necessária para resolvê-lo facilmente (ou estimular o progresso).

A massa crítica conversacional é alcançada e a colaboração se torna autoestimulante, com novas ideias sendo constantemente exploradas. No próximo capítulo, Capítulo 4, veremos muitos padrões colaborativos que podem ajudar a atingir esses objetivos, incluindo:

- Modularizar a colaboração, ou seja, descobrir maneiras de dividir a tarefa geral em subtarefas menores que possam ser abordadas de forma independente ou quase independente. Isso reduz as barreiras à entrada de novas pessoas e, assim, amplia o leque de expertise disponível. A modularidade costuma ser difícil de alcançar, exigindo um comprometimento consciente e incansável por parte dos participantes.

- Incentivar pequenas contribuições, novamente para reduzir as barreiras de entrada e ampliar a gama de conhecimentos especializados disponíveis.

- Desenvolver um acervo de informações rico e bem estruturado, para que as pessoas possam desenvolver trabalhos anteriores. Quanto mais fácil for encontrar e reutilizar trabalhos anteriores, mais rápido o acervo de informações crescerá.

No capítulo 5, examinaremos os limites da inteligência coletiva. Descobriremos que, para que a inteligência coletiva seja bem-sucedida, os participantes devem estar comprometidos com um conjunto compartilhado de métodos de raciocínio, para que os desacordos entre os participantes possam ser resolvidos e não causem rupturas permanentes.
Esse conjunto compartilhado de métodos está disponível em áreas como xadrez, programação e ciências, mas nem sempre em outras áreas. Por exemplo, artistas podem estar fundamentalmente divididos quanto a princípios estéticos básicos.
Essas divisões impedirão que a colaboração seja ampliada e, portanto, impedirão a serendipidade projetada e a massa crítica conversacional.

Como a colaboração online vai além Organizações Convencionais

Usar inteligência coletiva para resolver problemas não é novidade. Historicamente, os grupos têm utilizado três maneiras principais de resolver problemas criativos: (1) grandes organizações formais, como as centenas ou milhares de pessoas que podem estar envolvidas na criação de um filme, por exemplo, ou de um novo gadget eletrônico; (2) o sistema de mercado; e (3) conversas em pequenos grupos informais. No restante deste capítulo, investigaremos como as ferramentas online podem nos levar além dessas três maneiras existentes de resolver problemas em grupo.

Para entender como a colaboração online vai além das organizações convencionais, considere uma produção cinematográfica. Um filme de sucesso moderno pode empregar centenas ou até milhares de pessoas — o filme ***Avatar,*** de 2009 , empregou 2.000 pessoas. Mas, ao contrário dos participantes de Kasparov contra o Mundo ou do Projeto Polímata, cada funcionário tem seu próprio papel na produção. Um funcionário do departamento de arte do filme normalmente não dá conselhos a um violinista da orquestra.

No entanto, foi exatamente esse o tipo de tomada de decisão que ocorreu em Kasparov versus o Mundo. Lembre-se do movimento crítico número 26 sugerido por Yasha. Em termos cinematográficos, era como se um estranho desconhecido tivesse entrado no set, feito uma sugestão crucial ao diretor, mudando completamente o curso do filme, e depois sumido.

Claro, existem histórias assim no cinema. O ator Mel Gibson teve sua grande chance quando um amigo que estava fazendo o teste para o filme ***Mad Max*** pediu para ser levado de carro. Gibson não estava fazendo o teste, mas havia se envolvido em uma briga em uma festa na noite anterior e estava com hematomas por todo o rosto. O agente de elenco decidiu que aquele era o visual que o filme precisava, e Gibson foi convidado novamente, mudando completamente o filme e lançando-o no caminho para o estrelato internacional.

No mundo do cinema, esta é uma história incomum. Mas em Kasparov versus o Mundo, esse tipo de ocorrência não foi um acaso, mas sim a essência da forma como a Seleção Mundial jogava. Não havia uma divisão de trabalho pré-planejada e estática, como em uma organização convencional. Em vez disso, havia uma divisão de trabalho dinâmica, na qual cada jogador da Seleção Mundial tinha a oportunidade, pelo menos em princípio, de se envolver em cada mover.

Deixe-me ser mais preciso sobre o que quero dizer com uma divisão dinâmica de trabalho. É uma divisão de trabalho em que todos os participantes de uma colaboração podem responder aos problemas em questão, à medida que surgem. Zacary Brown viu o problema da ASSET e percebeu que poderia resolvê-lo. Yasha acompanhou o progresso da Equipe Mundial e percebeu que tinha uma visão especial na jogada 26. E todos os participantes do Projeto Polymath puderam acompanhar a conversa em rápida evolução e participar sempre que tivessem uma visão especial.
Em organizações offline convencionais, essas respostas flexíveis geralmente só são possíveis em grupos pequenos, se tanto. Em grupos maiores, diferentes membros se concentram em suas próprias áreas de responsabilidade pré-atribuídas. Ferramentas online mudam isso, possibilitando que grandes grupos explorem as áreas específicas de microespecialização de cada participante, na hora em que a necessidade dessa expertise surge. É isso que quero dizer com uma divisão dinâmica do trabalho. Idealmente, como vimos anteriormente, isso levará à serendipidade planejada. Mas mesmo quando isso não acontece, a divisão dinâmica do trabalho ainda é notavelmente diferente da divisão estática do trabalho convencional.

Nada disso nega o valor de uma divisão estática do trabalho. Alcançamos enormes melhorias em nossa capacidade de fabricar bens ao aprimorar a divisão estática do trabalho — pense na linha de montagem de Henry Ford ou mesmo na hipotética fábrica de alfinetes de Adam Smith. Mas, embora essa divisão do trabalho seja adequada à fabricação de bens, utilizando um processo previsível e repetitivo, ela tem sido menos útil na resolução de problemas criativos complexos. A razão é que, no trabalho criativo, muitas vezes são os insights e conexões não planejados e inesperados que mais importam.
Em muitos casos, o que torna uma ideia criativa importante é justamente o fato de ela combinar ideias que antes eram consideradas desconexas.
Quanto mais desconexas, mais importante a conexão — lembre-se da impressionante conexão que Stein e Grossmann fizeram entre a gravidade e a geometria riemanniana. Por isso, a maior obra criativa não pode ser planejada como parte de uma divisão de trabalho estática convencional. Ninguém poderia prever que Kasparov versus o Mundo se desenrolaria da maneira como se desenvolveu, e, portanto, não era possível prever que a microespecialização especial de Krush seria necessária para lidar com a situação que ocorreu no lance 10. E certamente não era possível prever a necessidade de Yasha no lance 26. Só foi possível realizar essa divisão de trabalho dinamicamente, conforme a situação surgisse.

A razão pela qual tudo isso importa é que, para problemas criativos complexos, até recentemente tínhamos que confiar na genialidade de indivíduos e pequenos grupos, além de interações ocasionais e fortuitas. Isso limita a gama de expertise que pode ser utilizada. Mesmo em uma tarefa como a produção de filmes

Com sua reputação de ser livre, as principais decisões criativas são tomadas principalmente por um pequeno número de pessoas. Vale ressaltar que as organizações modernas não estão completamente apegadas ao estilo estático de linha de montagem. Elas frequentemente alcançam uma divisão dinâmica do trabalho em pequena escala, com pequenos grupos trabalhando em equipes criativas. Isso acontece, por exemplo, em produções cinematográficas e também em muitas outras organizações criativas, incluindo organizações célebres como a Skunk Works, da Lockheed Martin, ou o Projeto Manhattan, que desenvolveu a bomba atômica. Técnicas de gestão como Gestão da Qualidade Total e manufatura enxuta incorporam ideias que ajudam a permitir uma divisão mais dinâmica do trabalho — um exemplo famoso é a maneira como a Toyota delega aos operários da fábrica grande responsabilidade por encontrar e corrigir defeitos de fabricação em tempo real.

A novidade sobre as ferramentas on-line é que elas tornam muito mais fácil fazer essa divisão dinâmica de trabalho em larga escala, trazendo a expertise de grupos muito maiores para lidar com problemas criativos difíceis.

A distinção entre divisão dinâmica e estática do trabalho também ilumina a diferença entre colaborações online e a colaboração científica convencional em larga escala. Considere, por exemplo, a colaboração de 138 físicos de partículas cujo trabalho levou à descoberta do bóson Z, uma nova partícula fundamental da natureza, em 1983, no acelerador de partículas CERN, na Europa. Ao contrário de Kasparov versus o Mundo ou do Projeto Polímata, cada um dos participantes da colaboração do CERN foi contratado para desempenhar um papel definido. Os papéis abrangiam muitas especialidades cuidadosamente escolhidas, desde engenheiros, cujo trabalho era resfriar o feixe de partículas, até estatísticos, cujo trabalho era dar sentido aos complexos resultados experimentais. Essas colaborações especializadas podem realizar feitos notáveis, mas, com seus papéis relativamente fixos e divisão estática do trabalho, deixam uma grande quantidade de microespecialização latente e demonstram pouca flexibilidade em seus propósitos. Sua inflexibilidade significa que, embora possam fazer ciência extremamente importante, não se trata de um modelo que possa ser facilmente adaptado aos fins mais fluidos característicos de grande parte do trabalho científico mais criativo.

Como a colaboração online vai além do Mercado

Uma das ferramentas mais poderosas da humanidade para amplificar a inteligência coletiva é o sistema de mercado, e podemos aprender muito sobre colaboração online comparando-a com o mercado. É claro que o mercado é tão familiar que é tentador considerá-lo garantido e focar apenas em exemplos em que ele amplifica a estupidez coletiva, como as crises de 2008 e 1929. Mas, na maioria das vezes, o mercado realmente amplifica nossa inteligência coletiva. Em seu livro " ***The Company of Strangers",*** o economista britânico Paul Seabright conta como, dois anos após a dissolução da União Soviética, se encontrou com um alto funcionário russo que estava visitando o Reino Unido para aprender sobre o livre mercado. "Por favor, entendam que estamos ansiosos para avançar em direção a um sistema de mercado", disse o funcionário russo, "mas precisamos entender os detalhes fundamentais de como tal sistema funciona. Diga-me, por exemplo: quem é responsável pelo fornecimento de pão para a população de Londres?"

A resposta familiar, mas ainda surpreendente, a essa pergunta é que, em uma economia de mercado, todos estão no comando. À medida que o preço de mercado do pão sobe e desce, ele influencia nosso comportamento coletivo: se devemos plantar um novo campo de trigo ou deixá-lo em pousio; se devemos abrir aquela padaria nova que você está pensando em abrir na esquina; ou simplesmente se devemos comprar dois ou três pães esta semana. Os preços são sinais para ajudar a coordenar as ações de fornecedores e consumidores: à medida que a demanda por um bem aumenta, o preço também aumenta, motivando novos fornecedores a entrar no mercado.

O resultado é uma dança maravilhosa de ações que coloca comida em nossas mesas, carros em nossas garagens e smartphones em nossos bolsos. A familiaridade nos faz tomar isso como certo, mas a dança é, na verdade, uma colaboração em massa milagrosa, mediada tão suavemente pelo mercado que só é notada quando ausente.

O que torna os preços úteis é que, como enfatizado pelo economista Friedrich von Hayek, eles agregam uma enorme quantidade de conhecimento oculto — conhecimento que, de outra forma, não seria aparente para todas as pessoas interessadas na produção ou no consumo de bens. Ao usar os preços para agregar esse conhecimento e orientar ações futuras, o mercado produz resultados superiores até mesmo aos indivíduos mais inteligentes e bem informados. Isso permite uma divisão dinâmica do trabalho: se uma enchente destruir a safra de trigo em grande parte dos Estados Unidos, o preço aumentará e outros fornecedores de trigo responderão trabalhando arduamente para aumentar a oferta.

Os mercados e o sistema de preços apresentam, portanto, muitas das propriedades que identificamos na colaboração online. Em contraste com a colaboração offline convencional

Organizações, elas utilizam tanto uma divisão dinâmica do trabalho quanto a serendipidade planejada. Mas colaborações online como o Projeto Polymath vão além dos mercados presenciais na complexidade dos problemas em consideração e na velocidade com que problemas imprevistos podem ser levantados e resolvidos. Mesmo que você não tenha interesse em matemática, é fácil apreciar o rico sabor desta "pergunta boba" proposta pelo participante do Polymath, Ryan O'Donnell:

> Alguém pode me ajudar com essa pergunta idiota?
> Suponha que ***A*** = ***B*** sejam a família de conjuntos que não inclui o último elemento n. Então ***A*** e ***B*** têm densidade de cerca de 1/2 dentro de ***K*** Nn,n/2ÿk/2. (Estamos pensando em ***k(n)*** ÿ ÿ, k(n)/n ÿ 0 aqui, certo?) [. . .]

Esse é apenas o começo da questão; está muito longe de "Qual é o preço do pão?". A pergunta de O'Donnell é muito específica e dependente do contexto para ser abordada por um mercado offline convencional. Ele poderia, talvez, ter publicado um anúncio em um periódico de matemática pedindo ajuda, mas o incômodo teria sido maior do que o benefício.

Em uma colaboração online como o Projeto Polymath, tal pergunta pode ocorrer a alguém, ser transmitida a outros participantes e respondida, tudo em minutos ou horas. As ferramentas online combinam, portanto, a divisão dinâmica do trabalho e a serendipidade projetada encontradas nos mercados com a flexibilidade e a espontaneidade das conversas cotidianas. Essa combinação as torna um grande avanço em relação aos mercados offline e, em particular, as torna adequadas para abordar problemas criativos complexos.

Até agora, concentrei-me nos mercados offline convencionais. É claro que, nos últimos anos, os mercados adotaram a internet e outras tecnologias de comunicação modernas e, com isso, mudaram e se tornaram mais complexos. Cada vez mais, eles também podem ser usados para abordar questões muito especializadas e dependentes do contexto. Nesse sentido, as ferramentas online estão gradualmente subsumindo e expandindo os mercados. Algo semelhante também está acontecendo nas organizações convencionais que discutimos na última seção: as ferramentas online são cada vez mais usadas como infraestrutura de comando e controle nessas organizações. E, assim, as ferramentas online podem subsumir e expandir tanto os mercados convencionais quanto as organizações convencionais. E, como veremos em breve, elas também podem subsumir e expandir a terceira forma histórica de colaboração: a conversa em pequenos grupos. Em cada caso, as ferramentas online estão possibilitando arquiteturas de atenção que vão além do que é possível em métodos offline de colaboração.

Como a colaboração online se compara à offline Conversa em pequenos grupos

Em muitos aspectos, colaborações online como o Projeto Polímata e Kasparov versus o Mundo assemelham-se a conversas presenciais em pequenos grupos. Como veremos, em alguns aspectos, a conversa presencial é realmente melhor do que a colaboração online, enquanto em outros, é nitidamente inferior. Mas antes de compararmos os dois, vamos primeiro esclarecer as coisas, descartando dois argumentos comuns, porém falaciosos, que pretendem relacionar a colaboração online à conversa presencial.

A primeira falácia é pensar que a colaboração online é de alguma forma semelhante ao trabalho monótono de um comitê. Às vezes, as pessoas ouvem falar de um projeto como o Projeto Polímata e sua mente salta para os estereótipos nada lisonjeiros que associamos a comitês — "Um camelo é um cavalo projetado por um comitê" e assim por diante. É verdade que muitos comitês sufocam a criatividade e o comprometimento. Mas isso não significa que a colaboração online tenha os mesmos problemas. Quando você analisa atentamente projetos como o Projeto Polímata e Kasparov versus o Mundo, eles não se parecem muito com comitês disfuncionais. Em vez disso, são comunidades vibrantes, repletas de criatividade e comprometimento.

Como essas colaborações escapam dos problemas de comitês disfuncionais? Entender por que alguns grupos funcionam bem e outros não é um problema complexo, e não vou abordar essa questão de forma abrangente aqui. Mas há dois fatores poderosos que ajudam a explicar por que a colaboração online frequentemente funciona bem onde um comitê não funcionaria.

Em primeiro lugar, os comitês são frequentemente compostos por pessoas que foram forçadas a participar deles, enquanto colaborações como o Projeto Polymath são repletas de voluntários entusiasmados. Esse comprometimento apaixonado faz uma grande diferença. Em segundo lugar, embora um comitê possa ser bastante prejudicado por alguns membros obstrutivos, as colaborações online podem frequentemente ignorar essas pessoas. No Projeto Polymath, por exemplo, era fácil para participantes bem informados ignorarem as ocasionais contribuições bem-intencionadas, mas inúteis. Colaborar online simplesmente não é a mesma coisa que trabalhar em comitê.

Uma segunda falácia, por vezes apresentada pelos céticos da colaboração online, é que é sempre possível substituir colaborações online por colaborações offline equivalentes. Por exemplo, eles podem argumentar que, com paciência suficiente e uma sala cheia de matemáticos, seria possível...

Uma "simulação" offline do Projeto Polímata. Há dois problemas com esse argumento. O primeiro é que, na prática, é muito mais fácil reunir-se online do que offline. Portanto, a objeção é um pouco como dizer que a invenção do automóvel ou do trem de passageiros não mudou nada nas viagens, porque as pessoas sempre foram capazes, em princípio, de usar uma charrete para percorrer longas distâncias. A observação é verdadeira, mas tem pouca importância prática para o comportamento real das pessoas. O segundo problema é que o comportamento humano em uma sala cheia de matemáticos seria, na prática, dramaticamente diferente do que no Projeto Polímata. Para escolher um dos muitos exemplos de diferenças: offline, se alguém fala com você quando você está cansado e irritado, você pode não entender o que a pessoa disse; online, você pode ler e reler quando quiser, quando estiver alerta e entusiasmado. Por causa dessas e de muitas outras diferenças, você só pode fazer a simulação offline se fizer suposições irrealistas sobre como os humanos se comportariam na sala. Isso não quer dizer que uma sala cheia de matemáticos não pudesse colaborar para realizar um trabalho notável. Mas não usaria um processo no estilo Polymath, mas sim uma arquitetura de atenção diferente. Ferramentas online realmente nos permitem colaborar de novas maneiras.

Com essas duas falácias esclarecidas, o que dizer das maneiras pelas quais a conversa offline é genuinamente superior à colaboração online? Uma se destaca em especial: a rica natureza do contato presencial. Linguagem corporal, expressão facial, tom de voz e contato informal regular são tremendamente importantes para uma colaboração eficaz e não podem ser substituídos. Com pessoas de quem você gosta, a conversa presencial é agradável e estimulante, e a colaboração online perde algo em contraste. É claro que essa perda está sendo gradualmente compensada por tecnologias colaborativas mais expressivas — uma ferramenta como o bate-papo por vídeo do Skype é notavelmente eficaz como forma de colaboração. A longo prazo, ideias como mundos virtuais e realidade aumentada podem até tornar o contato online melhor do que o contato presencial. Ainda assim, hoje em dia, a experiência online de colaboração direta entre pessoas carece muito da riqueza da colaboração offline. É tentador concluir que a colaboração online não pode ser tão boa quanto a offline.

O problema com essa conclusão é que ela ignora a questão de como encontrar a pessoa certa para trabalhar em primeiro lugar. Isso talvez se deva ao fato de que encontrar a pessoa certa historicamente tem sido um problema tão difícil que geralmente não nos damos ao trabalho. Offline, pode levar meses para encontrar um novo colaborador com experiência que complemente a sua da maneira certa. Mas isso muda quando você pode fazer uma pergunta em um

fórum online e receba uma resposta dez minutos depois de um dos maiores especialistas do mundo no tópico sobre o qual você perguntou. Na resolução criativa de problemas, muitas vezes é melhor ter uma interação concisa de vinte minutos, apenas por texto, com um especialista que pode resolver seu problema com facilidade, em vez de semanas de uma agradável discussão presencial com alguém cujo conhecimento não é muito diferente do seu. E, em qualquer caso, você não precisa fazer essa escolha. Na prática, você pode usar ferramentas relativamente impessoais para encontrar a pessoa ou pessoas certas para o problema em questão, e ferramentas mais expressivas, como bate-papo por vídeo, mundos virtuais e realidade aumentada, para tornar o trabalho com essa pessoa ou pessoas o mais eficaz possível.

Em outras palavras, as grandes vantagens da colaboração online em relação à conversa offline estão na escala e na diversidade cognitiva. Imagine que o pessoal da ASSET Índia reunisse um grupo para fazer um brainstorming de ideias para roteadores sem fio. A menos que tivessem muita sorte, o grupo não teria ninguém com o mesmo tipo de expertise que Zacary Brown. Ao aumentar a escala da colaboração, as ferramentas online expandem o leque de expertise disponível, reduzindo a chance de o grupo ser bloqueado por um problema que ninguém no grupo consegue resolver. Idealmente, a serendipidade planejada e a massa crítica conversacional ocorrerão, permitindo que o grupo explore em profundidade uma gama muito mais ampla de ideias do que seria possível em um grupo pequeno, com sua expertise limitada.

Como as ferramentas online permitem que a conversa seja ampliada? A resposta óbvia é que as ferramentas online facilitam a reunião de especialistas em todo o mundo como parte de um grupo. Isso é importante, mas é apenas uma pequena parte do que está acontecendo. De fato, ao usar uma arquitetura de atenção cuidadosamente projetada, as ferramentas online permitem que as colaborações envolvam muito mais pessoas do que é prático em conversas offline. Deixe-me descrever como isso funcionou no Projeto Polymath. Superficialmente, o formato do Projeto Polymath, baseado em comentários em blogs, parece semelhante a discussões sobre matemática em conversas presenciais. Mas vai além em três aspectos importantes. Primeiro, ao trabalhar online, as pessoas pré-filtram seus comentários mais do que em conversas matemáticas comuns.

Em conversas offline, até os melhores matemáticos fazem longas pausas, precisam voltar atrás e, ocasionalmente, ficam confusos. No Projeto Polímata, a maioria dos comentários destilou um ponto de forma relativamente concisa.

Em segundo lugar, como leitor, é fácil pular rapidamente os comentários do blog. Quando você está cara a cara, se não entender o que alguém está dizendo, pode acabar ouvindo essa pessoa falar de forma incompreensível por dez minutos.

Mas em um blog você pode dar uma olhada em um comentário por alguns segundos e tomar nota

da ideia geral e seguir em frente. Terceiro, quando você pula um comentário, sempre sabe que pode retornar a ele mais tarde. Ele é arquivado e facilmente encontrado em mecanismos de busca. O efeito geral dessas três diferenças é aumentar o número de pessoas que podem participar da conversa. Ao aumentar a escala da conversa, o blog nos dá acesso às melhores ideias de um conjunto cognitivamente mais diverso de participantes, e assim a serendipidade planejada e a massa crítica conversacional têm maior probabilidade de ocorrer.

No entanto, há uma compensação inerente ao aumento da colaboração. Por um lado, uma colaboração deve envolver o maior e mais diverso grupo possível de participantes, em termos cognitivos. Por outro lado, uma vez que a colaboração se torne grande o suficiente, os participantes não conseguem prestar atenção a tudo o que está acontecendo. Em vez disso, eles ***precisam ,*** forçosamente , começar a prestar atenção apenas a algumas das contribuições. Idealmente, a arquitetura da atenção direcionará os participantes para os lugares onde seus talentos específicos são mais adequados para dar o próximo passo — onde eles têm a máxima vantagem comparativa. Assim, cada participante vê apenas parte da colaboração maior. Como um exemplo simples, a InnoCentive classifica os Desafios em áreas temáticas, para ajudar os participantes a encontrar os Desafios de maior interesse para eles. No próximo capítulo, veremos algumas maneiras mais sofisticadas de ajudar as pessoas a decidir para onde direcionar sua atenção. Dessa forma, é possível escalar além do ponto em que cada participante deve prestar atenção a toda a colaboração. Em outras palavras, a arte de escalar é filtrar as contribuições para que cada participante veja apenas as contribuições que pessoalmente considere mais valiosas e estimulantes; o importante não é o que vemos, mas o que podemos ignorar. Quanto melhores os filtros, melhor a nossa atenção é adaptada às oportunidades de contribuição. Em resumo, uma arquitetura ideal de atenção permite que o maior e mais diverso grupo cognitivamente utilize da melhor forma a limitada atenção disponível, de modo que, a qualquer momento, cada participante esteja maximizando sua vantagem comparativa. Colaborações como o Projeto Polymath contribuem apenas parcialmente para esse objetivo. Ao utilizar uma arquitetura de atenção melhor, é possível escalar a colaboração ainda mais além do Projeto Polymath. No próximo capítulo, examinamos diversos padrões que podem ser usados para escalar as colaboraç

CAPÍTULO 4

Padrões de colaboração online

Em 26 de agosto de 1991, às 2h12, um estudante finlandês de programação de 21 anos chamado Linus Torvalds publicou uma breve nota em um fórum online para programadores. A mensagem dizia, em parte:

> Estou desenvolvendo um sistema operacional (gratuito) (apenas um hobby, não será grande e profissional como o GNU) para clones de AT 386(486)... Gostaria de saber quais recursos a maioria das pessoas gostaria. Sugestões são bem-vindas, mas não prometo que as implementarei :-)

Apenas 14 minutos depois, outro usuário respondeu com as palavras "Conte-nos mais!" e fez várias perguntas. Quase seis semanas depois, em 5 de outubro, Torvalds publicou uma segunda nota, anunciando que o código do seu sistema operacional — que em breve seria chamado de Linux — estava agora disponível publicamente. Ele escreveu no anúncio:

> Este é um programa para hackers, criado por um hacker. Eu me diverti fazendo isso, e alguém pode gostar de dar uma olhada e até mesmo modificá-lo para atender às suas próprias necessidades. Ele ainda é pequeno o suficiente para ser compreendido, usado e modificado, e aguardo ansiosamente seus comentários.

Torvalds era um desconhecido, um estudante que trabalhava em relativo isolamento na Universidade de Helsinque, e não fazia parte de alguma startup descolada do Vale do Silício. Mesmo assim, o que ele havia anunciado era interessante para muitos hackers. O sistema operacional é o centro nervoso de um computador, a peça que faz todo o resto funcionar. Entregar o código de um sistema operacional a um hacker hardcore é como dar a um artista as chaves da Capela Sistina e pedir que ele redecore tudo. Logo após a publicação de Torvalds, uma lista de discussão de ativistas do Linux foi criada e, apenas três meses depois, a lista havia crescido para 196 membros.

Torvalds não só disponibilizou gratuitamente o código do seu sistema operacional, como também encorajou outros programadores a enviar-lhe por e-mail o código para

possível incorporação ao Linux. Ao fazer isso, Torvalds iniciou a formação e o rápido crescimento de uma comunidade de desenvolvedores Linux — programadores que coletivamente o ajudaram a melhorar o Linux. Em março de 1994, 80 pessoas foram nomeadas como contribuidoras no arquivo Linux Credits, e as pessoas estavam contribuindo com código a uma taxa astronômica. Em 1995, a empresa Red Hat foi formada, comercializando uma das primeiras versões comercialmente bem-sucedidas do Linux; em 1999, a Red Hat abriu o capital na Bolsa de Valores de Nova York, com um valor de mercado de 3 bilhões de dólares ao final de seu primeiro dia de negociação. No início de 2008, o kernel Linux — a parte central do sistema operacional Linux — continha quase 9 milhões de linhas de código, escritas em colaboração com mais de 1.000 pessoas. É um dos artefatos de engenharia mais complexos já construídos.

O Linux se tornou tão difundido que é fácil considerá-lo garantido. Embora o Microsoft Windows continue sendo o sistema operacional dominante para uso doméstico e de escritório, em muitas outras áreas o Linux o supera. Empresas como Google, Yahoo! e Amazon usam enormes clusters Linux, contendo dezenas ou centenas de milhares de computadores.
Nas empresas de animação e efeitos visuais de Hollywood, o Linux é o sistema operacional dominante, superando o Windows e o MacOS e desempenhando um papel importante na Pixar, Dreamworks e Industrial Light and Magic. Na indústria de eletrônicos de consumo, empresas como Sony, Nokia e Motorola usam Linux em tudo, de celulares a televisores. Essa ubiquidade faz com que seja fácil esquecer o quão notável é a história do Linux. Imagine que, em 1991, um estudante de programação finlandês de 21 anos se aproximasse de você, dizendo que havia escrito o núcleo de um sistema operacional e planejava lançar o código e, a propósito, esperava recrutar um exército de programadores voluntários para melhorá-lo. Você acharia isso ridículo. Era ridículo. Tão ridículo que nem o próprio Torvalds imaginou que aconteceria.

O Linux é um exemplo de software de código aberto. Projetos de software de código aberto têm dois atributos principais. Primeiro, o código é disponibilizado publicamente, para que qualquer pessoa possa experimentá-lo e modificá-lo, não apenas o programador original. Segundo, outras pessoas são incentivadas a contribuir com melhorias no código. Isso pode significar enviar um relatório de bug quando algo dá errado, ou talvez sugerir uma alteração em uma única linha de código, ou até mesmo escrever um módulo de código principal contendo milhares de linhas de código. Os projetos de código aberto mais bem-sucedidos recrutam um grande número de colaboradores, que juntos podem desenvolver software muito mais complexo do que qualquer programador individual conseguiria desenvolver.

seus próprios. Para se ter uma ideia da escala, em 2007 e 2008, os desenvolvedores Linux adicionaram uma média de 4.300 linhas de código por ***dia*** ao kernel Linux, excluíram 1.800 linhas e modificaram 1.500 linhas. Essa é uma taxa de mudança impressionante — em um grande projeto de software, um desenvolvedor experiente normalmente escreve alguns milhares de linhas de código por ***ano.***

É claro que a maioria dos projetos de código aberto tem menos colaboradores do que o Linux. Um repositório popular de projetos de código aberto chamado SourceForge abriga mais de 230.000 projetos de código aberto. Quase todos esses projetos têm apenas um ou alguns colaboradores. Mas um pequeno número de projetos cativou a imaginação dos programadores, atraindo dezenas, centenas ou milhares de colaboradores.

O código aberto começou no mundo da programação, mas não se trata fundamentalmente de programação. Em vez disso, o código aberto é uma metodologia geral de design que pode ser aplicada a qualquer projeto que envolva informações digitais. Se você é arquiteto, por exemplo, pode fazer arquitetura de código aberto: basta compartilhar os projetos de seus edifícios livremente e incentivar outras pessoas a contribuírem com melhorias. Em 2006, um arquiteto chamado Cameron Sinclair e uma jornalista chamada Kate Stohr lançaram a Open Architecture Network, que está criando uma comunidade online para arquitetura de código aberto — uma espécie de Source-Forge para arquitetura. No início de 2010, o site continha mais de 4.000 projetos, muitos com plantas baixas, discussões sobre materiais de construção, fotografias de edifícios concluídos e assim por diante, todos disponíveis para reutilização e aprimoramento por outros.
O site concentra-se especialmente em projetos para uso em países em desenvolvimento, e Sinclair e Stohr esperam que ele ajude a disseminar as melhores ideias e inovações arquitetônicas mais rapidamente. Um exemplo é mostrado na figura 4.1, o projeto para uma escola primária construída em Gando, uma cidade de 3.000 habitantes no pequeno país de Burkina Faso (anteriormente conhecido como Alto Volta), na África Ocidental. O projeto inclui plantas baixas, elevações e muitos outros detalhes, além de fotos da escola concluída.

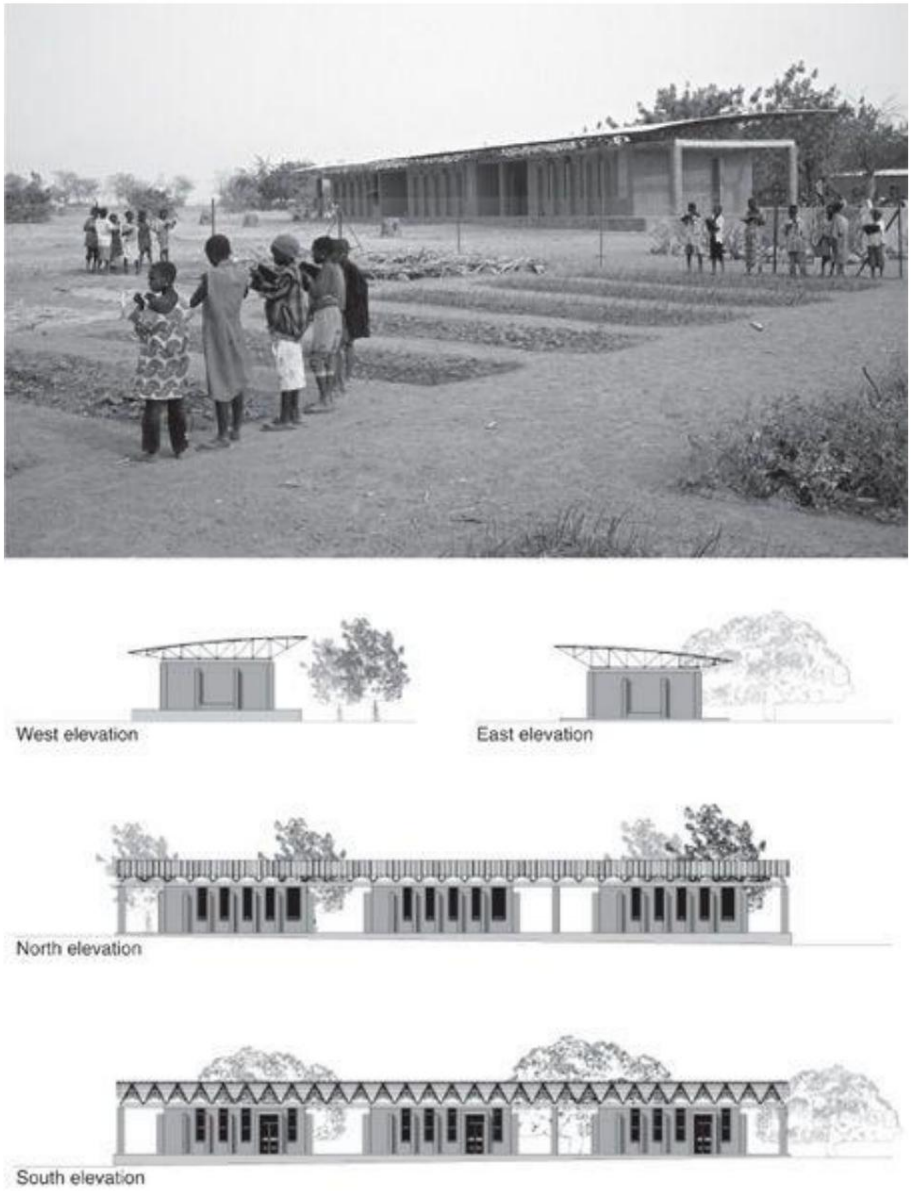


Figura 4.1. Acima: uma escola primária na cidade de Gando, no Burkina Faso, na África Ocidental. Abaixo: um dos vários documentos de projeto para a escola, disponíveis gratuitamente para download na Open Architecture Network. Outras pessoas podem usar os documentos de projeto e modificá-los conforme suas próprias necessidades. Crédito: Siméon Duchaud / Prêmio Aga Khan de Arquitetura.

Não é só a arquitetura que pode ser de código aberto. Se você é um artista digital, pode fazer arte de código aberto: compartilhe os arquivos da sua arte digital livremente e incentive outros a contribuírem com melhorias. Se você é biólogo, pode fazer biologia de código aberto: compartilhe projetos de DNA de seres vivos e incentive outros a contribuírem com melhorias. Existe uma comunidade de biólogos fazendo exatamente isso. Se você está escrevendo um...

enciclopédia, você pode compartilhar o texto dos seus artigos enciclopédicos livremente e incentivar outras pessoas a contribuírem com melhorias. É assim que a Wikipédia é escrita: a Wikipédia é um projeto de código aberto. O padrão subjacente em todos esses projetos é o mesmo: compartilhe seu design digital e incentive outras pessoas a contribuírem com mudanças. O Projeto Polymath não segue exatamente esse padrão, mas utiliza ideias semelhantes, criando um espaço online onde as pessoas podem compartilhar suas ideias e trabalhar para aprimorar as ideias de outras pessoas.

Até agora neste livro, analisamos vários exemplos que mostram como ferramentas online podem tornar grupos mais inteligentes. Colaborações de código aberto geralmente têm propósitos diferentes: visam dar às pessoas a liberdade de aprimorar e modificar o trabalho de outras pessoas e — para grandes projetos, como Linux e Wikipédia — permitir que grupos criem projetos mais complexos do que qualquer indivíduo conseguiria criar sozinho. Essa diferença de propósito se reflete no fato de que, embora a Wikipédia seja impressionante, em muitos assuntos os maiores especialistas do mundo poderiam escrever artigos melhores.
Da mesma forma, o código para Linux precisa apenas ser bom o suficiente para funcionar; não precisa ser da mais alta qualidade em todos os aspectos. Mas, apesar dessa diferença de intenção em relação aos nossos exemplos anteriores, o código aberto ainda pode nos ensinar muito sobre como ampliar a inteligência coletiva. Em particular, as colaborações de código aberto têm sido extremamente eficazes na expansão, aumentando assim a diversidade cognitiva e a gama de microespecialização disponíveis para a colaboração. Neste capítulo, identificaremos quatro padrões poderosos que as colaborações de código aberto têm usado para escalar. (1) um compromisso implacável com o trabalho modular, encontrando maneiras inteligentes de dividir a tarefa geral em subtarefas menores; (2) incentivar pequenas contribuições para reduzir as barreiras de entrada; (3) permitir a fácil reutilização de trabalhos anteriores por outras pessoas; e (4) usar mecanismos de sinalização, como pontuações, para ajudar as pessoas a decidir para onde direcionar sua atenção. Esses padrões podem ser incorporados a qualquer arquitetura de atenção e, portanto, usados para ampliar a inteligência coletiva.

A importância de ser modular

Para entender como as colaborações de código aberto escalam, vamos olhar para uma época em que a colaboração Linux quase falhou em escalar, uma época em que

A comunidade de desenvolvedores Linux quase se dividiu em dois grupos distintos, trabalhando em duas versões distintas do Linux. O incidente começou inofensivamente, em 29 de setembro de 1998, com uma postagem na lista de discussão do kernel Linux do desenvolvedor Michael Harnois. Harnois escreveu dizendo que estava tendo problemas com parte do sistema de exibição do Linux. Esse tipo de reclamação não era incomum — aliás, tais reclamações são a matéria-prima que os desenvolvedores Linux usam para melhorar o código — e um respeitado desenvolvedor Linux chamado Geert Uytterhoeven respondeu rapidamente a Harnois.
Uytterhoeven disse a ele para não perder tempo, que o problema já havia sido corrigido e que a única razão pela qual Harnois estava tendo problemas era porque o código que corrigia o problema ainda não estava incluído na base de código oficial do Linux, mantida por Linus Torvalds.

Até então, tudo corria como sempre. Mas o que Uytterhoeven acrescentou em seguida provocou uma grande explosão. Ele disse a Harnois que, embora a correção para o seu problema ainda não estivesse na base de código oficial, ele poderia obter uma cópia da correção em um site chamado VGER. O VGER era um serviço que começou como um espelho (ou seja, uma cópia) do código oficial do Linux, um local alternativo onde as pessoas podiam baixar o Linux, caso o site principal estivesse fora do ar ou fosse difícil de acessar. Mas alguns desenvolvedores Linux estavam ficando insatisfeitos com Torvalds, acreditando que ele não estava integrando suas contribuições com rapidez suficiente à base de código oficial do Linux. O grupo de voluntários que administrava o VGER, por outro lado, estava aceitando algumas dessas contribuições, e sabia-se discretamente que o "VGER Linux" estava começando a ficar à frente do Linux oficial em aspectos cruciais.

Menos de duas horas após a publicação de Uytterhoeven, Linus Torvalds respondeu à lista de discussão com uma mensagem concisa, afirmando que Harnois "não estava perdendo tempo" e que o VGER era irrelevante para o desenvolvimento do Linux. A publicação de Torvalds desencadeou uma avalanche de respostas, com alguns dos mais respeitados colaboradores do Linux reclamando veementemente que esta não era a primeira vez que ele deixava de integrar uma contribuição importante ao código oficial do Linux. Vários reclamaram que haviam enviado contribuições de código a Torvalds diversas vezes sem receber nenhum reconhecimento, às vezes até mesmo por trabalhos realizados a seu pedido. Torvalds, por sua vez, também expressou frustração:

> Francamente, esta discussão em particular (e outras anteriores) só me deixou irritado e ESTÁ ACRESCENDO pressão. Em vez disso, sugiro que, se você tiver alguma reclamação sobre como eu lido com [contribuições], pense no que eu terei que lidar por cinco minutos.

> Vão embora, pessoal. . . . Não estou interessado, estou de férias e não quero mais saber disso. Resumindo, saia da minha caixa de correio.

Para ter sucesso, uma colaboração deve dividir o problema que está atacando em tarefas que podem ser realizadas por indivíduos individualmente. Na época dessa explosão, a comunidade Linux havia crescido tanto que a tarefa de revisar e integrar os códigos enviados estava além de Torvalds (ou provavelmente de qualquer pessoa individualmente). Acrescente as palavras de um dos desenvolvedores Linux envolvidos no imbróglio, Larry McVoy: "O Linus não escala". Como resultado, a comunidade de desenvolvimento Linux não estava mais trabalhando de forma eficaz e corria o risco de se fragmentar em duas ou mais comunidades separadas. Isso não ocorreu porque Torvalds ou qualquer outra pessoa estivesse fazendo algo errado. Em vez disso, foi uma consequência do sucesso: a comunidade havia crescido tanto que a antiga maneira de fazer as coisas não funcionava mais.

A maneira óbvia de resolver o problema era dividir a tarefa de aprovação das contribuições de código entre várias pessoas. Mas alguns desenvolvedores Linux temiam que o amplo conhecimento de Torvalds sobre o kernel Linux pudesse ser essencial para revisar e aprovar as contribuições de código.
Permitir que outros aprovem contribuições pode realmente prejudicar o Linux? Talvez alguma funcionalidade essencial, mas até então tácita, da colaboração Linux possa ter se perdido. Felizmente, esses temores não se confirmaram.
Após uma acalorada discussão online e uma reunião presencial de alguns dos principais desenvolvedores Linux, incluindo Torvalds e o criador do VGER, Dave Miller, Torvalds concordou em delegar mais tomada de decisões aos tenentes, e isso ocorreu sem quaisquer efeitos nocivos evidentes.

Em algumas colaborações, é fácil dividir o problema em tarefas menores. Lembre-se do projeto de classificação de galáxias Galaxy Zoo, que conhecemos no capítulo de abertura. O Galaxy Zoo pede aos colaboradores que respondam a perguntas sobre apenas uma galáxia por vez, dividindo o problema de classificação de galáxias em milhões de pequenas tarefas. Essa é uma maneira simples, mas eficaz, de dividir o problema geral do Galaxy Zoo.

Às vezes, porém, esse tipo de modularidade é muito mais difícil de alcançar. No Projeto Polymath, o trabalho era realizado por meio de comentários em postagens de blog. Nos primeiros dias do projeto, era fácil para matemáticos interessados participarem da discussão. Mas o número de comentários aumentou rapidamente, chegando a 800 comentários e

170.000 palavras. Para quem estava de fora, essa era uma barreira assustadora, já que os comentários não estavam organizados de forma a permitir que participassem da discussão sem antes compreender a maior parte das contribuições anteriores. Embora o Projeto Polymath tenha sido uma colaboração grande para os padrões convencionais da matemática, com contribuições de 27 pessoas, provavelmente teria sido ainda maior se a discussão tivesse sido menos monolítica e mais modular. Isso, por sua vez, teria aumentado a diversidade cognitiva, disponibilizando uma gama maior de expertise para a colaboração.

Esse estilo narrativo monolítico é uma característica inevitável de colaborações como o Projeto Polymath? Ou é possível conceber uma abordagem mais modular que divida a colaboração em subprojetos? Podemos compreender essas questões analisando mais de perto grandes projetos de código aberto, como o Linux. Esses projetos não alcançaram a modularidade facilmente ou por acaso, mas sim com muito, muito esforço. Eles assumiram um compromisso consciente de serem modulares e, em seguida, cumpriram esse compromisso incansavelmente, mesmo quando isso exigiu muito trabalho. Vimos um exemplo disso na forma como a comunidade Linux respondeu à crise do VGER. Mas ainda mais impressionante, embora de forma mais discreta, é o compromisso diário da comunidade de desenvolvedores Linux com a modularidade. Por exemplo, a base de código original do kernel Linux não tinha o tipo de estrutura modular simples que facilitaria o envolvimento de potenciais desenvolvedores na melhoria do código. Para a versão 2.0 do Linux, todo o código Linux foi substancialmente reescrito e reorganizado para torná-lo modular. Isso pode parecer fácil no papel, mas exigiu um enorme esforço coordenado dos desenvolvedores Linux. Veja como Torvalds explicou:

> Com o kernel Linux, ficou claro muito rapidamente que queríamos um sistema o mais modular possível. O modelo de desenvolvimento de código aberto realmente exige isso, porque, caso contrário, não seria fácil ter pessoas trabalhando em paralelo.
>
> . . .
>
> Com o kernel 2.0, o Linux realmente evoluiu bastante. Foi nesse ponto que adicionamos módulos carregáveis do kernel. Isso obviamente melhorou a modularidade, criando uma estrutura explícita para a escrita de módulos. Os programadores podiam trabalhar em diferentes módulos sem risco de interferência. Eu podia manter o controle sobre o que estava escrito no kernel. Então, mais uma vez, gerenciar pessoas e gerenciar código levou à mesma decisão de design. Para manter o número de pessoas

> Trabalhando em conjunto no Linux, precisávamos de algo como módulos do kernel. Mas, do ponto de vista do design, também era a coisa certa a fazer.

Esse padrão de modularidade consciente e implacável é observado na maioria das grandes colaborações de código aberto. Frequentemente, é necessário até mesmo em projetos onde a modularidade parece fácil de alcançar, como a Wikipédia. Superficialmente, a Wikipédia parece ser apenas uma coleção de artigos de enciclopédia, com uma estrutura modular simples e natural: a escrita é naturalmente dividida entre os diferentes artigos. Mas essa modularidade superficial é apenas parte da história. Escrever uma enciclopédia envolve muitas tarefas além da edição dos artigos, e essa complexidade adicional se reflete na estrutura da Wikipédia. Talvez o exemplo mais simples seja que cada artigo da Wikipédia tenha uma página de "Discussão" associada. Se você não sabe o que é uma página de Discussão da Wikipédia, abra seu navegador e veja o artigo "Geologia" da Wikipédia (http://en.wikipedia.org/wiki/Geology). No topo da página, você verá uma aba chamada "Discussão". Clique na aba e você será levado para a página de Discussão do artigo "Geologia". É lá que a discussão ***sobre*** o artigo acontece entre os editores da Wikipédia: discussão sobre deficiências no artigo, discussão sobre como o artigo pode ser melhorado e até mesmo discussão sobre se o artigo deveria existir em primeiro lugar. Essas páginas de Discussão são um local para conversas sobre muitas tarefas que são essenciais para o funcionamento adequado da Wikipédia, mas que não podem ser realizadas nas páginas de artigos. Além das páginas de Discussão, a Wikipédia também tem uma vasta gama de outras páginas especiais, cada uma voltada para tarefas específicas. A página "Village Pump", por exemplo, é para discussão sobre a política da Wikipédia, questões técnicas e assim por diante. Há uma página listando artigos que estão sendo considerados para exclusão da Wikipédia. Muitas páginas da Wikipédia tratam de tópicos de interesse apenas para a própria comunidade da Wikipédia. Algumas dessas páginas são engraçadas: há um teste de 1.181 perguntas para ver se você é um Wikipedista viciado (para quem se dispuser a fazer o teste inteiro, acho que a resposta é obviamente "sim"); uma lista de artigos com títulos bizarros ("22,86 Centimetre Nails", a versão métrica da banda "Nine Inch Nails", agora infelizmente excluída); e muitos outros. Algumas das páginas são tristes: há uma página listando wikipedistas falecidos, com links para suas páginas de usuário, onde você frequentemente encontrará comunidades de luto de amigos e familiares. A Wikipédia não é uma enciclopédia. É uma cidade virtual, uma cidade cujo principal produto de exportação para o mundo são seus artigos de enciclopédia, mas com uma vida interna própria. Todas essas páginas — as páginas de discussão, os eciges, as páginas da comunidade e as

Os próprios artigos — refletem tarefas vitais da Wikipédia e ajudam a dividir o enorme problema de administrar uma enciclopédia em muitas tarefas menores. E, como em uma cidade bem administrada, essa divisão não foi determinada antecipadamente por algum comitê central, mas surgiu organicamente, em resposta às necessidades e desejos dos "residentes" da Wikipédia; os editores que escrevem a Wikipédia.

Quando esse padrão de modularidade consciente e implacável não é utilizado, a colaboração em código aberto não se expande. Houve, por exemplo, muitas tentativas fracassadas de usar wikis e uma abordagem de código aberto para escrever um romance de boa qualidade. Uma tentativa de destaque foi o projeto Million Penguins, executado pela editora Penguin em fevereiro e março de 2007. A ideia era recrutar escritores para produzir um romance colaborativo usando software wiki. A julgar pelo número de pessoas que contribuíram (1.500), o projeto foi um sucesso. Mas essas pessoas nunca conseguiram trabalhar juntas de forma eficaz e, como obra literária, o resultado foi um fracasso. No início do projeto, um dos coordenadores, Jon Elek, escreveu: "Ficarei feliz, desde que consiga evitar se tornar uma espécie de assassinos zumbis robóticos contra ninjas africanos no espaço narrados por uma tiara papal". O romance em si era muito mais estranho. Aqui está uma pequena amostra para você ter uma ideia:

> Não havia possibilidade de dar um passeio naquele dia... nadar, talvez, mas não caminhar — pois Artie era uma baleia, uma baleia jubarte, para ser mais preciso, pelo menos naqueles momentos. Era um dia ensolarado, e Artie teria usado seus óculos de sol, mas ser uma baleia significava que ele não tinha orelhas, o que dificultava a fixação dos óculos. Não importava, pensou, pelo menos ele era jovem e forte.

É fácil entender por que a Penguin realizou esse experimento. Wikis têm sido usados com sucesso para produzir não apenas uma enciclopédia, mas também muitas outras obras de referência, desde a fabulosa Muppet Wiki (muppet.wikia.com) até a Intellipedia da Comunidade de Inteligência dos EUA (sem URL publicamente acessível para essa, desculpe!). Superficialmente, um romance se parece bastante com uma enciclopédia ou outra obra de referência. Mas o grau de modularidade suficiente para produzir uma enciclopédia não é suficiente para escrever um romance de primeira linha, porque deixa algumas tarefas essenciais sem serem realizadas. Cada frase em um romance tem uma relação potencial com todas as outras frases, uma relação potencial com cada arco narrativo dentro do romance e uma relação com o arco narrativo geral. Um bom autor está ciente de todas essas relações e as utiliza para alcançar ressonância e

reforço entre diferentes partes da história e evitar dissonância e incoerência. Para escrever um bom romance, uma das tarefas sempre à sua frente é comparar a frase que você está escrevendo ***agora*** com todas essas outras partes do romance, pensando se ela aprimora ou prejudica o todo do romance. Para que a escrita colaborativa tenha sucesso, é preciso manter o controle de todas essas relações possíveis. No entanto, os wikis não oferecem nenhuma maneira natural de resolver o problema de manter o controle dessas relações. Portanto, embora os wikis possam funcionar bem para artigos curtos e independentes, como os que aparecem em uma obra de referência, eles não funcionam como um meio colaborativo para textos mais longos. Ainda assim, a tecnologia colaborativa está em seus primórdios. Minha aposta é que, em breve, uma tecnologia para colaboração online será desenvolvida, provavelmente não muito diferente de um wiki, mas facilitando o controle das relações entre as diferentes partes de um romance. Isso será um grande passo em direção ao primeiro bom romance escrito por uma colaboração de código aberto. (É claro que gerenciar essas relações é apenas parte do desafio; no próximo capítulo, encontraremos mais dificuldades.)

Vimos como a Wikipédia e wikis de referência semelhantes usam uma estrutura de página cuidadosamente escolhida para modularizar. Outra abordagem à modularidade é ilustrada pela forma como o trabalho no navegador Firefox é organizado. Se você não conhece o Firefox, ele é uma alternativa popular ao navegador Internet Explorer. Assim como o Linux, o Firefox é um projeto de código aberto. Mas os desenvolvedores do Firefox organizam seu trabalho usando uma abordagem diferente daquela usada tanto no Linux quanto na Wikipédia. Em particular, eles organizam grande parte do seu trabalho usando uma ferramenta conhecida como ***rastreador de problemas.*** Para entender como o rastreador de problemas funciona, imagine que você é um usuário do Firefox que encontrou um bug. Por exemplo, um bug que notei às vezes é este: na minha lista de favoritos do Firefox, as pequenas imagens (chamadas favicons) ao lado dos meus favoritos às vezes se misturam. Ou seja, a imagem errada aparece ao lado de um favorito, ou favicons aparentemente aleatórios de outros sites aparecem sem motivo aparente. Não tenho ideia de por que isso acontece, e é apenas um pequeno incômodo em um produto excelente, mas pode ser um pouco confuso. De qualquer forma, tendo notado esse bug, você decide ajudar o projeto Firefox reportando-o. Para isso, você acessa o rastreador de problemas online do Firefox, um site onde você pode inserir uma descrição do problema e quaisquer outros detalhes que possam ser úteis para quem estiver tentando corrigir o bug: qual página da web você estava navegando quando notou o bug, qual sistema operacional você usa, qual versão do Firefox e assim por diante.

Pedi que você imaginasse fazer isso, mas, na verdade, você não precisa imaginar. Verifiquei o rastreador de problemas do Firefox e alguém chamado Bob fez exatamente o que acabei de descrever em 11 de janeiro de 2008.
Assim que ele enviou seu relatório sobre o bug do favicon, ele rapidamente foi incluído na lista de "Hot Bugs" do rastreador de problemas. A lista de Hot Bugs é a Firefox Central Station, com muitos dos desenvolvedores que trabalham no Firefox acompanhando a lista de perto. Quando veem um bug que acham que podem ajudar a corrigir, eles se mobilizam. Para o bug do favicon de Bob, um tópico de discussão rapidamente se iniciou. Lendo a discussão, você descobre que o bug é surpreendentemente sutil e, na verdade, envolve mais de um problema no código do Firefox. Dezenas de pessoas acabaram se envolvendo antes que o bug fosse corrigido de forma conclusiva.

O rastreador de problemas não serve apenas para corrigir bugs, ele também é usado para propor e implementar novos recursos. Se você quiser sugerir um novo recurso no Firefox, pode acessar o rastreador de problemas, sugerir o recurso e uma conversa será iniciada. Se um número suficiente de pessoas desejar o recurso, alguém começará a codificá-lo. O rastreador de problemas, portanto, funciona como uma miscelânea de problemas e ideias, cada um com seus próprios tópicos de conversa anexados. É uma ótima maneira de modularizar o trabalho: ao organizar a atenção dos participantes em torno de questões individuais, o rastreador de problemas limita o escopo da conversa e, portanto, a quantidade de atenção que as pessoas precisam investir para participar. Em vez de ter que entender toda a discussão anterior, como no Projeto Polymath, os participantes precisam apenas entender a questão em questão. Isso permite que muito mais pessoas se envolvam e que a colaboração se beneficie de uma gama muito mais ampla de expertise. Em outras palavras, a recompensa da modularidade implacável e consciente é que ninguém precisa entender todo o projeto em detalhes, mas pode contribuir onde for mais capaz. O efeito geral é como um estaleiro virtual. Muitas pessoas diferentes estão espalhadas por todos os lugares, contribuindo para as diferentes partes do navio, em esforços separados, cada um modesto em tamanho e escopo. Mas o produto agregado é notável.

É claro que a modularidade não é o fim da história. É apenas um padrão único que ajuda a ampliar a colaboração. As unidades modulares são os átomos de atenção a partir dos quais a arquitetura da atenção é construída. O ideal, como vimos, é criar uma arquitetura em que essas unidades modulares sejam organizadas de forma que cada participante veja as tarefas em que tem maior vantagem comparativa e, portanto, possa dar a maior contribuição. Ferramentas existentes, como blogs, wikis e rastreadores de problemas, fazem isso apenas de forma imperfeita. Mas, a longo prazo, veremos gradualmente a

surgimento de uma ciência de design da atenção, que nos ajuda a construir ferramentas que melhor utilizem a atenção especializada disponível.

E o Linux? Linus Torvalds desistiu há muito tempo de tentar acompanhar toda a comunidade de desenvolvedores do kernel Linux. Em maio de 2000, um usuário da lista de discussão do kernel Linux reclamou que Torvalds não estava respondendo às suas postagens. Torvalds respondeu o seguinte:

> Observe que ninguém lê todas as postagens no linux-kernel. Aliás, ninguém que espera ter tempo de sobra para realmente trabalhar no kernel lerá nem metade. Exceto Alan Cox [um dos tenentes de Torvalds], mas ele não é humano, e sim cerca de mil gnomos trabalhando em cavernas subterrâneas em Swansea. Nenhum dos gnomos lê todas as postagens, eles apenas trabalham muito bem juntos.
>
> De qualquer forma, alguns de nós nem conseguem ler todos os nossos e-mails pessoais, simplesmente porque recebem muitos. Eu faço o meu melhor.

O Linux cresceu muito desde que Torvalds escreveu aquele post. Hoje, ninguém, nem mesmo o super-humano Alan Cox, consegue acompanhar todo o trabalho em andamento. A beleza da colaboração Linux é que ela é organizada de forma que ninguém precise.

Reutilização Radical e o Information Commons

A modularidade é importante, mas existe um padrão ainda mais básico de colaboração subjacente ao código aberto: a capacidade dos programadores de código aberto de reutilizar e modificar o trabalho uns dos outros. Isso pode parecer tão óbvio a ponto de ser indigno de consideração, mas tem algumas consequências surpreendentes. O impacto óbvio, é claro, é que os programadores não precisam começar do zero, mas podem desenvolver e aprimorar gradativamente o que outros já fizeram. Efetivamente, os programadores de código aberto estão construindo um patrimônio comum de informações compartilhadas publicamente. Esse patrimônio comum não está localizado em nenhum lugar específico, mas consiste em todo o código-fonte aberto distribuído em uma miríade de locais pela internet. Isso permite uma divisão dinâmica do trabalho, na qual o código de uma pessoa pode posteriormente ser aprimorado por outras pessoas que ela nunca conheceu, com

expertise e necessidades das quais talvez nunca tenham ouvido falar. Quanto mais rico o acervo de informações comuns se torna, mais poderosa é a base para a colaboração.

Em conjunto, a comunidade de programadores de código aberto está criando um acervo de informações notavelmente ativo e rico. Um estudo realizado por dois cientistas da empresa de software SAP, Amit Deshpande e Dirk Riehle, mostra que o acervo agora contém mais de um bilhão de linhas de código disponíveis publicamente e está crescendo a uma taxa de mais de 300 milhões de linhas por ano. Quer adicionar chamas ao seu filme caseiro como efeito especial? Existem pacotes de software de código aberto para isso. Quer controlar seu telescópio robótico doméstico? Dependendo do seu telescópio, pode muito bem haver software de código aberto para isso. Softwares de código aberto estão disponíveis para realizar uma gama quase inimaginavelmente ampla de tarefas.

O surgimento desse rico acervo de informações mudou radicalmente a maneira como os programadores trabalham. Antes, os programadores escreviam seus programas em grande parte do zero. Seus heróis eram pessoas que conseguiam, em poucos dias, criar um programa que programadores menos experientes levariam meses para escrever. Para lhe dar uma ideia de quais habilidades eram valorizadas naquela época, considere esta história de um dos grandes pioneiros da computação moderna, Alan Kay, ganhador do Prêmio Turing, a mais alta honraria da ciência da computação. É uma história admirável sobre a proeza de programação de Donald Knuth, outra lenda da computação e ganhador do Prêmio Turing:

> Quando eu estava em Stanford com o projeto [de inteligência artificial] [no final da década de 1960], uma das coisas que costumávamos fazer todo Dia de Ação de Graças era um concurso de programação com pessoas em projetos de pesquisa na região da Baía de São Francisco. O prêmio, eu acho, era um peru.
>
> [Pioneiro da inteligência artificial e professor de Stanford, John] McCarthy costumava inventar os problemas. No ano em que Knuth participou, ele ganhou o melhor tempo para executar o programa e também o melhor tempo de execução do algoritmo. Ele fez isso no pior sistema [...] E ele basicamente deu uma surra em todo mundo.

Hoje em dia, a programação mudou. Hoje em dia, um grande programador não é apenas alguém que consegue resolver um problema rapidamente do zero. Um grande programador é alguém que também domina a informação.

commons, alguém que, quando solicitado a resolver um problema, sabe como montar e adaptar rapidamente código extraído dos commons e como equilibrar isso com a necessidade de escrever código adicional do zero. Esse mestre pode se basear no trabalho de outros para resolver problemas de forma mais rápida e confiável do que outros programadores menos experientes. É um tipo de colaboração passiva, cuja eficácia aumenta à medida que o commons de informação cresce. Antes mesmo de escreverem uma única linha de código, os programadores de hoje frequentemente se baseiam no trabalho de milhares de outros programadores. Como alguns programadores gostam de dizer: "Bons programadores codificam; grandes programadores reutilizam o código de outras pessoas".

Na programação, o commons de informação decolou no início da década de 1990, com a ampla adoção da internet. Mas, em uma forma mais primitiva, as ideias de reutilização e commons de informação foram pioneiras séculos antes, na ciência. Quando alguém publica uma descoberta científica — digamos, o famoso artigo de Einstein contendo a fórmula $E = mc^2$ — outros cientistas podem reutilizar esse resultado em seus próprios artigos, simplesmente citando a derivação original. Isso permite que os cientistas desenvolvam o trabalho anterior sem ter que repeti-lo. A citação credita o descobridor original e fornece um elo em uma cadeia de evidências. Se alguém quiser saber por que $E = mc^2$, basta seguir a citação até o artigo original de Einstein. O resultado é que, como na programação moderna, um grande cientista não é meramente uma pessoa capaz de penetrar enormemente em insights sobre a natureza, mas alguém que também domina o commons de informação — conhecimento científico já publicado — e a capacidade de desenvolver esse conhecimento. A ciência é, nesse sentido, uma grande colaboração, construída sobre o commons de informação.

O recurso comum de informação baseado em citações da ciência é poderoso, mas trabalhoso e lento quando comparado, digamos, ao padrão rápido de reutilização em um projeto como a Wikipédia ou o Linux. Um cientista que utilizasse o padrão Wikipédia e Linux — reutilizando o texto de outra pessoa palavra por palavra, mas fazendo algumas melhorias aqui e ali — provavelmente receberia uma nota de indignação (ou pior) do autor original. No entanto, tais melhorias são a força vital de muitas colaborações online, permitindo melhorias iterativas extremamente rápidas, com as pessoas focadas exclusivamente em avançar, não em refazer o que já é conhecido. Um artigo moderadamente ativo da Wikipédia pode ser modificado 20 ou 30 vezes por uma dúzia de pessoas diferentes em uma semana. Obter o mesmo acúmulo cumulativo de ideias em muitas áreas da ciência pode levar anos. Projetos como o Projeto Polímata aceleram o processo de construção cumulativa da ciência convencional, criando um espaço compartilhado onde os cientistas podem rapidamente construir sobre

as ideias uns dos outros. A citação é talvez a técnica mais poderosa para a construção de um patrimônio de informação comum que poderia ser criado com a tecnologia do século XVII. Mas, como demonstra o Projeto Polímata, e como exploraremos com mais detalhes posteriormente, as tecnologias modernas agora possibilitam uma maneira melhor.

A Competição MathWorks

Em 1998, uma empresa de software chamada MathWorks começou a realizar um concurso de programação de computadores duas vezes por ano, aberto a qualquer pessoa no mundo. Para cada concurso, a MathWorks propõe um problema de programação aberto. Para lhe dar uma ideia do concurso, considere o problema usado no primeiro concurso, em 1998, um problema chamado problema de empacotamento de CD: escrever um programa que, ao receber uma longa lista de músicas, escolha uma sublista que se aproxime o máximo possível de preencher a duração de 74 minutos de um CD. Por exemplo, seu programa pode ser solicitado a escolher músicas do catálogo antigo do Pink Floyd. Você executa seu programa e ele encontra uma lista de músicas do catálogo que deixa apenas 35 segundos de espaço extra no CD. Mas se seu programa tivesse uma maneira melhor de selecionar músicas, você poderia se ver com apenas 15 segundos restantes no CD.

O problema da compactação de CDs parece artificial. Poucas pessoas precisam gravar CDs o mais próximo possível do tamanho ideal. Apesar disso, o problema é exatamente o tipo de desafio que muitos programadores apreciam. É um problema simples e de fácil compreensão, mas que pode ser abordado de diversas maneiras. Como todas as competições da MathWorks, a competição original foi muito popular, atraindo mais de 100 participantes do mundo todo.

Cada programa inscrito em um concurso da MathWorks recebe uma pontuação que reflete tanto a rapidez com que o programa é executado (programas mais rápidos recebem pontuações melhores) quanto a eficácia com que resolve o problema. No caso do empacotamento de CDs, os programas que chegaram mais perto de preencher o CD receberam pontuações melhores. Os participantes podem inscrever seus programas a qualquer momento durante a semana de competição e podem enviar várias inscrições ou versões da mesma inscrição. As inscrições são pontuadas automaticamente assim que enviadas, e as pontuações são imediatamente colocadas em uma tabela de classificação.
(Voltaremos a falar sobre como a pontuação automatizada é feita em breve.)

Os elogios são concedidos a quem estiver no topo da tabela de classificação, mesmo que brevemente. Assim, em vez de esperar até o final da semana para enviar suas inscrições, as pessoas as enviam ao longo de toda a semana. O vencedor geral do concurso é a pessoa que estiver no topo da tabela de classificação ao final da competição.

O que torna a competição MathWorks especial é que cada vez que alguém envia uma inscrição, o código do programa é imediatamente disponibilizado para que outras pessoas baixem e reutilizem. Ou seja, qualquer um pode entrar, "roubar" o código de outra pessoa, alterá-lo para obter uma melhoria e, em seguida, reenviá-lo como seu, possivelmente ultrapassando a outra pessoa no ranking. Essa capacidade de reutilizar o código de outras pessoas tem consequências espetaculares. Os programas líderes são constantemente ajustados por pequenas alterações, muitas vezes alterando apenas uma única linha de código em uma inscrição anterior. As mudanças ocorrem de forma rápida e intensa, e alguns participantes ficam viciados, movidos pelo feedback instantâneo e pela sensação de que estão a apenas uma ideia do topo do ranking.
Um concorrente escreveu:

> Comecei a ficar "obcecado". Em casa, embora eu seja pai de três filhos, meu trabalho em tempo integral era trabalhar no concurso. Eu trabalhava talvez 10 horas depois do expediente todos os dias. Na quinta-feira, ficou claro que eu não conseguiria trabalhar a sério (pelo meu trabalho), então tirei um dia de folga na sexta-feira.

É semelhante ao ciclo rápido de feedback que torna os jogos de computador viciantes. Você sempre pode ter mais uma chance de fazer uma pequena melhoria. É discutível se isso é sempre bom — o participante citado parece precisar tirar umas férias do computador —, mas esse foco incansável também produz resultados incríveis.

O progresso do concurso é vividamente ilustrado pelo gráfico da figura 4.2. O eixo horizontal representa o tempo, enquanto o eixo vertical representa a pontuação: para o problema de embalagem de CDs, pontuações mais baixas são melhores. Cada ponto no gráfico representa uma inscrição na competição. As pontuações caíram tão drasticamente durante o concurso que o eixo vertical foi redimensionado — as pontuações no topo são centenas de vezes maiores do que as pontuações na parte inferior. A linha contínua marca a melhor pontuação em um determinado momento. Como você pode ver, há grandes passos ocasionais na linha, indicando ideias inovadoras que melhoram substancialmente a melhor pontuação. Após esse avanço, geralmente há um período em que as pessoas fazem muitos pequenos ajustes na pontuação princip

entrada, encontrando pequenas melhorias que otimizem ainda mais o programa e os deixem com a melhor pontuação.

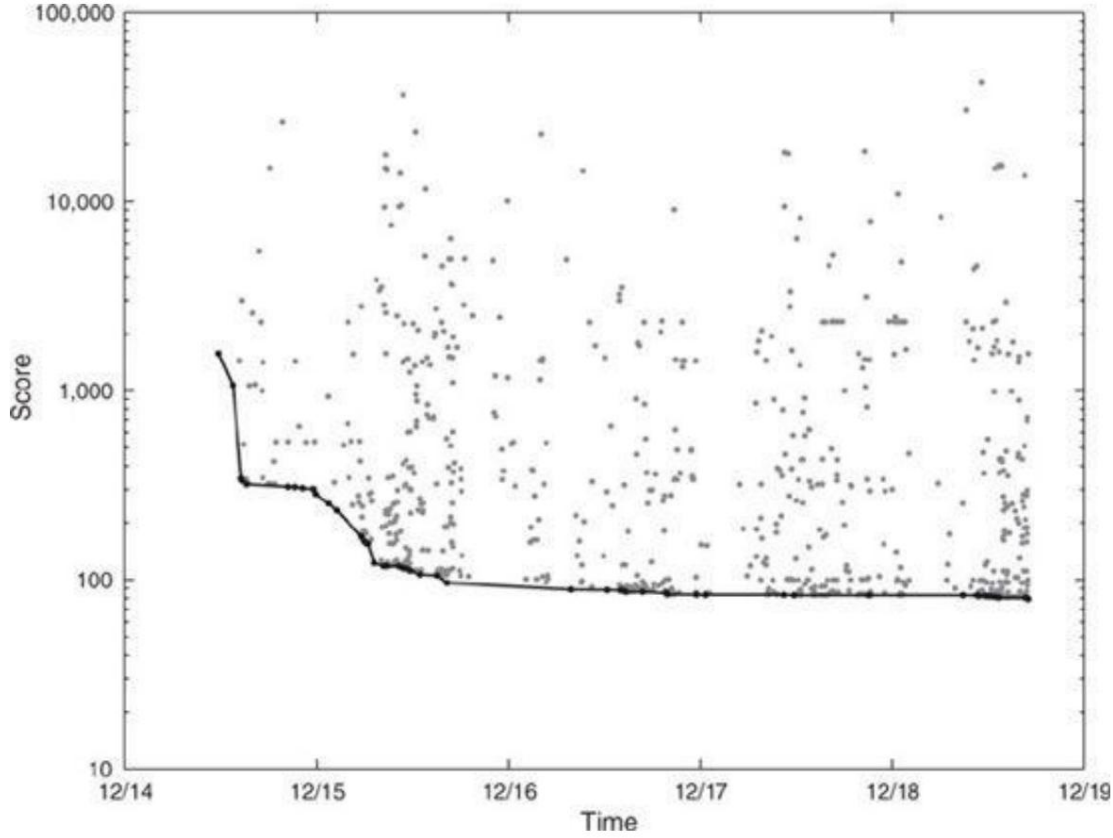


Figura 4.2. Progresso das pontuações na competição de programação da MathWorks. Pontuações mais baixas são melhores. Crédito: Copyright 2011 The MathWorks, Inc. Usado com permissão. Agradecimentos a Ned Gulley por fornecer a figura.

A diferença entre as melhores inscrições iniciais e o vencedor final é drástica. No concurso de empacotamento de CDs, as melhores inscrições iniciais foram executadas rapidamente, mas deixaram seis minutos de espaço no CD sem uso. O programa vencedor foi executado aproximadamente na mesma velocidade, mas deixou apenas 20 segundos sem uso, uma melhoria de quase 20 vezes. Ele utilizou contribuições de pelo menos nove pessoas, em dezenas de inscrições separadas. Embora seja uma competição, o concurso MathWorks funciona, portanto, em muitos aspectos, como uma colaboração em larga escala. O organizador do concurso, Ned Gulley, disse sobre o programa vencedor: "nenhuma pessoa no planeta poderia ter escrito um talgoritmo tão oimizado. No entanto, ele apareceu no final do concurso, esculpido do nada por pessoas de todo o mundo, a maioria das quais nunca se conheciam antes". Isso não foi um acaso. O concurso de empacotamento de CDs foi o primeiro de mais de vinte competições MathWorks realizadas até o momento. Cada concurso vê o mesmo surgimento gradual de

um programa cuja construção está, sem dúvida, além da capacidade de qualquer um dos concorrentes individuais.

Microcontribuição

A competição Mathworks ilustra vividamente um padrão que pode ser usado para escalar a colaboração online: microcontribuição. O tipo mais comum de inscrição na competição MathWorks é aquela que altera apenas uma ***única*** linha de código em alguma inscrição anterior. Isso mesmo, alguém entra e altera apenas ***uma*** linha de código em uma inscrição anterior — muito possivelmente a inscrição de outra pessoa! — e a reenvia como se fosse sua. O segundo tipo mais comum de inscrição altera apenas duas linhas. E assim por diante. O resultado é que, mesmo com as pessoas competindo, a evolução das inscrições principais parece quase uma conversa, com muitas idas e vindas, à medida que o bastão da liderança passa de um participante para outro. É uma troca criativa de ideias que impulsiona a melhoria gradual ao longo do tempo, com diferentes pessoas contribuindo da melhor forma possível.

O mesmo padrão de microcontribuição é usado em muitas colaborações online. Na Wikipédia, a edição mais comum em um artigo altera apenas uma única linha desse artigo. No Linux, as contribuições mais comuns alteram apenas uma única linha de código. Um estudo realizado por dois cientistas da empresa de software SAP, Oliver Arafat e Dirk Riehle, mostrou que esse padrão é bastante geral: na maioria dos projetos de software de código aberto, a alteração mais comum é em apenas uma única linha de código, a segunda alteração mais comum é em duas linhas, e assim por diante. No Projeto Polymath, o líder do projeto, Tim Gowers, pediu aos participantes que compartilhassem apenas uma única ideia em cada contribuição e que resistissem à tentação de desenvolver ideias extensivamente por conta própria.

A microcontribuição reduz a barreira à contribuição, incentivando mais pessoas a se envolverem e também aumentando a gama de ideias contribuídas por cada pessoa em particular. Consequentemente, aumenta a gama de expertise disponível para a colaboração. Lembre-se de Yasha, o membro da Equipe Mundial que contribuiu com o lance crucial número 26. Yasha teria se perdido jogando contra Kasparov sozinho. Mas foi muito útil, talvez vital, para a Equipe Mundial ter acesso à pequena contribuição de Yasha. Pequenas contribuições despertam ideias e insights, à medida que as pessoas

compartilhar ideias que eles não conseguiriam desenvolver sozinhos, mas que podem inspirar outros. Se um participante do Projeto Polymath ou da competição MathWorks estivesse sem ideias, bastava esperar algumas horas, observando novas ideias que o estimulassem e desafiassem. Ou então, ele poderia vasculhar os arquivos em busca de novos estímulos a partir de ideias antigas. A microcontribuição, portanto, ajuda a construir uma comunidade vibrante, uma sensação de que algo está acontecendo, de que o progresso está sendo feito, de que mesmo quando você, pessoalmente, está bloqueado, outras pessoas estão fazendo as coisas avançarem. A microcontribuição é um padrão poderoso de colaboração, em suma, porque as pequenas contribuições ajudam a colaboração a explorar rapidamente uma gama muito mais ampla de ideias do que seria possível de outra forma.

Pontuações como sinais para coordenar especialistas Atenção

Eu disse anteriormente que as inscrições na competição MathWorks são pontuadas automaticamente assim que são enviadas, mas deixei passar como isso é feito. Imagine que você é um dos organizadores da competição e um dos competidores acaba de enviar seu programa. Como você deve avaliá-lo?

A coisa óbvia a fazer (e a maneira como isso é realmente feito) é executar o programa em algumas entradas de teste. Você pode testá-lo em (digamos) três entradas de teste: o catálogo dos Beatles, uma coleção de peças de jazz e uma coleção de música dance. Assim, na primeira execução, o programa tentaria preencher um CD com músicas escolhidas do catálogo dos Beatles, na segunda execução, usaria músicas da coleção de jazz e, na terceira execução, músicas da coleção dance. Você então atribuiria ao programa uma pontuação determinada pela rapidez com que ele é executado e pela eficiência com que ele preenche todo o CD em cada uma das três entradas de teste. É claro que não há necessidade de um organizador fazer isso manualmente. Tudo pode ser feito automaticamente assim que as inscrições são enviadas, para que a pontuação possa ser calculada imediatamente. A única ressalva é que, para que isso funcione, os organizadores precisam manter as entradas de teste em segredo — se os concorrentes soubessem, por exemplo, que seu programa seria usado no catálogo dos Beatles, eles poderiam adaptá-lo especificamente para o catálogo dos Beatles, anulando o objetivo da competição. Mas, desde que

os organizadores têm o cuidado de manter as informações dos testes em segredo, eles podem pontuar as inscrições automaticamente assim que elas são enviadas.

A pontuação automatizada é importante porque ajuda os participantes a concentrar sua atenção onde ela trará mais benefícios. Se alguém altera um programa e causa um grande salto (ou até mesmo uma pequena melhoria) na pontuação, outras pessoas percebem e verificam o que foi alterado: talvez essa pessoa tenha uma ótima ideia nova. A pontuação automatizada facilita, portanto, para os programadores acompanharem as melhores ideias uns dos outros — mesmo que o número de participantes seja muito grande — e identificar oportunidades de usar sua própria experiência para fazer melhorias adicionais, ultrapassando assim os outros. Alguns dos programadores, por exemplo, são especialistas nos detalhes da linguagem de programação (chamada MATLAB) usada na competição. Eles observam os programas de outras pessoas cuidadosamente e usam seu conhecimento de MATLAB para fazer pequenas otimizações, muitas vezes alterando apenas uma ou duas linhas de código MATLAB para torná-lo mais eficiente, reduzindo assim uma fração de milissegundo do tempo de execução. Outros competidores se especializam de outras maneiras. Alguns vasculham a literatura científica em busca de inspiração. Outros fazem brainstormings de abordagens completamente novas. E alguns trabalham na hibridização de abordagens existentes. Em meio a todas essas abordagens diferentes, a pontuação automatizada desempenha um papel semelhante ao dos preços em um mercado, fornecendo informações que podem ser usadas para embasar a tomada de decisões pelos participantes do concurso. Embora seja impraticável conduzir uma conversa envolvendo as mais de 100 pessoas inscritas na competição MathWorks — ninguém tem tempo para prestar atenção a mais de 100 vozes distintas —, a pontuação ajuda as pessoas a tomarem boas decisões sobre onde concentrar sua atenção e, assim, impulsiona uma rápida melhoria.

A pontuação do MathWorks não é perfeita como forma de coordenar a atenção. Como as mesmas informações de pontuação são fornecidas a todos, isso leva os competidores a concentrarem sua atenção de maneiras semelhantes. Por exemplo, se alguém salta para o topo da tabela de classificação, muitos participantes imediatamente desviarão sua atenção para essa entrada. É claro que alguma concentração de atenção é boa, mas se todos seguirem a mesma liderança, o grupo como um todo pode negligenciar direções promissoras para exploração. Você poderia imaginar mecanismos de sinalização mais complexos que espalhariam a atenção de forma mais ampla e levariam a uma melhor alocação de conhecimento. Por exemplo, pessoas com experiência em otimização de código MATLAB poderiam ser direcionadas para programas cuja estrutura bruta estivesse mudando rapidamente, mas cujos detalhes finos ainda não tivessem sido otimizados. Ou talvez pudesse haver alguma maneira de detectar grupos de programas que fazem uso de

de ideias semelhantes. Os participantes que gostaram de hibridizar diferentes abordagens poderiam usar essas informações para ajudá-los a escolher os melhores programas em cada grupo e tentar hibridizá-los.

Deixando essas limitações de lado, a pontuação do MathWorks faz um ótimo trabalho ajudando a coordenar a atenção e, portanto, a escala de colaboração do MathWorks. Como forma de direcionar a atenção, ela funciona muito mais efetivamente do que, por exemplo, qualquer mecanismo disponível no Projeto Polímata, que dependia da perspicácia dos indivíduos para avaliar quais contribuições valiam a pena serem acompanhadas. Podia levar horas ou dias para os polímatas identificarem as melhores ideias novas. Isso é rápido, especialmente quando comparado ao ritmo normal da pesquisa científica, mas lento em comparação com a imediatez da pontuação do MathWorks. A situação em Kasparov versus o Mundo foi semelhante à do Projeto Polímata, embora ferramentas como a árvore de análise de Krush ajudassem a coordenar a atenção. Quanto melhor a arquitetura da atenção for em direcionar a atenção dessa maneira, mais a inteligência coletiva será amplificada.

Convertendo Insights Individuais em Coletivos

Entendimento

Além de coordenar a atenção, a pontuação do MathWorks também serviu ao importante propósito de ajudar a transformar os insights de participantes individuais em insights coletivos de todo o grupo. Cada vez que alguém tinha uma ideia que melhorava um programa, isso se refletia em sua pontuação, tornando o valor da nova ideia imediatamente evidente para todos os participantes. Para que a colaboração seja bem-sucedida, deve haver alguma maneira de converter insights individuais em insights coletivos. Em outras palavras, a colaboração precisa saber o que a colaboração sabe.

Kasparov versus o Mundo mostra o que acontece quando uma colaboração converte apenas imperfeitamente insights individuais em insights coletivos. Como vimos, a Equipe Mundial contou com Irina Krush e seus colegas para identificar e divulgar as melhores ideias da Equipe Mundial. Sem a habilidade de Krush em avaliar e comparar análises, a Equipe Mundial provavelmente teria se saído muito pior em agregar as melhores ideias. É claro que, embora Krush e seus colegas tenham se esforçado bastante, seu manual

A abordagem não foi tão rápida ou objetiva quanto a pontuação automatizada na competição MathWorks. Como resultado, grande parte da expertise disponível na Equipe Mundial foi desperdiçada. Muitos enxadristas experientes participaram da Equipe Mundial e, embora alguns tenham gostado da experiência, outros se sentiram alienados, acreditando que suas ideias se perdiam no ruído geral da discussão. Anos depois da partida, um participante escreveu em um fórum online:

> Se alguma coisa na minha vida da qual participei e que eu poderia rotular como um exemplo perfeito de como uma comunidade NÃO deve resolver um problema, foi a partida KvW (da qual participei bastante e sou um mestre (fide [classificação de xadrez] 2276)).

Tal descontentamento ocorreu porque Krush e alguns colegas estavam integrando manualmente as melhores ideias de milhares de pessoas. Seus esforços foram notáveis, mas, é claro, eles só conseguiram realizar o trabalho de forma imperfeita.
Isso causou frustração ocasional na Equipe Mundial e, quase certamente, algumas oportunidades perdidas. Esta é uma regra geral: quanto mais eficaz uma colaboração converter insights individuais em insights coletivos, mais eficaz será a colaboração.

De fato, o sistema da Equipe Mundial para converter insights individuais em insights coletivos falhou gravemente em um ponto crucial da partida. Como mencionei anteriormente, até o lance 51, a partida oscilava entre Kasparov e o Mundo, sem que nenhum dos lados obtivesse vantagem decisiva. No lance 51, Kasparov estava em uma posição ligeiramente mais forte, e a Equipe Mundial lutava por um empate. Infelizmente, no lance 51, um membro da Equipe Mundial chamado José Unodos alegou ter conseguido burlar o sistema de votação da Microsoft e ter manipulado a votação em favor de uma jogada da qual ele gostava pessoalmente, mas isso não foi considerado uma jogada forte por Krush e pela maioria dos outros jogadores de ponta da Equipe Mundial.
O lance preferido de José Unodos venceu a votação, a primeira vez desde o lance 9 que a recomendação de Krush não foi jogada pela Equipe Mundial. O evento ajudou a pender a balança do jogo a favor de Kasparov e prejudicou o moral da Equipe Mundial. Onze lances depois, Kasparov venceu, num triste final para uma das grandes partidas da história do xadrez. Quando as maneiras de um grupo converter a percepção individual em percepção coletiva falham, a inteligência coletiva deixa de funcionar. No próximo capítulo, veremos que, em alguns campos, tais falhas impõem limites fundamentais à inteligência coletiva.

CAPÍTULO 5

Os Limites e o Potencial do Coletivo Inteligência

A inteligência coletiva não é uma panaceia para a resolução de problemas. Neste capítulo, identificaremos um critério fundamental que separa os problemas em que a inteligência coletiva pode ser aplicada dos problemas em que ela não pode. Em seguida, usaremos esse critério para entender por que os problemas científicos são especialmente adequados para o ataque da inteligência coletiva.

Para entender o critério, vamos primeiro nos voltar para um experimento realizado em 1985 pelos psicólogos Garold Stasser e William Titus. O que Stasser e Titus mostraram é que grupos que discutem um determinado tipo de problema — uma decisão política — frequentemente se saem surpreendentemente mal no uso de todas as informações que possuem. Isso talvez não pareça tão surpreendente: afinal, a discussão política cotidiana nem sempre é muito informativa. Mas o que Stasser e Titus mostraram foi muito além: a discussão em grupo às vezes torna as decisões políticas das pessoas piores do que teriam sido se tivessem tomado essas decisões individualmente.

Stasser e Titus começaram criando perfis escritos de três candidatos fictícios à presidência do grêmio estudantil da Universidade de Miami, onde Stasser era docente. Os perfis continham informações sobre as políticas dos candidatos em relação a questões de interesse dos alunos — horários de visitação nos dormitórios, leis locais sobre consumo de bebidas alcoólicas e assim por diante. Stasser e Titus construíram deliberadamente os três perfis de forma que um dos candidatos fosse claramente mais desejável do que os outros dois. Fizeram isso primeiro entrevistando os alunos para descobrir quais características eles consideravam desejáveis e, em seguida, construindo os perfis de acordo. Daremos um nome a esse candidato extra-desejável: o chamaremos de "Melhor".

Na primeira versão do experimento, cada aluno recebeu perfis completos dos três candidatos e foi solicitado a decidir quem era seu candidato preferido. Não surpreendentemente, 67% dos alunos escolheram o Melhor. Stasser e Titus então dividiram os alunos em pequenos grupos de quatro pessoas cada e pediram aos grupos que discutissem qual candidato

deveria ser presidente. Ao final da discussão, os alunos foram novamente questionados sobre seu candidato preferido. O apoio a Best aumentou para 85%.

Até agora, sem surpresas. Mas Stasser e Titus também realizaram uma segunda versão do experimento. Desta vez, alteraram os perfis para que cada aluno recebesse apenas informações ***parciais*** sobre os três candidatos: removeram algumas das informações positivas sobre Best — coisas que se esperava que os alunos gostassem — e também removeram algumas das informações negativas sobre um dos candidatos indesejáveis. De fato, qualquer perfil ***parcial*** sugeria que um dos candidatos indesejáveis era, na verdade, melhor que Best. Não surpreendentemente, quando solicitados a escolher um candidato com base nesses perfis parciais, 61% dos alunos preferiram o candidato indesejável, enquanto apenas 25% preferiram Best. Depois disso, Stasser e Titus dividiram novamente os alunos em pequenos grupos de quatro e pediram que os grupos discutissem qual candidato deveria ser presidente. Mas aqui está a parte inteligente: quando Stasser e Titus estavam construindo os perfis parciais, eles tiveram o cuidado de remover informações ***diferentes*** de diferentes perfis, para que cada ***grupo*** de alunos ainda tivesse ***todas*** as informações sobre os três candidatos. Assim, cada grupo ainda tinha todas as informações necessárias para identificar Best como o verdadeiro melhor candidato. Observe que os alunos foram avisados com antecedência de que nem todos no grupo tinham necessariamente as mesmas informações sobre os três candidatos.

Agora, nesta segunda versão do experimento, seria de se esperar que a porcentagem de Best aumentasse após a discussão em grupo, à medida que as pessoas compartilhassem o que sabiam e percebessem que Best era realmente o melhor candidato. Mas não foi isso que aconteceu. De fato, após a discussão, foi o candidato indesejável cuja porcentagem aumentou, de 61% para 75%. A porcentagem de Best, na verdade, diminuiu, de 25% para 20%. Os grupos não estavam tanto compartilhando informações, mas sim reforçando as ideias preconcebidas dos alunos. Em outras palavras, a discussão em grupo não melhorou as decisões dos grupos, mas as piorou. Foi um caso de estupidez coletiva, não de inteligência coletiva.

O que estava acontecendo? Vimos muitos exemplos mostrando como grupos podem usar sua inteligência coletiva para ter um desempenho melhor do que qualquer indivíduo no grupo. No entanto, o experimento Stasser-Titus mostra que a discussão às vezes faz com que os grupos se saiam pior do que a média de seus membros. Além disso, o experimento Stasser-Titus faz parte de um conjunto muito mais amplo de descobertas em psicologia de grupo que mostram que grupos — mesmo

pequenos grupos, ou grupos de especialistas — muitas vezes têm dificuldade em aproveitar seu conhecimento coletivo.

Por exemplo, em 1989, em um acompanhamento do experimento original de Stasser-Titus, as discussões em grupo foram gravadas para que os experimentadores pudessem entender melhor como os grupos chegaram às suas decisões. O que eles descobriram foi que, em vez de explorar todas as informações disponíveis, os grupos passavam a maior parte do tempo discutindo informações que tinham em ***comum.*** Assim, por exemplo, se várias pessoas soubessem que Best tinha uma posição impopular sobre (por exemplo) visitas a dormitórios, era provável que houvesse uma discussão relativamente longa sobre esse fato, e a informação provavelmente seria mencionada novamente na discussão. Mas quando alguém no grupo tinha uma informação ***única*** sobre um candidato, uma informação que só ele sabia, a discussão dessa informação geralmente era superficial. Isso importava, porque no experimento original de Stasser-Titus, informações negativas sobre Best eram frequentemente compartilhadas por vários membros do grupo, enquanto informações positivas eram frequentemente compartilhadas por apenas um único membro.

Em 1996, outro experimento de acompanhamento foi realizado, desta vez em um hospital universitário, solicitando a grupos que fizessem diagnósticos médicos com base em videoclipes de entrevistas com pacientes. Novamente, as informações eram parciais: cada pessoa do grupo assistiu a apenas parte da entrevista em vídeo. Os grupos que tomaram as decisões incluíam três pessoas de diferentes status: um residente médico, um interno e um estudante. De forma alarmante, mas talvez não surpreendente, os grupos prestaram muito mais atenção às informações exclusivas do residente médico de alto status. Informações exclusivas dos internos e estudantes tinham muito mais probabilidade de serem ignoradas.

Esses e muitos outros estudos pintam um quadro sombrio para a inteligência coletiva. Eles mostram que os grupos frequentemente não aproveitam bem seu conhecimento coletivo. Em vez disso, concentram-se no conhecimento que possuem em comum, no conhecimento de membros de alto status do grupo e, frequentemente, ignoram o conhecimento de membros de baixo status do grupo. Por isso, não conseguem converter o conhecimento individual em conhecimento coletivo compartilhado pelo grupo.
E isso é uma má notícia se você está tentando usar inteligência coletiva.

Os limites da inteligência coletiva

Por que projetos como o Projeto Polímata, Kasparov versus o Mundo e a competição MathWorks são tão bem-sucedidos, enquanto os grupos nos experimentos Stasser-Titus e relacionados apresentam desempenho tão baixo? Em termos mais precisos, por que os grupos nos projetos bem-sucedidos conseguiram converter seus melhores insights individuais em insights coletivos, enquanto os grupos nos experimentos Stasser-Titus e relacionados não conseguiram essa conversão? A diferença se deveu meramente a diferenças nos processos utilizados nos respectivos casos? Ou existe alguma diferença mais fundamental, uma diferença que não pode ser resolvida por um processo aprimorado, talvez devido à natureza dos problemas em discussão?

Para responder a essas perguntas, quero que você considere um pequeno quebra-cabeça. Darei uma descrição verbal do quebra-cabeça, mas ele é bastante visual, e você pode achar útil consultar a explicação pictórica apresentada na imagem e na legenda na próxima página. Você recebe um tabuleiro de xadrez vazio de oito por oito e precisa cobri-lo com peças de dominó de um por dois, de modo que apenas dois quadrados permaneçam descobertos: o quadrado inferior esquerdo e o quadrado superior direito. Você consegue fazer isso? Se sim, como?
Se não, por que não? Não é permitido empilhar dominós, quebrá-los ou deixá-los pendurados na borda do tabuleiro — tudo na descrição do quebra-cabeça deve ser interpretado da maneira usual.
Para simplificar ainda mais as coisas, também exigiremos que cada dominó cubra dois quadrados adjacentes no tabuleiro — dominós colocados obliquamente não são permitidos.

A maioria das pessoas não acha isso um quebra-cabeça fácil. Mas vale a pena lutar com ele por alguns minutos antes de continuar lendo. Se você fizer isso, e você tentar colocar dominós imaginários (ou reais) em um tabuleiro de xadrez, você descobrirá que não importa o quanto você tente, você não consegue fazer isso. É como se houvesse uma obstrução invisível que está de alguma forma impedindo você de ter sucesso. Na verdade, não há nenhuma maneira de cobrir o tabuleiro da maneira solicitada. Aqui está o porquê. A chave é notar que se você colocar um dominó no tabuleiro, não importa onde você o coloque, ele cobrirá um total de um quadrado preto e um quadrado branco. Então, se você colocar dois dominós, haverá um total de dois quadrados pretos cobertos e dois quadrados brancos.

Três dominós significam três quadrados pretos cobertos e três quadrados brancos cobertos. E assim por diante. Não importa quantos dominós você coloque, a quantidade total de preto e branco cobertos será a mesma. Mas observe que tanto os quadrados inferior esquerdo quanto superior direito do tabuleiro são pretos. Então, para chegar a uma situação em que eles sejam os únicos quadrados

Sem cobertura, você precisa cobrir de alguma forma 32 quadrados brancos e 30 quadrados pretos. É um número desigual, então não tem como.

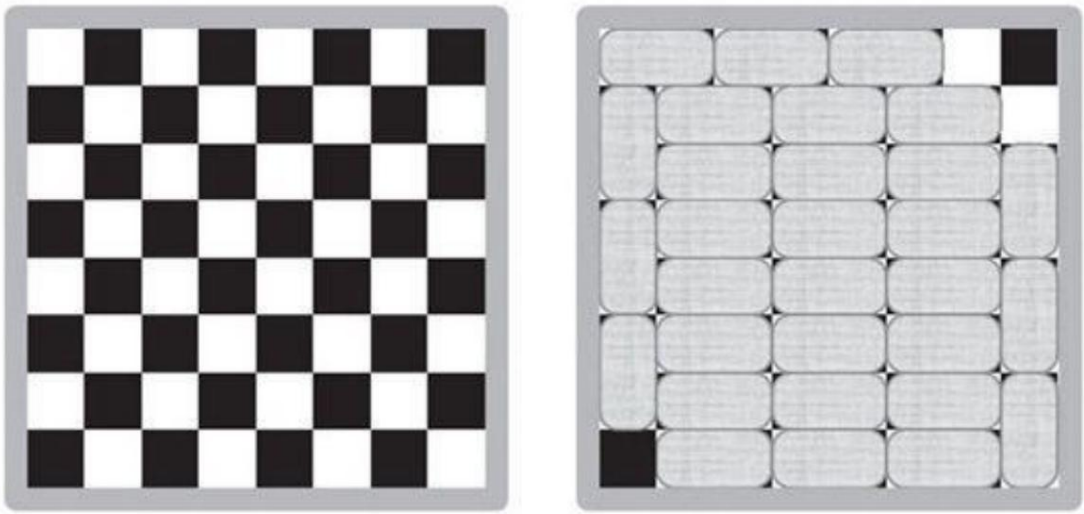


Figura 5.1. O quebra-cabeça começa com um tabuleiro de xadrez vazio de oito por oito, como mostrado à esquerda. Pergunta-se se é possível cobrir o tabuleiro com dominós de um por dois, de modo que apenas os quadrados inferior esquerdo e superior direito permaneçam descobertos. À direita, mostrei uma tentativa frustrada de fazer isso, que deixa dois quadrados extras no canto superior direito descobertos.

Embora a maioria das pessoas ache difícil resolver esse quebra-cabeça, quando a solução é explicada, elas rapidamente dizem: "Ahá, entendi!". É muito mais fácil reconhecer o insight que resolve o problema do que tê-lo. Em outras palavras, existe uma lacuna entre a dificuldade de reconhecer o insight e a dificuldade de tê-lo em primeiro lugar. Uma lacuna semelhante está presente em exemplos como o Projeto Polímata, Kasparov versus o Mundo e a competição MathWorks. Considere a competição MathWorks. É preciso uma engenhosidade tremenda para escrever programas que rapidamente enchem CDs quase cheios de músicas. Mas, como vimos, é fácil reconhecer quando alguém escreveu um bom programa: basta executar o programa em algumas entradas de teste e verificar se ele roda rápido e deixa pouco espaço livre no CD. É essa lacuna entre a dificuldade de escrever programas e a facilidade de avaliá-los que alimenta o progresso coletivo na competição MathWorks. No xadrez, reconhecer insights valiosos não é tão simples, mas um enxadrista competente como Krush consegue reconhecer e compreender uma análise excepcionalmente perspicaz de uma posição específica, mesmo que não pudesse ter elaborado a análise sozinho. As melhores análises podem até estimular o mesmo sentimento de "Aha, que inteligente!" que no quebra-cabeça de dominó. Krush não sabe jogar

consistentemente no nível de Kasparov, mas ela é boa o suficiente para reconhecer quando outras pessoas estão (pelo menos momentaneamente) jogando nesse nível e para entender suas análises. E no Projeto Polímata, os participantes conseguiram reconhecer quando outros tinham insights matemáticos que superavam os seus e conseguiram incorporar esses insights ao seu conhecimento coletivo. Novamente, é aquela sensação de "Aha!" estimulada por um insight inteligente.
Cada projeto, portanto, usou essa lacuna entre nossa capacidade de ter e reconhecer insights úteis, a fim de converter insights individuais em insights coletivos.

O problema nos experimentos de Stasser-Titus é que as discussões em pequenos grupos não converteram de forma confiável o insight individual em insight coletivo. Intelectualmente, muitos dos alunos participantes dos experimentos sem dúvida concordariam que a solução seria reunir sistematicamente todas as suas informações e, então, tomar uma decisão com base nos perfis combinados assim construídos. Mas, na prática, não o fizeram.
E, dado o contexto, isso não é surpreendente. Nas discussões políticas cotidianas, a maioria de nós não avalia os políticos construindo um quadro completo de suas posições. Estamos ocupados demais tentando descobrir como suas posições se relacionam com nossos valores e interesses.

Suponha, no entanto, que os grupos no experimento Stasser-Titus tivessem de fato iniciado sua discussão reunindo sistematicamente todas as suas informações. Esse experimento nunca foi feito, que eu saiba, mas acho que podemos ter certeza de que mudaria drasticamente o resultado. Portanto, o problema nos grupos Stasser-Titus foi, em parte, uma falha de processo; um processo aprimorado resultaria em resultados dramaticamente melhores. Mas não foi apenas uma falha de processo. Mesmo que os grupos tivessem compartilhado informações sistematicamente, diferentes alunos ainda teriam diferenças de opinião insolúveis. Se um aluno adora beber e festejar, enquanto outro se opõe fortemente à bebida por motivos religiosos, eles podem nunca concordar em escolhas políticas, não importa quão bom seja o processo utilizado.

Isso aponta para um requisito fundamental que deve ser atendido se quisermos ampliar a inteligência coletiva: os participantes devem compartilhar um conjunto de conhecimentos e técnicas. É esse conjunto de conhecimentos e técnicas que eles usam para colaborar. Quando esse conjunto compartilhado existir, o chamaremos de ***práxis compartilhada,*** termo derivado da palavra ***práxis,*** que significa a aplicação prática do conhecimento. A disponibilidade de uma práxis compartilhada determina se a inteligência coletiva pode ser ampliada ou não.

Como exemplo de práxis compartilhada, imagine um grande grupo trabalhando em conjunto no problema do dominó. Assim que qualquer pessoa do grupo descobre que o problema do dominó é impossível de resolver, ela consegue convencer os outros rapidamente, porque cada passo do seu raciocínio é evidentemente correto: todos nós compartilhamos as mesmas habilidades básicas de raciocínio.

Este é um exemplo de práxis compartilhada. De forma semelhante, existe uma práxis compartilhada para o trabalho em matemática — todos os métodos padrão de raciocínio matemático e normas sobre o discurso matemático — e é por isso que os participantes do Projeto Polímata conseguiam reconhecer e concordar sobre quando o progresso matemático estava sendo feito. Da mesma forma, a pontuação na competição MathWorks definiu implicitamente uma práxis compartilhada: qualquer mudança em um programa que melhorasse a pontuação era entendida pelos participantes como progresso. No xadrez, a práxis compartilhada não é tão forte quanto na matemática e na programação de computadores: até mesmo os melhores enxadristas às vezes discordam sobre o valor de diferentes análises.

No entanto, há um grande corpo de conhecimento sobre xadrez que é amplamente aceito pelos jogadores fortes, e esse conhecimento compartilhado significa que os jogadores mais fortes da Equipe Mundial geralmente concordam sobre quais análises são melhores.

Todos esses são exemplos de problemas em que há uma práxis compartilhada. Mas, para muitos problemas, não há práxis compartilhada. Por exemplo, como vimos, não existe uma práxis compartilhada forte disponível na política. As pessoas podem facilmente discordar sobre valores básicos. E se um grupo não tiver essa práxis compartilhada, surgirão desacordos que não poderão ser resolvidos. Uma vez que surja um desacordo insolúvel, a comunidade começará a se fragmentar em torno desse desacordo, limitando a capacidade de ampliar a colaboração.

Agora, para problemas isolados — digamos, o de adivinhar o peso de um boi em uma feira rural inglesa (veja a página 7) — isso talvez não importe muito. Mas, para trabalhar em conjunto em múltiplas etapas para resolver um problema, essa fragmentação impõe limites fundamentais à escala da colaboração.

A política é apenas uma das muitas áreas que carecem de uma práxis compartilhada forte. O mesmo se aplica a muitas das belas-artes, onde a avaliação de obras criativas costuma ser altamente controversa. Por exemplo, para decidir qual de duas pinturas é melhor, utilizamos nossos próprios padrões estéticos, padrões que podem ser bem diferentes daqueles adotados por outras pessoas.

Da mesma forma, podemos discordar razoavelmente sobre qual das duas composições musicais é melhor. Isso não quer dizer que não exista a noção de um padrão objetivo nas artes. Praticamente todos concordam que os Beatles são melhores do que uma boy band qualquer. Mas, em

Comparando os Beatles a Bach, pessoas razoáveis podem discordar. Ao fazer essa afirmação, sem dúvida ofendi esnobes da música em todo o mundo. Mas a questão é que ofendi tanto os esnobes da música clássica, que não conseguem acreditar que os Beatles comecem a se comparar a Bach, quanto os esnobes da música pop, que acreditam que Bach pertence a uma tradição que já foi superada. Quando tais tradições coexistem, é extremamente difícil para as pessoas nas duas tradições colaborarem, porque não têm base para concordar quando estão fazendo progresso compartilhado. Este não é um julgamento negativo sobre tais campos — grandes músicos, pintores e políticos operam perto dos limites da capacidade humana —, mas é uma limitação importante sobre quando a inteligência coletiva pode ser usada.

Não são apenas a política e as belas-artes que não possuem uma forte práxis compartilhada. Muitas áreas acadêmicas também carecem de uma. Pense na crítica da literatura inglesa. Os críticos não vão um dia largar suas penas e chegar a um entendimento comum de Shakespeare. De fato, chegar a esse entendimento comum não é o objetivo. Em tais áreas, a pluralidade de pontos de vista é uma característica, não um defeito, e uma nova maneira de entender Shakespeare deve ser celebrada. Mas essa mesma pluralidade de pontos de vista dificulta o reconhecimento e a integração dos melhores insights de um grande grupo de pessoas. Qualquer tentativa de colaboração desse tipo inevitavelmente esbarra em discussões sobre valores básicos e questionamentos sobre o que torna uma contribuição valiosa. Em tais áreas, o acordo não é escalável, e isso limita severamente nossa capacidade de converter insights individuais em insights coletivos, impedindo, assim, a aplicação da inteligência coletiva.

Alguns campos situam-se perto do limite que divide os campos onde é possível escalar a inteligência coletiva e aqueles onde isso não é possível. Em economia, por exemplo, existem muitos métodos poderosos de raciocínio com os quais a maioria dos economistas concorda: a compreensão de como o comércio pode beneficiar a todos, a ideia de que imprimir mais dinheiro geralmente causa inflação, e assim por diante. Mas os economistas ***discordam*** sobre algumas das questões mais fundamentais da economia. Como diz a velha piada, se você colocar cinco economistas em uma sala, eles darão seis opiniões completament
O presidente dos EUA, Harry Truman, supostamente pediu um economista maneta, alguém que não pudesse dizer "Por outro lado". Portanto, embora haja uma práxis compartilhada na economia, ela não é tão forte quanto a práxis compartilhada em áreas como matemática, programação de computadores e xadrez. Como resultado, há muitas questões em economia que não podem ser atacadas pelos métodos da inteligência coletiva. Elas estão presentes apenas em algumas áreas da economia, como o estudo de alguns modelos matemáticos de finanças e

a economia, onde uma forte práxis compartilhada está disponível. É nessas áreas da economia que a inteligência coletiva pode ser ampliada.

A disponibilidade de uma práxis compartilhada não é o único desafio na aplicação da inteligência coletiva. Existem muitos outros problemas práticos. Um exemplo é o pensamento de grupo, em que os membros de um grupo podem estar mais interessados em se relacionar uns com os outros do que em avaliar ideias criticamente. Ou os grupos podem se tornar câmaras de eco, com os membros do grupo apenas reforçando as opiniões existentes uns dos outros. Em alguns grupos, as normas básicas de comportamento civilizado se rompem. Esse tipo de colapso destruiu muitas colaborações em software de código aberto e atormenta muitos fóruns mal projetados na web, que podem se tornar refúgios para trolls da internet e outros comportamentos antissociais. Os projetos que discutimos superaram esses e outros problemas semelhantes: alguns obtiveram sucesso com louvor (o Projeto Polímata), enquanto outros tiveram sucesso por pouco (as deliberações da Equipe Mundial às vezes oscilaram à beira do colapso por falta de civilidade). Problemas semelhantes também afligem grupos offline, e muito já foi escrito sobre os problemas e como superá-los — incluindo livros como " ***A Sabedoria das Multidões",*** de James Surowiecki, " ***Infotopia",*** de Cass Sunstein , e muitos outros livros sobre negócios e comportamento organizacional. Embora esses problemas práticos sejam importantes, eles frequentemente podem ser resolvidos com um bom processo. Mas, independentemente da qualidade do processo, permanece uma linha divisória fundamental: se uma práxis compartilhada está disponível. Em campos onde uma práxis compartilhada está disponível, podemos escalar a inteligência coletiva e obter grandes melhorias qualitativas no comportamento de resolução de problemas, como a serendipidade projetada e a massa crítica conversacional. Para campos sem uma práxis compartilhada, as ferramentas online não nos proporcionam a mesma mudança qualitativa.

As práticas compartilhadas da ciência

A ciência é adequada para a inteligência coletiva. A maioria dos campos da ciência possui um vasto repositório de técnicas poderosas, compartilhadas pelos cientistas que nela trabalham. Existem padrões amplamente aceitos para o que significa que um argumento, análise ou procedimento experimental esteja correto. Isso foi ilustrado vividamente pelo Projeto Polímata, onde a discussão foi conduzida em um tom notavelmente civilizado. Nesses raros

Nas ocasiões em que houve desacordo, geralmente era porque alguém havia cometido um erro flagrante de raciocínio. Alguém apontava o erro, sem rancor, e o autor da discussão imediatamente reconhecia o erro. Isso não quer dizer que os participantes nunca se envolveram em especulações, mas eles cuidadosamente marcaram suas especulações como tal e não as apresentaram como fatos incontestáveis. Em quase todas as questões cruciais, os participantes concordaram rapidamente sobre quando uma linha de argumentação estava certa e quando estava errada, e sobre quando uma ideia era promissora e quando não era. Foi esse rápido acordo que possibilitou a ampliação da colaboração.

Como ilustração de quão firmemente esses padrões são mantidos na ciência, considere o trabalho do jovem Albert Einstein, não o ícone científico que conhecemos hoje, mas um desconhecido funcionário de 26 anos que trabalhava no escritório de patentes suíço, incapaz de encontrar um emprego como físico profissional. Dessa posição de obscuridade, em 1905, Einstein publicou seus famosos artigos sobre a relatividade especial, mudando radicalmente nossas noções de espaço, tempo, energia e massa. Outros cientistas anteciparam parcialmente as conclusões de Einstein, mas nenhum expôs com tanta ousadia e força as consequências completas da relatividade especial. As propostas de Einstein eram surpreendentes, mas seus argumentos eram tão convincentes que seu trabalho foi publicado em um dos principais periódicos de física de sua época e rapidamente aceito pela maioria dos físicos renomados. Como é notável que um forasteiro, um virtualmente desconhecido, pudesse vir desafiar muitas de nossas crenças mais fundamentais sobre como o universo funciona. E, em pouco tempo, a comunidade de físicos essencialmente disse: "Sim, você está certo".

Como outro exemplo, considere a descoberta da estrutura do DNA. Esta descoberta foi feita por James Watson e Francis Crick, utilizando dados atribuídos em parte a Rosalind Franklin. Os três eram cientistas jovens e desconhecidos: Watson tinha 24 anos e Crick, 36, reestabelecendo-se após uma breve carreira na física e trabalho no Almirantado Britânico durante a Segunda Guerra Mundial. Franklin tinha 32. Quem os competiu na descoberta foi o principal químico do mundo, Linus Pauling. Mais de uma década antes, o brilhante Pauling fizera uma série de descobertas que lhe renderiam o Prêmio Nobel de Química. Se conseguisse desvendar a estrutura do DNA, outro prêmio certamente viria. Em certo momento da corrida, ele deu um tremendo susto em Watson e Crick, anunciando que havia encontrado a estrutura. Mas Watson e Crick conversaram com o filho de Pauling, Peter Pauling, que lhes mostrou a estrutura proposta por Pauling pai para o DNA. Para seu espanto, rapidamente perceberam que Pauling estava errado: o maior químico do mundo havia cometido um erro simples em química básica, um erro

Seus próprios livros didáticos deveriam tê-lo alertado. Watson e Crick retomaram seu trabalho com intensidade renovada e logo depois encontraram a estrutura correta. Quando isso aconteceu, não importava que Pauling fosse mundialmente famoso enquanto Watson, Crick e Franklin eram desconhecidos. A comunidade científica rejeitou o trabalho de Pauling e aclamou a dupla hélice como uma das descobertas científicas do século.

Os exemplos de Einstein e de Watson, Crick e Franklin ilustram a força da práxis compartilhada na ciência. De forma incomum em muitos aspectos da vida, na ciência, muitas vezes, quem vence é a pessoa com as melhores evidências e os melhores argumentos, e não a pessoa com a maior reputação e o maior poder. Pauling pode ter sido amplamente reconhecido como o principal químico do mundo, mas outros químicos podiam ver, com tanta certeza quanto Watson e Crick, que a estrutura de Pauling estava simplesmente errada. Essa forte práxis compartilhada torna a ciência bem adequada à inteligência coletiva.

Essa forte práxis compartilhada não significa que a ciência seja um processo limpo e simples. O processo cotidiano de fazer ciência é confuso, especulativo e repleto de erros e argumentos. O cientista Richard Feynman era tão cheio de ondas cerebrais irreprimíveis e "grandes" ideias, muitas das quais mais tarde se provaram erradas, que, segundo seu biógrafo, James Gleick, seus colegas mais astutos desenvolveram uma regra prática: "Se Feynman diz três vezes, está certo". O mesmo poderia ser dito de muitos cientistas. Frequentemente, um cientista inicia uma investigação com pouco mais do que um vislumbre de uma ideia, uma suspeita de que alguma hipótese seja verdadeira. Eles esboçam uma maneira de testá-la, muitas vezes vagamente no início, preenchendo gradualmente mais e mais detalhes. Os experimentos muitas vezes precisam ser realizados muitas vezes, com o desenho experimental gradualmente alterado e aprimorado, à medida que o cientista entende melhor quais evidências são necessárias para ser convincente. Tudo isso é um processo lento que envolve muita especulação, argumentação e falsos começos, à medida que o cientista gradualmente avança para argumentos e evidências cada vez mais robustos. O objetivo final, porém, é um conjunto de princípios e evidências que estejam de acordo com a prática compartilhada do campo.

E isso é bem diferente de uma discussão sobre Bach versus os Beatles, ou uma discussão política, ou uma discussão sobre Shakespeare, onde, no final, pode permanecer uma divisão fundamental sobre valores básicos. É claro que cientistas ainda publicam, às vezes, artigos errados, equivocados ou pouco convincentes. Mas mesmo quando um cientista publica tal resultado, outros cientistas podem voltar e repetir os experimentos para encontrar falhas ou apontar deficiências nos argumentos. Em suma, eles podem testar novamente os resultados em relação à prática compartilhada da área e considerá-los insuficientes. É isso

capacidade de errar de forma clara, o que permite o progresso. Nesse sentido, a ciência já é, como eu disse antes, uma grande colaboração, mantida unida por padrões comuns de evidência e raciocínio.

Existem áreas da ciência sem uma práxis compartilhada, áreas mais como a economia, por exemplo, onde os problemas são tão desafiadores que o campo ainda é uma protociência, com conhecimento e técnicas compartilhados apenas começando a emergir? Por exemplo, um dos grandes problemas em aberto da física é o de encontrar uma teoria quântica da gravidade — uma teoria única que unifique a mecânica quântica e a teoria da gravidade de Einstein. É um dos problemas mais difíceis da física, um problema que tem derrotado as melhores mentes por décadas. Na década de 1980, uma abordagem para o problema conhecida como teoria das cordas ganhou destaque e gradualmente passou a dominar o trabalho sobre gravidade quântica. Ao mesmo tempo, um número muito menor de físicos continuou a buscar outras abordagens para a gravidade quântica. Nos últimos anos, um debate chamado por alguns de "guerra das cordas" tem sido travado entre os defensores das diferentes abordagens. Muitos físicos afirmam que a teoria das cordas é a única abordagem razoável para a gravidade quântica. Outros, incluindo Stephen Hawking, Roger Penrose e Lee Smolin, acreditam que vale a pena buscar abordagens diferentes. Notavelmente, alguns proeminentes teóricos das cordas descartam os que não estudam a teoria das cordas não apenas como errados, mas como equivocados, ou até mesmo como tolos. Quando ocorre uma divisão tão fundamental, é quase impossível que grandes grupos colaborem entre si. A inteligência coletiva só pode ser aplicada dentro das respectivas tribos, onde há uma práxis compartilhada. E tais colaborações precisam ser cuidadosamente protegidas contra a interrupção pela tribo rival.

A situação na gravidade quântica é incomum. Na maioria das áreas da ciência, os cientistas podem comparar duas explicações concorrentes de um fenômeno a um experimento e perceber que uma explicação está correta (ou, pelo menos, não descartada pelo experimento) e a outra está errada. Ou um cientista pode apontar uma falha no procedimento experimental de outro, e todos concordarão que, sim, isso é realmente uma falha, não atende ao padrão esperado. Mas na gravidade quântica, os fenômenos em estudo são tão remotos que ainda não sabemos como fazer experimentos — ainda é tudo teoria. E desenvolver a teoria básica é tão desafiador que escolher premissas iniciais tornou-se, em certa medida, uma questão de gosto pessoal, de maneira semelhante às belas-artes. São essas condições altamente incomuns que impediram o desenvolvimento de uma práxis compartilhada. Em contraste, na maioria dos outros campos da ciência, existe uma práxis compartilhada fortemente arraigada. E assim a ciência nos dá uma oportunidade maravilhosa de ampliar nossa inteligência coletiva.

Usando a Inteligência Coletiva na Ciência

Na parte 2 deste livro, vimos como ferramentas online podem ser usadas para ampliar a inteligência coletiva, tornando os grupos mais inteligentes e mais inteligentes. Ao chegarmos ao final da parte 1, vamos usar essas ideias para imaginar algumas das maneiras pelas quais ferramentas online podem ser usadas para ampliar a inteligência coletiva na ciência. Adotaremos um ponto de vista pessoal, tentando imaginar algumas das maneiras pelas quais essas ferramentas podem impactar o cotidiano de um cientista. Nos próximos capítulos, veremos como alguns desses sonhos estão sendo realizados e até superados hoje. Também veremos como outras partes desses sonhos são bloqueadas pelas práticas sociais atuais na ciência — e como isso pode ser mudado.

Imagine que estamos alguns anos no futuro e você é um físico teórico trabalhando no Instituto de Tecnologia da Califórnia (Caltech), em Pasadena. Todas as manhãs, você começa seu trabalho sentando-se diante do seu computador, que apresenta uma lista de dez pedidos de ajuda, uma lista que foi elaborada especialmente para você a partir de milhões de pedidos semelhantes registrados durante a noite por cientistas do mundo todo. De todos esses pedidos, esses são os problemas em que você provavelmente terá a máxima vantagem comparativa. Hoje, um dos pedidos imediatamente chama sua atenção. Um cientista de materiais em Budapeste, Hungria, está trabalhando em um projeto para desenvolver um novo tipo de cristal. Durante o projeto, surgiu uma dificuldade inesperada envolvendo um tipo de problema muito especializado: descobrir o comportamento de partículas enquanto elas saltam aleatoriamente ("difusas") em uma rede triangular. Infelizmente para o cientista de materiais, a difusão é um assunto sobre o qual ele não sabe muito.

Você, por sua vez, não sabe muito sobre cristais, mas é um especialista em matemática da difusão e, de fato, já resolveu vários problemas de pesquisa semelhantes ao que intriga o cientista de materiais. Depois de refletir sobre o problema da difusão por alguns minutos, você tem certeza de que o problema se encaixará facilmente em técnicas matemáticas que você conhece bem, mas que o cientista de materiais provavelmente desconhece.

Você envia uma mensagem ao cientista de materiais com um esboço de uma solução para o problema dele. Nos dias seguintes, vocês se comunicam, desenvolvendo juntos uma solução, preenchendo muitos detalhes e traduzindo suas ideias matemáticas para a linguagem da ciência dos materiais. Ainda há muito trabalho a ser feito no projeto original, mas um gargalo crítico...

foi superado. Sua recompensa é um colaborador feliz, uma eventual coautoria em um artigo e o prazer de aprender um pouco sobre a física dos cristais e como ela se relaciona com sua expertise em difusão. A recompensa do seu colaborador é economizar centenas de horas que, de outra forma, teriam sido gastas para se tornarem especialistas o suficiente para resolver o problema da difusão. A comunidade como um todo também é recompensada: com sua ajuda, o problema foi resolvido muito mais rápido e a um custo menor do que seria de outra forma, os resultados científicos obtidos são mais sólidos e a explicação dos resultados no artigo publicado é mais clara. Todos se beneficiam da sua vantagem comparativa — você tem as habilidades para resolver rapidamente um problema que levaria semanas para o cientista de materiais resolver. Cada um de vocês faz o que sabe fazer de melhor — e a sociedade economiza milhares de dólares.

Na mesma manhã em que tudo isso começa, você percebe outro pedido marcante em sua lista de pedidos mais bem classificados. Vem de um estudante em Bangalore, Índia, que quer ajuda para aprender sobre pesquisas recentes sobre o uso de algoritmos computacionais para simular sistemas quânticos complexos. Ele não conhece nenhum especialista local e está aprendendo com artigos online, que consideram confusos em alguns pontos. Você recebeu o pedido porque é especialista em tais algoritmos e pode responder facilmente às perguntas do aluno. Além disso, você solicitou ao seu sistema que o alertasse sobre alguns pedidos de assistência de alunos a cada semana, adaptados a áreas em que você tem especialização. Uma rápida troca de mensagens com o aluno ocorre nos próximos dias, esclarecendo grande parte da confusão. Seu trabalho com o aluno é automaticamente registrado em um arquivo de sua atividade científica, juntamente com estatísticas que mostram sua contribuição para a divulgação pública.

Alguns outros pedidos também aparecem na sua lista de pedidos mais bem classificados, mas você decide que não tem tempo para ajudar. Entre eles, há vários outros pedidos de colaboração bastante semelhantes ao do cientista de materiais, embora com detalhes diferentes; um pedido de assistência de uma escola local; e um pedido de material de leitura de um aluno cujo tema de tese coincide com vários dos seus artigos antigos. Todos esses pedidos serão vistos por dezenas ou centenas de outras pessoas, a maioria das quais, como você, possui uma especialização intimamente relacionada aos pedidos.
A resposta é voluntária e nenhuma das solicitações é direcionada somente a você.

Tudo isso é possível graças a um algoritmo de classificação que prioriza os milhões de pedidos de assistência feitos diariamente, para que você veja apenas os pedidos em que você, pessoalmente, provavelmente tem o maior interesse e a maior vantagem comparativa. O algoritmo de classificação leva em consideração

Considere suas áreas de especialização, as solicitações às quais você já respondeu, o histórico de quem as fez e suas preferências, como seu desejo de ajudar os alunos. Ao selecionar as solicitações criteriosamente, você pode maximizar o impacto do seu trabalho.

Em todo o mundo, padrões semelhantes se repetem milhões de vezes. Uma cientista cognitiva em Ottawa está tentando replicar um experimento que mostra como uma ilusão de ótica específica pode ser suprimida pela mudança da cor de algumas partes da ilusão. Quando começou a trabalhar, ela tentou descobrir como replicar o experimento apenas a partir de uma compreensão ampla do experimento original. Ela obteve bons progressos, mas ocasionalmente travava, e então consultou vídeos online que mostravam o experimento sendo realizado em dois outros laboratórios.

Isso ajudou, mas ela ainda está com dificuldades para reproduzir os resultados. Depois de vários dias atolada, ontem à noite ela enviou um pedido de ajuda, na esperança de encontrar alguém com experiência tanto em ilusões de ótica quanto em como o sistema nervoso combina as informações de cores vindas dos diferentes cones do olho. Esta manhã, ela recebeu uma ligação de um psicofísico em Iowa, que enviou um esquema de cores modificado e algumas instruções sobre como recalibrá-lo, se necessário. Rapidamente, ela resolve o problema e o experimento está pronto para ser realizado.

Enquanto isso, em um laboratório de pesquisa em Xangai, na China, um biólogo trabalha até tarde da noite, sequenciando geneticamente uma cepa do vírus da gripe. Ao terminar o sequenciamento, ele consulta bancos de dados online para comparar a composição genética do vírus com todas as cepas virais conhecidas. Ele descobre que se trata, como suspeitava, de uma nova variação da gripe. Nas próximas semanas, ele desenvolverá uma vacina para o novo vírus. Para desenvolver a vacina, ele usa um software que extrai informações de dezenas de bancos de dados online, efetivamente perguntando e recebendo respostas para milhares de perguntas sobre vírus, seus genes, as proteínas que produzem e o efeito dessas proteínas. Mas, diferentemente de nossos exemplos anteriores, essas perguntas não são feitas como solicitações únicas, como informações. Em vez disso, o software faz as perguntas e recebe as respostas de forma automatizada, quase invisível para o cientista, entrelaçando o conhecimento adquirido por dezenas de milhares de biólogos e, em seguida, recombinando esse conhecimento para ajudar a fazer uma nova descoberta.

Milhões de conexões como essas estão sendo feitas em todo o mundo. Cientistas cujo trabalho está atualmente bloqueado por problemas científicos complexos estão sendo conectados a outros cientistas que têm a expertise necessária para resolvê-los rapidamente. É um mercado online de atenção especializada, uma espécie de mercado de colaboração que torna todos mais eficientes e

capazes, mais aptos a trabalhar em problemas nos quais têm vantagem comparativa, e deixando o restante do trabalho para outras pessoas. Nesse mercado de colaboração, o tipo de conexão que hoje só acontece por acaso, na verdade, acontece por design. Ao mesmo tempo em que essas conexões entre cientistas são feitas, uma troca de conhecimento mais silenciosa, porém muito maior, acontece em segundo plano, à medida que os cientistas baixam e processam vastas quantidades de dados, aproveitando assim o conhecimento previamente adquirido por milhares de outros cientistas. Este também é um mercado de colaboração, mas em vez de perguntas especializadas e pontuais, é para perguntas tão padronizadas que podem ser respondidas automaticamente.

Vamos voltar ao nível pessoal, de volta a Pasadena e Caltech.
Além da sua nova colaboração com o cientista de materiais na Hungria, você passa a maior parte do dia trabalhando em um de seus projetos em andamento, um empreendimento ambicioso para projetar um computador quântico.
Computadores quânticos são computadores hipotéticos que utilizam a mecânica quântica para resolver problemas inviáveis em computadores convencionais. Embora computadores quânticos de larga escala prometam ser dispositivos notáveis, construí-los é um enorme desafio, pois os estados quânticos são muito delicados. Para enfrentar esse desafio, há seis meses você e dois colegas iniciaram um projeto para projetar um computador quântico que realmente possa ser ampliado. Seu projeto envolve uma abordagem especial à computação quântica chamada computação quântica topológica, uma abordagem que se baseia em insights de diversos campos da ciência, desde o campo matemático da topologia até a física dos supercondutores, e da fabricação de semicondutores a todos os detalhes da teoria da computação quântica. O projeto cresceu rapidamente e agora envolve mais de 100 cientistas do mundo todo, colaborando online.

Alguns desses cientistas são teóricos, com expertise diversificada nas diversas áreas envolvidas. Mas a maioria é formada por experimentalistas, incluindo alguns dos maiores especialistas do mundo em supercondutores e semicondutores, bem como cientistas de materiais especializados na preparação de amostras de materiais de alta qualidade. Esses experimentalistas compartilham seus truques e dicas sobre o que é possível fazer nos laboratórios mais avançados, o tipo de conhecimento popular que distingue os laboratórios de vanguarda dos que estão um passo atrás.

A colaboração nem sempre progrediu suavemente em direção ao seu objetivo. Mas mesmo quando obstáculos aparentemente intransponíveis surgiram, muitas vezes foi possível superá-los usando o mesmo mercado de colaboração que permitiu o início da sua colaboração húngara.

esta manhã. Isso também ajudou a atrair novas pessoas e novos especialistas para a colaboração. À medida que a colaboração cresceu, tornou-se seu maior compromisso contínuo, e na maioria dos dias você dedica pelo menos uma ou duas horas ao projeto. Foi muito mais longe do que você imaginava inicialmente, pois a colaboração contornou obstáculos que você considerava intransponíveis e, à medida que as ideias dos colaboradores passaram de especulativas para mais viáveis, alguns dos laboratórios envolvidos estão começando a prototipar algumas dessas ideias.

Alguns leitores — especialmente, talvez, aqueles que trabalharam como cientistas — podem ler os parágrafos acima e achar que parecem uma utopia. "Por que", podem perguntar, "esses experimentalistas se ajudariam dessa maneira? No mundo real, eles nunca compartilhariam as ideias-chave que constituem sua vantagem competitiva". Hoje, isso é verdade, e voltaremos a esse problema sob diferentes formas repetidamente nos próximos capítulos. Mas, à medida que nossa compreensão se aprofunda, veremos que, embora seja um problema desafiador, não é intransponível. Por enquanto, porém, adiaremos a discussão.

Estas são apenas algumas ideias para estimular sua reflexão sobre como ferramentas online e inteligência coletiva podem ser usadas para mudar a ciência. É claro que muito mais é possível. Imagine abordagens totalmente de código aberto para a realização de pesquisas. Imagine uma rede online conectada de conhecimento científico que integra e conecta dados, códigos de computador, cadeias de raciocínio científico, descrições de problemas em aberto e muito mais. Essa rede de conhecimento científico poderia incorporar vídeos, mundos virtuais e realidade aumentada, bem como mídias mais convencionais, como artigos científicos.
E estaria fortemente integrado a uma rede social científica que direciona a atenção dos cientistas para onde ela é mais valiosa, liberando um enorme potencial de colaboração.

Na parte 2 deste livro, exploraremos, em termos concretos, como a era da ciência em rede está se configurando hoje. Veremos, por exemplo, como vastos bancos de dados contendo grande parte do conhecimento mundial estão sendo explorados em busca de descobertas que escapariam a qualquer ser humano sem auxílio. Veremos como as ferramentas online nos permitem construir novas instituições que atuam como pontes entre a ciência e o restante da sociedade de novas maneiras, e que podem ajudar a redefinir a relação entre ciência e sociedade. O lugar onde essas ideias estão sendo mais plenamente concretizadas é na ciência básica e, portanto, o foco da parte 2 é na ciência básica — em contraste, a ciência aplicada é frequentemente realizada por pequenos grupos trabalhando em segredo, dentro de empresas privadas, e esse sigilo limita sua capacidade de ampliar a colaboração. Mas mesmo na ciência básica, existem sérios obstáculos a serem superados. Ideias simples

Mercados colaborativos, como artigos de pesquisa de código aberto semelhantes a wikis e compartilhamento de dados e códigos de computador, enfrentam obstáculos culturais consideráveis. Desenvolveremos a ideia de que, para que a ciência em rede alcance seu pleno potencial, ela deve ser ***ciência aberta,*** baseada em uma cultura na qual os cientistas compartilhem aberta e entusiasticamente todos os seus dados e conhecimento científico. E, por fim, veremos como essa cultura científica mais aberta pode ser criada.

PARTE 2

Ciência em Rede

CAPÍTULO 6

Todo o conhecimento do mundo

Don Swanson parece uma pessoa improvável de fazer descobertas médicas. Cientista da informação aposentado, mas ainda ativo, da Universidade de Chicago, não possui formação médica, não realiza experimentos médicos e nunca teve um laboratório. Apesar disso, ele fez diversas descobertas médicas significativas. Uma das primeiras foi em 1988, quando investigou enxaquecas e descobriu evidências sugerindo que elas são causadas por deficiência de magnésio. Na época, a ideia surpreendeu outros cientistas que estudavam enxaquecas, mas a de Swanson foi posteriormente testada e confirmada em múltiplos ensaios terapêuticos por grupos médicos tradicionais.

Como é possível que alguém sem qualquer formação médica pudesse fazer tal descoberta? Embora Swanson não tivesse nenhuma das credenciais convencionais da pesquisa médica, o que ele tinha era uma ideia inteligente. Swanson acreditava que o conhecimento científico havia se tornado tão vasto que conexões importantes entre os assuntos estavam passando despercebidas, não porque fossem especialmente sutis ou difíceis de compreender, mas porque ninguém tinha um conhecimento científico amplo o suficiente para perceber essas conexões: em um palheiro grande o suficiente, até mesmo uma agulha de 15 metros pode ser difícil de encontrar. Swanson esperava descobrir essas conexões ocultas usando um mecanismo de busca médica chamado Medline, que permite pesquisar milhões de artigos científicos na área médica — você pode pensar no Medline como um mapa de alto nível do conhecimento médico humano. Ele começou seu trabalho usando o Medline para pesquisar na literatura científica conexões entre enxaquecas e outras condições. Aqui estão dois exemplos de conexões que ele encontrou: (1) as enxaquecas estão associadas à epilepsia; e (2) as enxaquecas estão associadas à formação de coágulos sanguíneos mais facilmente do que o normal. É claro que as enxaquecas têm sido objeto de muitas pesquisas, e essas são apenas duas de uma lista muito mais longa de conexões que ele encontrou. Mas Swanson não se limitou a essa lista. Em vez disso, ele analisou cada uma das condições associadas e, em seguida, usou o Medline para encontrar outras conexões com cada condição. Ele descobriu que, por exemplo, (1) a deficiência de magnésio aumenta a suscetibilidade a

epilepsia; e (2) a deficiência de magnésio faz com que o sangue coagule mais facilmente. Quando começou seu trabalho, Swanson não tinha ideia de que acabaria conectando enxaquecas à deficiência de magnésio. Mas, depois de encontrar alguns artigos sugerindo essas conexões em dois estágios entre deficiência de magnésio e enxaquecas, ele restringiu sua busca para se concentrar na deficiência de magnésio, eventualmente encontrando onze dessas conexões em dois estágios com enxaquecas. Embora esse não fosse o tipo de evidência tradicionalmente favorecido por cientistas médicos, ainda assim apresentou um argumento convincente de que enxaquecas estão conectadas à deficiência de magnésio. Antes do trabalho de Swanson, alguns artigos sugeriram, de forma provisória (e principalmente de passagem), que a deficiência de magnésio poderia estar conectada a enxaquecas.
Mas o trabalho anterior não foi convincente e foi ignorado pela maioria dos cientistas. Em contraste, as evidências de Swanson foram altamente sugestivas e logo foram seguidas por ensaios terapêuticos que confirmaram a conexão entre enxaqueca e magnésio.

Se você sofre de enxaquecas, sabe que a descoberta da conexão entre enxaqueca e magnésio não resultou em uma cura ou um tratamento infalível. Hoje, a deficiência de magnésio é apenas um dos muitos fatores que contribuem para as enxaquecas, e a causa primária das enxaquecas permanece obscura e objeto de debate. No entanto, descobrir a conexão entre enxaqueca e magnésio foi um passo significativo para entender o que causa as enxaquecas e como tratá-las. Além disso, a importância do trabalho de Swanson vai muito além da medicina. Embora tenha se tornado senso comum da nossa época lamentar a explosão de informações, como se o aumento maciço do nosso conhecimento fosse de alguma forma algo ruim, Swanson inverteu esse ponto de vista. Ele via o crescimento do conhecimento não como um problema, mas como uma oportunidade. Ele percebeu que ferramentas como o Medline expandem nossa capacidade de encontrar significado no conhecimento coletivo da humanidade e, assim, nos permitem descobrir padrões no todo que são invisíveis para humanos sem ajuda. Nenhuma mente humana jamais poderia abranger os milhões de experimentos indexados pelo Medline.

Felizmente, ninguém precisa disso. Trabalhando em simbiose com ferramentas como o Medline, podemos expandir nossas mentes para encontrar conexões ocultas em quantidades sobre-humanas de conhecimento. Efetivamente, essas ferramentas estão possibilitando um novo método de descoberta científica.

Procurando por Influenza

O método usado por Swanson para descobrir a conexão entre enxaqueca e magnésio é apenas uma das muitas novas maneiras de encontrar significado oculto no conhecimento existente. Uma abordagem diferente foi usada recentemente por cientistas do Google e dos Centros de Controle e Prevenção de Doenças (CDC) dos EUA para desenvolver uma maneira melhor de rastrear a propagação do vírus da gripe. A cada ano, a gripe mata entre 250.000 e 500.000 pessoas em todo o mundo. Governos e organizações de saúde monitoram cuidadosamente a propagação da gripe para que possam responder rapidamente a surtos e prevenir pandemias como a gripe espanhola de 1918, que matou mais de 50 milhões de pessoas. Nos Estados Unidos, a gripe é monitorada pelo CDC, que inscreve médicos em todo o país para participar de um programa de rastreamento. Quando um paciente relata sintomas semelhantes aos da gripe – febre e dor de garganta ou tosse – o médico relata a consulta ao CDC. Apenas uma pequena fração dos médicos participa do programa do CDC, mas o suficiente para permitir que o CDC construa um panorama regional e nacional preciso da gripe. Quando ocorre um surto, o CDC pode se mobilizar, intensificando os programas de vacinação na região e divulgando a informação na mídia. Mas um problema com o sistema é que leva de uma a duas semanas para que os casos de gripe apareçam nos relatórios do CDC.

Esse lapso de tempo é uma preocupação séria, porque os surtos de gripe podem crescer rapidamente em apenas alguns dias.

Na esperança de acelerar o sistema do CDC, os cientistas do Google e do CDC se perguntaram se as consultas de pesquisa inseridas pelos usuários no mecanismo de busca do Google poderiam ser usadas para rastrear instantaneamente onde a gripe está ocorrendo. A ideia é que, se houver um aumento repentino de pessoas na cidade de Atlanta pesquisando por (digamos) "remédio para tosse", é provável que tenha havido um aumento nos casos de gripe em Atlanta. Para obter bons resultados, os cientistas do Google e do CDC usaram os dados históricos de gripe do CDC de 2003 ao início de 2007 e procuraram correlações com consultas de pesquisa comuns do Google. Eles encontraram 45 consultas de pesquisa que estavam especialmente bem correlacionadas com os dados históricos de gripe.
Usando essas consultas, eles construíram um modelo que, esperavam, poderia ser usado para descobrir instantaneamente onde a gripe está ocorrendo, apenas monitorando as buscas no Google. Em seguida, testaram esse modelo comparando-o com um novo conjunto de dados, os dados do CDC referentes à temporada de gripe de 2007-2008. O modelo apresentou concordância quase perfeita (97%)! Em outras palavras, as consultas de pesquisa do Google podem ser usadas para determinar onde os surtos de gripe estão ocorrendo e qual a sua extensão, mas sem o atraso sofrido pelo CDC. Além disso, as consultas de pesquisa do Google podem ser usadas para rastrear a gripe não apenas nos Estados Unidos, mas em qualquer lugar onde um grande número de pessoas esteja usando o Google.

Google, incluindo locais onde não há uma organização como o CDC rastreando doenças. O Google criou um site chamado Google Flu Trends, que usa consultas de pesquisa para rastrear a gripe em 29 países.

Os resultados do Google Flu Trends exigem algumas ressalvas. Primeiro, muitos médicos nos Estados Unidos agora usam sistemas eletrônicos de registro médico, e o CDC firmou recentemente uma parceria com os fabricantes de um desses sistemas, a General Electric, para desenvolver um novo sistema de rastreamento que deve permitir o acompanhamento quase em tempo real de relatos de gripe de 14 milhões de pacientes. É possível, e talvez provável, que o novo sistema do CDC torne o Google Flu Trends obsoleto, pelo menos nos Estados Unidos.

Em segundo lugar, os dados do CDC usados para construir o sistema Google-CDC não rastreavam, a rigor, a gripe. Em vez disso, rastreavam doenças "semelhantes à gripe" a partir de relatos de sintomas como tosse e dor de garganta, frequentemente associados à gripe. Outras condições, como resfriados, podem produzir sintomas semelhantes. Um estudo de acompanhamento realizado em 2010 confirmou que, sem surpresa, o Google Flu Trends é significativamente melhor no rastreamento de doenças semelhantes à gripe do que no rastreamento de casos reais de gripe confirmados em laboratório. É uma ferramenta de diagnóstico útil, não uma maneira perfeita de rastrear a gripe.

Usar o Google para prever a gripe é interessante, mas ainda mais interessantes são as outras possibilidades que ele sugere, possibilidades que vão além da medicina e se estendem a todos os aspectos da vida. Pesquisas subsequentes já mostraram que consultas de busca podem ser usadas para prever tendências de desemprego e preços de imóveis, e até mesmo para prever o desempenho de músicas nas paradas musicais. O que mais seria possível? O Google poderia descobrir quais consultas de busca preveem mudanças no preço das ações de alguma empresa, digamos, a Microsoft? E quanto ao comportamento do Índice Industrial Dow Jones? Ou qual startup de tecnologia é o melhor alvo para aquisição? Ou o resultado da próxima eleição presidencial nos EUA? Ou um golpe de estado em um país instável? Suponha que o Google estivesse rastreando as buscas de assistentes jurídicos que trabalham na Suprema Corte dos EUA — seria possível prever decisões judiciais? Ou talvez descobrir quais preocupações um juiz tem enquanto um caso está sendo ouvido? Suponha que um usuário do Google esteja fazendo buscas que sugerem que ele está planejando um assalto a banco. O Google deveria notificar as autoridades policiais? Em uma coletiva de imprensa em Abu Dhabi, em 2010, o CEO do Google, Eric Schmidt, disse: “Um dia, tivemos uma conversa em que imaginamos que poderíamos simplesmente tentar prever o mercado de ações. E então decidimos que era ilegal. Então, paramos de fazer isso.” É difícil saber se devemos ficar tranquilos ou horrorizados. É claro que não é só o Google que está em posição de fazer esse tipo de análise de dados.

Mineração. Muitas outras organizações — bancos, empresas de cartão de crédito e sites populares como Facebook e Twitter — têm acesso a fontes de dados que podem ser usadas para entender e até mesmo prever o comportamento humano. Se você tem acesso aos dados e aos meios para entendê-los, dados são poder.

Encontrando significado em todo o conhecimento do mundo

Durante quase toda a história registrada, nós, seres humanos, vivemos nossas vidas isolados dentro de minúsculos casulos de informação. Os mais brilhantes e bem-informados de nossos ancestrais frequentemente tinham acesso direto a apenas uma pequena fração do conhecimento humano. Então, nas décadas de 1990 e 2000, em um período de apenas duas décadas, nosso acesso direto ao conhecimento se expandiu talvez mil vezes. Ao mesmo tempo, uma segunda expansão, ainda mais importante, vem ocorrendo: uma expansão em nossa capacidade de encontrar significado em nosso conhecimento coletivo. Vemos essa expansão no uso do Medline por Swanson para encontrar conexões ocultas em nosso conhecimento médico coletivo, ou na maneira como o Google e o CDC combinaram o conhecimento existente (mas inadequado) do CDC sobre gripes relatadas com os dados de pesquisa do Google, para descobrir uma maneira melhor de rastrear a gripe. Também vemos exemplos em nossa vida cotidiana, como a capacidade do Google de responder às nossas perguntas, encontrando a página da web, a notícia, o artigo científico ou o livro certo. Ferramentas como Google e Medline redefinem nossa relação com o conhecimento, oferecendo-nos maneiras de encontrar significados anteriormente ocultos, todos os "conhecimentos desconhecidos" implícitos no conhecimento humano existente, mas que ainda não são apreendidos devido à escala massiva desse conhecimento. Anteriormente neste livro, vimos como a inteligência coletiva pode ser amplificada pela reestruturação da atenção especializada, a fim de aproveitar melhor a expertise disponível. Neste capítulo, discutiremos uma abordagem complementar para amplificar a inteligência coletiva: construir ferramentas que executem tarefas cognitivas diretamente, operando sobre o próprio conhecimento, buscando significado e conexões

O restante deste capítulo está dividido em duas partes. A primeira parte conta a história de um projeto de astronomia chamado Sloan Digital Sky Survey (SDSS). O SDSS está mapeando o universo, assim como os primeiros cartógrafos mapeavam a Terra, usando um telescópio robótico para explorar o céu de forma ampla.

Até agora, tirando imagens de 930.000 galáxias. Essas imagens não são apenas belas imagens; elas estão sendo mineradas por astrônomos para responder a todos os tipos de perguntas sobre o nosso universo. Aprenderemos como o SDSS foi usado para encontrar a maior estrutura conhecida no universo, uma cadeia gigante de galáxias com 1,37 bilhão de anos-luz de comprimento; para descobrir novas galáxias anãs perto da nossa galáxia, a Via Láctea; e para encontrar um par de buracos negros em órbita. Mas, embora essas descobertas sejam fascinantes por si só, há uma razão mais profunda para o nosso interesse no SDSS. Isso porque, embora o acesso ao conhecimento humano tenha se expandido enormemente nas últimas duas décadas, uma grande quantidade de conhecimento científico ainda ***não é*** acessível ao público, e há uma luta para torná-lo mais acessível. E assim, a primeira parte do capítulo conta a história da expansão dos recursos comuns de informação na ciência, usando o SDSS como um exemplo concreto para entender tanto os benefícios quanto os desafios dessa expansão. Essa compreensão concreta nos prepara para a segunda parte do capítulo, onde ampliamos nosso foco para pensar no panorama geral. Quais são as implicações de tornar todo o conhecimento mundial disponível abertamente? E que novos métodos de descoberta isso possibilitará?

Explorando o Universo Digital

A maior estrutura conhecida no universo é uma cadeia de galáxias chamada Grande Muralha de Sloan. Ela tem 1,37 bilhão de anos-luz de comprimento, contém milhares de galáxias e está a cerca de 1 bilhão de anos-luz de distância da Terra. É tão distante que essas galáxias são tênues demais para serem vistas a olho nu, mas se você pudesse vê-las, a Grande Muralha de Sloan se estenderia por quase um terço do céu, desde a constelação de Virgem, passando por Leão, até Câncer. É uma bela visão imaginar todas essas galáxias brilhando no céu noturno!

A Grande Muralha de Sloan foi descoberta em 2003, quando uma equipe de oito cientistas, liderada por J. Richard Gott III, da Universidade de Princeton, decidiu criar um mapa visual de todo o universo conhecido. Parece grandioso, mas eles o fizeram pelo mesmo motivo que criamos mapas de cidades e países: exibir nosso conhecimento visualmente pode facilitar a compreensão do que sabemos. Imagine como a geografia seria difícil se não tivéssemos mapas e tivéssemos que nos basear inteiramente em descrições verbais. Problemas

Problemas fáceis de resolver visualmente, como descobrir quantos continentes existem, de repente se tornariam problemas de pesquisa complexos. Imagina-se que os primeiros geógrafos realizassem conferências de pesquisa sobre "Resolução do Número de Massas Terrestres Continentais", talvez com argumentos acirrados sobre questões como se a Ásia e a América do Norte são continentes realmente separados.

Uma grande dificuldade em fazer um mapa do universo é saber o que está lá fora. Os telescópios modernos nos permitem ver trilhões de objetos, mas a maioria dos astrônomos se concentra em observar apenas uma pequena fração desses objetos. Isso pode parecer surpreendente, mas imagine que você é um astrônomo: você não preferiria gastar seu tempo observando algo que você já sabe ser extremamente interessante, como o buraco negro supermassivo no núcleo da Via Láctea, em vez de alguma estrela aleatória em alguma galáxia aleatória? A maioria dos astrônomos, portanto, passa a maior parte do tempo observando objetos já conhecidos por serem interessantes. É como a diferença entre explorar uma cidade amplamente para encontrar novos lugares interessantes e a tentação de revisitar apenas lugares familiares. Para encontrar novos objetos interessantes no céu, alguém precisa se aventurar e explorar o céu amplamente.

É aqui que entram os levantamentos celestes. Em vez de observar objetos conhecidos em detalhes exaustivos, os telescópios usados em levantamentos celestes varrem sistematicamente todo o céu, construindo uma imagem ampla do universo. Os levantamentos celestes são a base da astronomia, frequentemente nos dando as primeiras pistas sobre quais objetos observar com mais profundidade. Um dos primeiros levantamentos celestes foi o ***Almagesto,*** escrito pelo astrônomo Ptolomeu de Alexandria no século II d.C. Ptolomeu não tinha um telescópio, mas usava o olho nu para compilar todo tipo de informação útil sobre o que via no céu, desde uma descrição de como os planetas se movem até um catálogo detalhado de 1.022 estrelas. O ***Almagesto*** permaneceu como a obra padrão da astronomia na Europa e no Oriente Médio pelos 800 anos seguintes.

Como você talvez já tenha adivinhado, o ***Almagesto*** moderno é o Sloan Digital Sky Survey (SDSS), em homenagem à Fundação Alfred P. Sloan, que fornece grande parte do financiamento. O SDSS realiza seu trabalho usando um telescópio soberbo localizado nos arredores da pequena cidade de Sunspot, no alto das Montanhas Sacramento, no Novo México. O telescópio captura a luz usando um grande espelho de 2,5 metros de diâmetro. A excelente localização e o grande espelho permitem que o SDSS obtenha imagens muito boas e possa enxergar até os limites do universo conhecido. As imagens não são tão boas quanto as dos maiores telescópios do mundo, como o enorme

Gran Telescopio Canarias, de 10,4 metros, nas Ilhas Canárias. Mas o telescópio SDSS tem uma grande vantagem sobre a maioria dos telescópios maiores: possui uma lente grande angular especial que lhe permite fotografar rapidamente grandes seções do céu. Em uma única imagem, ele pode capturar uma área oito vezes maior que a da Lua cheia. Em contraste, o Gran Telescopio Canarias só consegue capturar uma área equivalente a um dezesseis avos do tamanho da Lua, tornando-o inadequado para a ampla exploração exigida por um levantamento do céu. Desde o início de sua operação em 2000, o SDSS já pesquisou mais de um quarto do céu, capturando imagens de 930.000 galáxias ao longo do caminho. E, como veremos, essas imagens têm sido usadas em milhares de outros projetos científicos, incluindo o projeto de Gott e colaboradores para fazer um mapa do universo.

Como se faz um mapa do universo? É um problema surpreendentemente complicado. Idealmente, gostaríamos que um mapa mostrasse tanto objetos relativamente próximos em termos astronômicos, como as estrelas próximas, que estão a apenas alguns anos-luz de distância, quanto as galáxias mais distantes, que estão a bilhões de anos-luz de distância. É difícil fazer as duas coisas no mesmo mapa. O problema da cartografia é ainda mais complicado pelo fato de o universo ser tridimensional, enquanto os mapas comuns são bidimensionais. É claro que existem muitas maneiras de tentar resolver essas complicações, mas isso leva a outro problema: das muitas maneiras de fazer seu mapa, qual é a melhor? Uma característica que é notavelmente óbvia em uma maneira de visualizar o universo pode ser quase invisível em outra. E se você fizer a escolha errada? Mapear a superfície da Terra é um problema muito mais fácil, mas os primeiros cartógrafos ainda tentavam muitas projeções diferentes para dar sentido à Terra.

Da mesma forma, Gott e seus colaboradores experimentaram diversas maneiras de elaborar seu mapa. Um dos mapas que eles criaram utilizou os dados de galáxias do SDSS para visualizar a distribuição de galáxias no universo. Esse mapa é mostrado na Figura 6.1. Não é um mapa comum como um roteiro, e por isso exige um pouco de esforço para ser compreendido, mas vale a pena ler a legenda em detalhes para entender o que está sendo mostrado.
O ponto-chave é a concentração de galáxias no canto superior esquerdo do mapa, uma concentração muito mais densa do que no restante do mapa. Foi o primeiro vislumbre da Grande Muralha de Sloan que a humanidade teve.

A Grande Muralha de Sloan é apenas uma das milhares de descobertas científicas feitas com o SDSS. Para dar uma ideia melhor do impacto do SDSS, deixe-me descrever brevemente mais duas dessas descobertas. Você talvez já saiba que nossa galáxia, a Via Láctea, tem duas galáxias vizinhas, a Grande e a Pequena Nuvem de Magalhães. Estas são anãs.

galáxias, com a maior das duas contendo cerca de 30 bilhões de estrelas, em comparação com as centenas de bilhões da nossa Via Láctea. Se você nunca esteve no hemisfério sul, talvez nunca tenha visto as Nuvens de Magalhães, pois elas estão muito ao sul no céu para serem visíveis de grande parte do hemisfério norte. Mas elas são visíveis em uma noite escura no hemisfério sul, onde aparecem como manchas no céu.

De acordo com nossa melhor compreensão atual da formação de galáxias, a Via Láctea deveria ter dezenas ou centenas de galáxias anãs próximas. Mas, antes do SDSS, apenas algumas galáxias anãs, além das Nuvens de Magalhães, haviam sido descobertas, e era um quebra-cabeça onde todas as outras anãs desaparecidas estavam. Quando as imagens do SDSS ficaram disponíveis, vários astrônomos pesquisaram as imagens em busca de mais galáxias anãs. Eles não fizeram isso manualmente — levaria muito tempo para examinar todas as imagens. Em vez disso, eles usaram algoritmos de computador para procurar novas galáxias anãs nas imagens do SDSS. O que eles encontraram até agora foram nove novas galáxias anãs perto da Via Láctea, contribuindo em grande parte para a solução do quebra-cabeça das anãs desaparecidas.

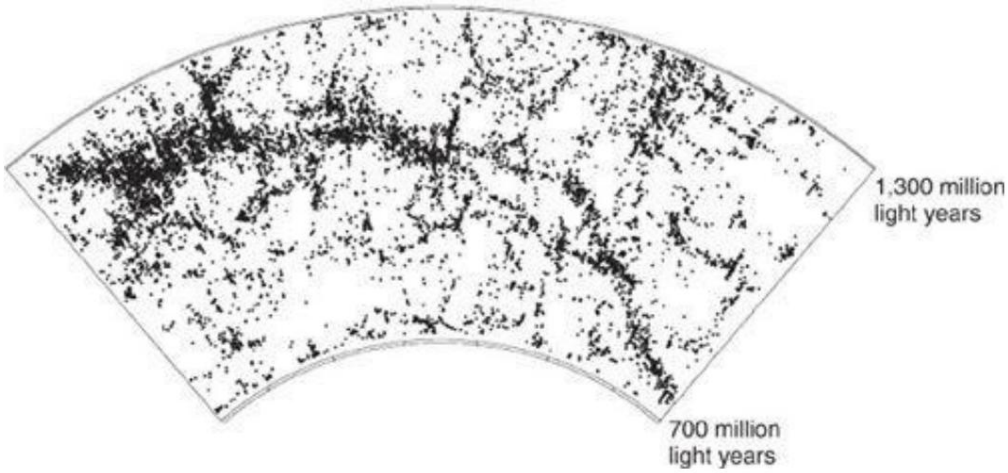


Figura 6.1. Uma parte ampliada de um dos mapas do universo feitos por Gott e colaboradores. Você notará que o mapa se assemelha a um pedaço de torta. Imagine-se na Terra, bem no centro da torta, olhando para o universo. Cada ponto no mapa representa uma única galáxia do SDSS. A direção radial indica a distância até a galáxia, com as galáxias mais próximas no gráfico a cerca de 700 milhões de anos-luz de distância e as mais distantes a cerca de 1,3 bilhão de anos-luz (conforme marcado no lado direito). Todas as galáxias mostradas no gráfico estão muito próximas do equador celeste, o grande arco que atravessa o céu, diretamente acima do equador real da Terra, e circunavegando a Terra. O que você está vendo aqui, então, são as galáxias em uma fina fatia do céu, todas muito próximas do equador celeste. A direção angular no

O gráfico mostra onde a galáxia está localizada ao longo do equador celeste. Os eixos do lado esquerdo do mapa estão na direção da constelação de Virgem, as galáxias do meio estão na direção de Leão e as galáxias da direita estão na direção de Câncer. A densa cadeia de galáxias concentrada no canto superior esquerdo é a Grande Muralha de Sloan. Crédito: Reproduzido com permissão da Sociedade Astronômica Americana.

Como outro exemplo de descoberta possibilitada pelo SDSS, em 2009 os astrônomos Todd Boroson e Tod Lauer usaram o SDSS para descobrir dois buracos negros orbitando um ao redor do outro. A maneira como Boroson e Lauer encontraram os buracos negros pareados foi — você não ficará surpreso! — usando um computador para pesquisar imagens de galáxias no SDSS. Buracos negros não têm cor e não aparecem diretamente em fotos. Mas buracos negros são cercados por enormes quantidades de matéria brilhante que está caindo, e então, em certo sentido, é possível "ver" buracos negros nas imagens de galáxias, dizer que eles têm uma cor, e assim por diante. A chave para o trabalho de Boroson e Lauer foi uma suposição inteligente que eles fizeram, que foi que, se dois buracos negros estivessem orbitando um ao redor do outro, eles pareceriam ter cores ligeiramente diferentes. A razão pela qual eles fizeram essa suposição é interessante. Quando objetos estão se movendo a uma velocidade alta o suficiente — precisa ser uma fração considerável da velocidade da luz — sua cor aparente muda consideravelmente.

Por que isso acontece é uma longa história, na qual não entraremos em detalhes, mas, como exemplo, um objeto vermelho que se move muito rapidamente em direção à Terra parece, na verdade, um pouco mais azul. Boroson e Lauer concluíram que dois buracos negros orbitando um ao outro teriam velocidades diferentes em relação à Terra e, portanto, um seria ligeiramente mais azul que o outro.

Munidos dessa ideia de dupla coloração, Boroson e Lauer usaram seu computador para caçar nos dados do SDSS. Sua esperança se concretizou quando encontraram uma galáxia a quatro bilhões de anos-luz de distância com exatamente a assinatura de dupla coloração que procuravam. Eles prosseguiram com um exame mais detalhado da galáxia, confirmando a presença dos buracos negros em órbita e revelando que ambos são incrivelmente grandes, 20 milhões e 800 milhões de vezes a massa do Sol, respectivamente, e separados por um terço de ano-luz, orbitando um ao outro aproximadamente uma vez a cada 100 anos. A descoberta despertou grande interesse e também desencadeou um debate, com outros astrônomos se perguntando se poderia haver alguma outra explicação para o que Boroson e Lauer estão observando. No momento, a teoria do buraco negro em órbita continua sendo a principal candidata entre várias explicações possíveis. Mas não importa qual seja a verdade, não

duvida-se que Boroson e Lauer tenham descoberto algo notável.

Todas essas descobertas são impressionantes, mas não transmitem completamente o enorme impacto que o SDSS teve na astronomia. Uma maneira de compreender esse impacto é observar quantas vezes os resultados do SDSS foram citados (ou seja, referenciados) em outros artigos científicos. A maioria dos artigos em astronomia é citada apenas algumas vezes, se é que o são. Um artigo citado dezenas de vezes é bastante bem-sucedido, enquanto um artigo citado centenas de vezes é famoso ou está em vias de se tornar famoso. O artigo original do SDSS foi citado em outros artigos mais de 3.000 vezes. Isso é mais citações do que muitos cientistas altamente bem-sucedidos recebem ao longo de toda a sua carreira. Para lhe dar uma ideia da conquista que isso representa, Stephen Hawking, provavelmente o cientista mais famoso do mundo, tem apenas um artigo com mais de 3.000 citações. O artigo de Hawking, publicado em 1975, contava, na verdade, com pouco mais de 4.000 citações em 2011. Em contraste, o artigo do SDSS foi publicado em 2000 e já conta com mais de 3.000 citações. Em breve, ele alcançará e ultrapassará o artigo de Hawking. Vários artigos subsequentes, descrevendo outros aspectos do SDSS, também receberam mais de 1.000 citações. Quando comparei o SDSS ao ***Almagesto*** de Ptolomeu , não estava brincando. O SDSS é um dos empreendimentos mais bem-sucedidos de toda a história da astronomia, digno de um lugar ao lado do trabalho de Ptolomeu, Galileu, Newton e outros grandes nomes de todos os tempos.

Dados Abertos

O que tornou o SDSS um sucesso tão grande? Já discutimos alguns dos motivos: o SDSS possui um excelente telescópio e ampla cobertura do céu. Mas esses não podem ser os únicos motivos. Nas décadas de 1940 e 1950, astrônomos usaram o gigantesco telescópio Palomar de 5 metros, nos arredores de San Diego, Califórnia, para realizar o Palomar Observatory Sky Survey. Mas, embora o telescópio Palomar seja, em alguns aspectos, ainda melhor do que o telescópio SDSS, o levantamento Palomar teve um impacto muito menos dramático na astronomia. Por que isso acontece? O principal motivo é que o levantamento Palomar produziu chapas fotográficas volumosas, que são caras para mover e duplicar e, portanto, só podiam ser acessadas por algumas pessoas.

pessoas. Em contraste, o SDSS tem usado a internet para compartilhar seus dados com toda a comunidade mundial de astrônomos. Desde 2001, o SDSS realizou sete grandes lançamentos de dados, disponibilizando suas imagens (e outros dados) na web, onde qualquer pessoa pode baixá-los. Se desejar, você pode acessar agora mesmo o SkyServer online do SDSS e baixar imagens impressionantes de galáxias distantes. Qualquer pessoa pode fazer isso, e o site foi projetado para ser usado não apenas por astrônomos profissionais, mas também pelo público em geral. As ferramentas do site variam de passeios pelas mais belas vistas do céu à capacidade de enviar consultas sofisticadas ao banco de dados que retornarão imagens com características específicas desejadas. O site ainda contém tutoriais que explicam como fazer coisas como encontrar asteroides ou regiões de formação estelar em outras galáxias.

Esse compartilhamento aberto de dados pelo SDSS parece uma pequena inovação quando comparado às abordagens radicais de inteligência coletiva que vimos em exemplos como o Projeto Polímata e Kasparov versus o Mundo. Mas o impacto do compartilhamento aberto de dados pelo SDSS é enorme. Isso significa que pessoas como Todd Boroson e Tod Lauer — pessoas que não são membros da colaboração do SDSS — podem vir e fazer perguntas fundamentais que ninguém jamais havia pensado em fazer antes: "Podemos usar os dados do SDSS para procurar pares de buracos negros em órbita?" Na ciência, as descobertas são frequentemente limitadas pelos limites do nosso conhecimento. Mas experimentos como o SDSS produzem uma riqueza de conhecimento tão extraordinária — mais de 70 terabytes de dados, muito além da capacidade de compreensão de qualquer ser humano — que eles confundem essa expectativa. Confrontados com tamanha riqueza de dados, em muitos aspectos não somos tão limitados pelo conhecimento quanto somos limitados pelas perguntas.

Somos limitados pela nossa capacidade de formular as perguntas mais engenhosas, ultrajantes e criativas. Ao disponibilizar seus dados para o mundo inteiro, o SDSS permitiu que pessoas como Boroson e Lauer fizessem tais perguntas, perguntas que talvez nunca tivessem sido feitas se o acesso aos dados fosse limitado. É o mesmo que vimos na descoberta de Swanson da conexão entre enxaqueca e magnésio: Swanson não utilizou fatos que já não fossem conhecidos, mas, ao formular uma nova pergunta sobre o conhecimento existente, fez uma descoberta valiosa. É uma variação da serendipidade planejada que vimos na parte 1. Em vez de transmitir uma pergunta ao mundo e esperar por uma resposta, projetos como o SDSS transmitem dados ao mundo, na crença de que as pessoas farão perguntas inesperadas que levarão a novas descobertas.

O compartilhamento de dados do SDSS não é importante apenas pelas descobertas que possibilita. É também importante porque o compartilhamento de dados neste

Essa forma, por mais simples e óbvia que pareça, na verdade é um passo radicalmente ousado para os cientistas envolvidos. A maioria dos cientistas guarda seus dados com zelo. Seus dados são o registro bruto de observações experimentais e podem levar a novas descobertas importantes. É sua vantagem especial sobre seus colegas e concorrentes. Por mais incomum que seja para eles revelarem abertamente seus dados, é ainda mais incomum que incentivem seus colegas a fazer análises independentes e, talvez, descobertas independentes. Você pode entender um pouco do que está em jogo observando alguns casos famosos em que dados foram parcialmente revelados. Por exemplo, mencionei anteriormente ***o Almagesto de Ptolomeu,*** uma das grandes obras científicas da antiguidade. Mas talvez eu devesse ter colocado "Ptolomeu" entre aspas, porque muitos historiadores da ciência — não todos, mas muitos — acreditam que Ptolomeu plagiou muitas das posições das estrelas em seu catálogo do astrônomo Hiparco, que havia feito seu próprio levantamento do céu quase 300 anos antes. De fato, a história da ciência está repleta de exemplos de cientistas roubando dados uns dos outros.

Nos primórdios da ciência moderna, o astrônomo Johannes Kepler descobriu que os planetas se movem em elipses ao redor do Sol usando dados que roubou de seu falecido mentor, o astrônomo Tycho Brahe. James Watson e Francis Crick descobriram a estrutura do DNA com a ajuda de dados que tomaram emprestados de uma das principais cristalógrafas do mundo, Rosalind Franklin. Digo emprestados porque isso foi feito sem o conhecimento dela, embora com a ajuda de um colega de Franklin que, sem dúvida, estava em seus direitos. Esses são, reconhecidamente, exemplos extremos, mas mostram por que a maioria dos cientistas se dá ao trabalho de manter seus dados em segredo.

Então surge um enigma aqui: por que o SDSS compartilha dados tão abertamente? Pense na situação do ponto de vista dos membros da colaboração do SDSS. É quase certo que existam descobertas importantes que eles poderiam ter feito, mas que foram superadas por alguém de fora da colaboração que utilizou dados do SDSS. Em termos puramente egoístas, embora dados abertos possam ser bons para a ciência, podem ser considerados ruins para as carreiras dos membros da colaboração do SDSS. Por que eles defendem isso? Por que o SDSS não mantém os dados em segredo?

Na verdade, o SDSS bloqueia parcialmente os dados. Quando o telescópio do SDSS captura imagens, elas não são imediatamente disponibilizadas ao público. Em vez disso, por um breve período — normalmente de alguns meses a pouco mais de um ano — elas ficam disponíveis apenas para membros oficiais da colaboração do SDSS.
Somente após esse período os dados serão disponibilizados gratuitamente para todos no mundo.

Há uma abertura parcial semelhante quanto à participação na colaboração do SDSS. Embora a maioria dos experimentos científicos ainda envolva apenas um pequeno número de participantes, a colaboração do SDSS conta com 25 instituições acadêmicas participantes e inclui também 14 cientistas adicionais que não fazem parte de nenhuma delas. No total, cerca de 200 cientistas são membros oficiais da colaboração, muito mais do que o cientificamente necessário para colocar o SDSS em funcionamento. A página inicial do site para a fase atual do SDSS (estágio III) até incentiva "consultas de partes interessadas para participar da colaboração". A astronomia é uma comunidade pequena, com apenas alguns milhares de astrônomos profissionais no mundo. Como resultado, muitos, talvez a maioria, dos astrônomos profissionais têm um amigo ou colega que faz parte da colaboração do SDSS e com quem podem potencialmente colaborar usando dados do SDSS, mesmo durante o período inicial, quando os dados não estão abertos.

Essas explicações esclarecem o processo que o SDSS usa para compartilhar dados, mas não respondem à nossa pergunta inicial: por que o SDSS torna seus dados parcialmente abertos? Por que não simplesmente bloquear os dados para sempre? E por que a colaboração do SDSS não é deliberadamente mantida o menor possível, para aumentar os benefícios recebidos por membros individuais? Antes de responder a essas perguntas, quero descrever brevemente vários outros exemplos de experimentos que tornam seus dados disponíveis abertamente. Esses exemplos nos ajudarão a entender por que e quando os cientistas tornam seus dados disponíveis abertamente e por que dados abertos são importantes.

Construindo o Commons de Informação Científica

Em setembro de 2009, uma organização chamada Ocean Observatories Initiative começou a construir uma rede de alta velocidade para dados e eletricidade no fundo do Oceano Pacífico. Eles estão estendendo a internet até o fundo do oceano, com o plano final sendo instalar 1.200 quilômetros (750 milhas) de cabos, desde a costa do Oregon até a Colúmbia Britânica. Essa internet subaquática terá um alcance de mais de 100 quilômetros (60 milhas) da costa. Quando estiver concluída, todos os tipos de dispositivos serão conectados à rede, de câmeras a veículos robóticos e sistemas de monitoramento genômico.

Equipamento de sequenciamento. Imagine um vulcão em erupção subaquática e um equipamento de sequenciamento genômico próximo sendo ligado para capturar imagens genéticas de micróbios nunca antes vistos, liberados durante a erupção. Ou imagine uma rede de termômetros e outros sensores mapeando o clima subaquático, da mesma forma que o SDSS mapeia o universo. Mas a Iniciativa de Observatórios Oceânicos vai ainda mais longe do que o SDSS, disponibilizando seus dados abertamente desde o início, para que qualquer pessoa no mundo possa baixá-los imediatamente, buscando novos padrões e fazendo novas perguntas. Que novas descobertas serão feitas com esse conhecimento sem precedentes do fundo do oceano?

Não são apenas os oceanos e o universo que estão sendo mapeados. Esforços estão em andamento para construir um mapa do cérebro humano. Por exemplo, cientistas do Instituto Allen de Ciência do Cérebro, sediado em Seattle, estão construindo o Atlas Cerebral Allen, mapeando o cérebro próximo ao nível de células individuais e determinando quais genes são ativados em quais regiões do cérebro. É um passo importante no caminho para a compreensão de como os genes formam uma mente e tem o potencial de ser tremendamente útil para entender como nossas mentes funcionam. Cientistas do Instituto Allen fatiaram 15 cérebros em centenas de milhares de fatias, cada fatia com apenas alguns micrômetros de espessura. Eles então analisam cada fatia, determinando quais genes são ativados e onde. Tudo isso é feito por uma equipe de cinco robôs que trabalham 24 horas por dia, cada robô analisando 192 fatias cerebrais por dia, todos os dias. Os cientistas do Instituto Allen esperam concluir seu mapa do cérebro até 2012, quando os resultados serão disponibilizados como dados abertos, para qualquer pessoa no mundo baixar e analisar.

Um esforço anterior do Instituto Allen, concluído em 2007, já nos forneceu um mapa de acesso aberto de como os genes são expressos no cérebro de camundongos. Além disso, este trabalho do Instituto Allen faz parte de um movimento maior na neurociência, em direção a um objetivo ainda mais ambicioso: mapear todo o conectoma humano — a posição de cada neurônio, cada dendrito, cada axônio e cada sinapse no cérebro. É possível que, um dia, em um futuro não muito distante, tenhamos um modelo detalhado e de acesso público de todo o cérebro humano.

O que vemos em exemplos como o SDSS, a Iniciativa de Observatórios Oceânicos e o Atlas Cerebral Allen é o surgimento de um novo padrão de descoberta. O SDSS está mapeando todo o universo. A Iniciativa de Observatórios Oceânicos fará observações abrangentes do fundo do oceano. O Atlas Cerebral Allen está mapeando o cérebro humano. Outros projetos visam construir mapas detalhados da atmosfera terrestre, da superfície terrestre, do clima terrestre, da linguagem humana e da

Composição genética de todas as espécies. Para praticamente qualquer fenômeno complexo na natureza, é provável que haja um projeto em andamento para mapeá-lo em detalhes. Em muitos casos, não se trata de um único projeto, mas de uma série de projetos que fornecem conhecimento cada vez mais detalhado. Vimos isso com a genética humana, onde o Projeto Genoma Humano mapeou o modelo genético humano básico; foi seguido pelo mapa de haplótipos, que mapeou as variações na genética humana; hoje, projetos subsequentes estão obtendo informações ainda mais detalhadas sobre as variações genéticas em grupos humanos específicos. Na astronomia, o SDSS em breve será sucedido pelo Large Synoptic Survey Telescope (LSST), que realizará um levantamento superior ao SDSS em quase todos os aspectos. O LSST, que está sendo construído nos Andes do Chile, será um dos maiores telescópios do mundo, com um diâmetro de espelho efetivo de 6,68 metros, muito maior que o espelho do SDSS, produzindo, portanto, imagens muito melhores. O telescópio terá um campo de visão tão amplo que mapeará todo o céu visível a cada quatro dias, em vez de levar anos para mapear uma fração do céu. Novamente, todos os dados serão imediatamente disponibilizados online.

Em conjunto, esses e outros projetos semelhantes estão mapeando nosso mundo com detalhes incríveis e sem precedentes. É claro que projetos de levantamento semelhantes foram realizados ao longo de toda a história da ciência, desde o ***Almagesto*** até os grandes botânicos dos séculos XVIII e XIX. Mas o que está acontecendo hoje é especial e sem precedentes. A internet expandiu drasticamente nossa capacidade de compartilhar e extrair significado dos modelos que estamos construindo. Isso causou um aumento correspondente em seu impacto científico, como o SDSS ilustra vividamente. O resultado é uma explosão no número e na ambição desses esforços, dando início a uma grande era de descobertas, muito semelhante à era dos exploradores dos séculos XV a XVIII. Mas enquanto esses exploradores foram até os limites da geografia da Terra, os novos descobridores estão explorando e mapeando os limites do nosso mundo científico.

À medida que mais dados são compartilhados online, a relação tradicional entre fazer observações e analisar dados está mudando. Historicamente, observação e análise têm sido interligadas: a pessoa que realizou o experimento também foi a pessoa que analisou os dados. Mas hoje está se tornando cada vez mais comum que as análises mais valiosas sejam feitas por pessoas de fora do laboratório original. Em algumas áreas da ciência, a divisão do trabalho está mudando, com algumas pessoas se especializando na construção do aparato experimental e na coleta de dados, enquanto outras

especializar-se na análise dos dados desses experimentos. Na biologia, por exemplo, surgiu uma nova geração de biólogos, o bioinformata, cuja principal habilidade não é cultivar culturas de células ou as outras habilidades tradicionais do laboratório de biologia, mas sim combinar as habilidades de programador de computador e biólogo para analisar dados biológicos existentes. De forma semelhante, a química viu o surgimento da quimioinformática, e a astronomia, o surgimento da astroinformática. Essas são disciplinas em que a ênfase principal não está em realizar novos experimentos, mas sim em encontrar um novo significado nos dados existentes.

Por que os dados estão se tornando abertos?

Voltemos ao enigma de por que e quando os cientistas disponibilizam seus dados abertamente. Uma pista vem do tamanho dos experimentos. O SDSS, a Iniciativa de Observatórios Oceânicos e o Atlas Cerebral Allen custaram (ou custarão) dezenas ou centenas de milhões de dólares e envolvem centenas ou milhares de pessoas. Nossos exemplos anteriores de dados abertos, como o Projeto Genoma Humano e o mapa de haplótipos, também foram projetos enormes. Mas a maioria dos experimentos científicos é bem menor. E nos experimentos menores, dados abertos são a exceção, não a regra.
Antes de me interessar por dados abertos, trabalhei por 13 anos como físico. Nesse período, vi centenas de experimentos, quase todos pequenos experimentos, realizados em laboratórios modestos. Até onde sei, nenhum desses experimentos fez qualquer esforço sistemático para tornar seus dados abertos. Vimos algo semelhante no capítulo inicial, na relutância inicial dos cientistas em compartilhar dados genéticos em bancos de dados online como o GenBank. Isso só mudou devido a grandes esforços cooperativos, como o Acordo das Bermudas sobre o compartilhamento de dados genéticos humanos. Na ciência, a situação atual está mudando, com alguns periódicos científicos e agências de fomento promulgando políticas que incentivam ou exigem que os dados sejam disponibilizados abertamente após a publicação dos experimentos. Mas dados abertos continuam sendo a exceção, não a regra. Se você for à sua universidade local e entrar em um pequeno laboratório, provavelmente descobrirá que os dados são mantidos a sete chaves, às vezes literalmente.

Parece, então, que grandes projetos científicos têm maior probabilidade de tornar seus dados abertos do que projetos menores. Por que isso acontece? Parte da

A explicação é política. Pense no SDSS. Um pequeno projeto típico de astronomia pode custar "apenas" algumas dezenas ou centenas de milhares de dólares. É muito dinheiro, mas é uma pequena quantia em comparação com os bilhões de dólares que nossa sociedade gasta em astronomia. Se as pessoas que realizam o experimento guardam os dados para si mesmas, não é uma grande perda para outros astrônomos. Além disso, esses outros astrônomos não estão em posição de reclamar, pois eles também estão mantendo os dados de seus experimentos em segredo. É uma situação estável e pouco cooperativa. Mas o tamanho do SDSS o torna especial e diferente. É tão grande que consome grande parte de todo o orçamento mundial para astronomia. Se os dados são mantidos em segredo, para astrônomos fora da colaboração do SDSS é como se toda aquela quantia de dinheiro simplesmente tivesse desaparecido do orçamento para astronomia.

Eles têm todos os motivos para insistir que os dados sejam abertos. E, portanto, se grandes projetos não se comprometerem com uma abertura, pelo menos parcial, seus pedidos de financiamento correm o risco de serem rejeitados por pessoas da mesma área, mas de fora da colaboração. Isso motiva grandes projetos científicos a tornarem seus dados, pelo menos parcialmente, abertos.

Há outro fator que inibe os dados científicos abertos: mesmo que você esteja disposto a compartilhar seus dados, pode ser difícil fazê-lo de uma forma que seja útil para os outros. Você pode tirar todas as fotos de galáxias que quiser e compartilhá-las com outras pessoas, mas essas fotos têm uso científico limitado sem todos os tipos de informações adicionais. Quais filtros de cor você usou? A imagem foi processada de alguma forma, por exemplo, para remover pixels ruins ou danificados? Havia alguma névoa onde as fotos nhe foram tiradas, o que poderia obscurecer a imagem? E assim por diante. Em muitas áreas da ciência, é difícil entender dados experimentais sem informações detalhadas de calibração. E mesmo com os dados ***e*** as informações de calibração, outros cientistas ainda precisam de uma compreensão extremamente detalhada do experimento para fazer uso dos dados. Adicione a isso problemas como garantir que todos estejam usando a terminologia técnica exatamente da mesma maneira, conversão de formato de arquivo e assim por diante. Individualmente, todos esses são problemas solucionáveis, mas juntos são um obstáculo formidável para o compartilhamento de dados de uma forma útil.

Essas questões sobre o compartilhamento de dados fazem parte de uma história mais profunda, uma história sobre por que e quando o conhecimento científico é compartilhado. Anteriormente neste livro, mencionei várias vezes que os cientistas constroem sua reputação e carreira com base nos artigos que escrevem. A reputação de escrever excelentes artigos lhes garantirá um bom emprego científico e apoio financeiro contínuo. Grande parte do desafio do compartilhamento de dados reside no fato de que as recompensas que os cientistas recebem por compartilhar seus dados são muito mais incertas do que as recompensas por escreve

artigos. É verdade que algumas grandes colaborações, como o SDSS, ganharam amplo reconhecimento por compartilhar dados. Mas, em muitas áreas da ciência, existem poucas normas estabelecidas sobre como e quando o uso de dados de outra pessoa deve ser reconhecido. E isso significa que compartilhar dados é arriscado para um cientista. Simplesmente não é algo pelo qual os cientistas são normalmente bem recompensados, apesar de ser extremamente valioso. E, portanto, dados abertos continuam incomuns, especialmente em laboratórios menores. Retornaremos à questão de como entusiasmar os cientistas com o compartilhamento de dados (e outras questões relacionadas) nos capítulos 8 e 9. Para os propósitos do restante deste capítulo, basta que já exista uma quantidade considerável (e crescente) de dados científicos disponíveis abertamente, por meio de projetos como o SDSS e o Projeto Genoma Humano.

Sonhando com a Web de Dados

Até agora neste capítulo, adotamos uma perspectiva concreta e de curto prazo, analisando projetos existentes como o SDSS. Mas a internet é uma plataforma infinitamente flexível e extensível para manipular o conhecimento humano, com um potencial ilimitado. Para compreender esse potencial, precisamos expandir nosso pensamento e adotar uma visão de longo prazo que veja a internet não como uma revolução de dez ou vinte anos, mas como uma revolução de cem ou mil anos. Precisamos imaginar um mundo onde a construção de um espaço comum de informação científica tenha se concretizado. Este é um mundo onde todo o conhecimento científico foi disponibilizado online e é expresso de uma forma que pode ser compreendida por computadores. Imagine, além disso, que os dados não estejam isolados em pequenas ilhas de conhecimento, como estão hoje, com descrições separadas e isoladas de fenômenos fundamentalmente conectados na natureza, fenômenos como aminoácidos, genes, proteínas, medicamentos e registros médicos humanos. Em vez disso, teremos uma rede interligada de dados que conectará todas as partes do conhecimento. Em vez de minerar esse conhecimento de forma fragmentada, seremos capazes de realizar inferências automatizadas sobre todo o conhecimento humano, encontrando conexões ocultas em uma escala que ofusca o trabalho de Swanson ou mesmo do SDSS. Daremos um nome a esse sonho: o chamaremos de sonho da rede de dados.

A rede de dados parece grandiosa. Mas, como vimos, já demos muitos pequenos passos em direção a ela, por meio de projetos como o SDSS e o Projeto Genoma Humano. O que está emergindo gradualmente é uma rede de conhecimento online que se destina a ser lida por máquinas, não por humanos. Essas máquinas encontrarão significado nessa rede de conhecimento e nos ajudarão a explicá-lo. No restante deste capítulo, perguntaremos como a rede de dados será construída e o que isso significará.

Há, no entanto, uma dificuldade na discussão, uma dificuldade que atormenta qualquer discussão sobre o potencial dos computadores para encontrar significado no conhecimento: quanto mais se especula sobre esse potencial, mais se avança na direção de uma discussão sobre inteligência artificial em plena expansão, o cenário de ficção científica do tipo "a internet acorda para dominar o mundo". É muito divertido falar sobre isso, mas é muito fácil se atolar em perguntas especulativas: "Então, as máquinas podem se tornar conscientes? E o que é consciência, afinal?" ou "Bem, sim, talvez um dia a internet acorde e assuma o controle, e daí?". Tudo isso já foi trilhado muitas vezes antes. Em vez de repetir essas discussões, exploraremos um meio-termo entre os projetos de curto prazo discutidos anteriormente neste capítulo e a inteligência artificial em plena expansão. Este futuro intermediário é conceitualmente rico, fascinante e estranhamente pouco discutido, talvez porque os sonhos da inteligência artificial exercem uma forte atração sobre a imaginação dos curiosos em tecnologia. O que faremos é sintetizar as ideias atuais da ciência da computação para entender o que acontece quando pegamos os algoritmos atuais e imaginamos um futuro em que eles possam ser aplicados a todo o conhecimento científico. Como veremos, os resultados prováveis são espetaculares.

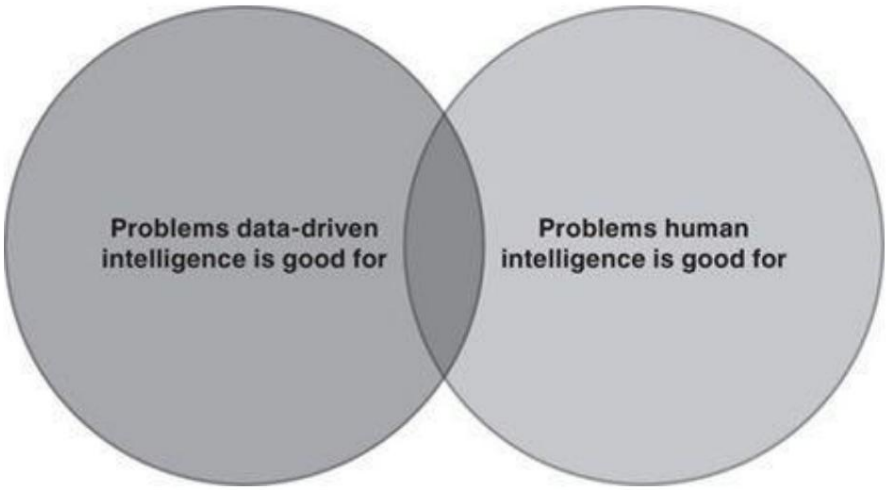
Inteligência orientada por dados

Para entender o que a rede de dados pode fazer, é útil dar um nome à capacidade dos computadores de extrair significado dos dados. Chamarei essa capacidade de ***inteligência orientada por dados.*** Exemplos de inteligência orientada por dados incluem os algoritmos usados nas buscas no Medline que Don Swanson fez para descobrir a conexão entre enxaqueca e magnésio, os algoritmos usados para correlacionar as buscas do Google com os dados do CDC sobre gripe e os algoritmos usados

explorar o SDSS em busca de galáxias anãs e buracos negros em órbita, e descobrir a Grande Muralha de Sloan.

O termo “inteligência orientada por dados” não é novo. Mas, atualmente, é usado principalmente em um sentido mais restrito do que o que estou propondo, para descrever abordagens orientadas por dados para a tomada de decisões corporativas — por exemplo, a maneira como as companhias aéreas coletam dados sobre o não comparecimento de passageiros para saber quanto overbooking devem fazer em seus voos. Proponho usar o termo de uma forma muito mais geral, como uma categoria ampla de inteligência, semelhante à forma como usamos termos como “inteligência humana” e “inteligência artificial”. Nesse sentido geral, “inteligência orientada por dados” é um termo muito necessário, em parte devido ao grande e crescente número de exemplos de inteligência orientada por dados. Mas o que é ainda mais importante é que o termo destaca uma abordagem específica para encontrar significado, uma abordagem para a qual os computadores são extremamente adequados e que é diferente e complementar à maneira como nós, humanos, encontramos significado.

É claro que um chauvinista humano poderia se opor ao meu uso do termo "inteligência" em "inteligência orientada por dados", argumentando que não há nada de muito inteligente em um computador pesquisando dez milhões de artigos científicos ou pesquisando galáxias anãs no SDSS. É apenas trabalho rotineiro e mecânico, embora em uma escala muito além da capacidade humana. Mas aqui está o ponto: esses são problemas que nós, humanos, não conseguimos resolver. Quando se trata de extrair significado de terabytes ou petabytes (milhares de terabytes) de dados, não somos muito melhores do que qualquer outro animal. Temos, na melhor das hipóteses, algumas habilidades muito especializadas nesse domínio, como a capacidade de processar imagens visuais, e praticamente nenhuma capacidade geral de processamento de dados em larga escala. Então, quem somos nós para julgar computadores nesse domínio? A capacidade de um humano sem ajuda de processar grandes conjuntos de dados é comparável à capacidade de um cachorro de fazer cálculos aritméticos, e não muito mais valiosa. Portanto, embora esses problemas talvez não exijam que os computadores sejam muito inteligentes, neste domínio eles são muito mais inteligentes do que os humanos. Esse ponto de vista é capturado no diagrama mostrado nesta página.



É da natureza humana focar nos problemas à direita do diagrama, os problemas onde a habilidade e a engenhosidade humanas são mais valiosas. E é preconceito humano normal subestimar os problemas à esquerda, o domínio onde a inteligência orientada por dados realmente brilha. Mas vamos deixar esse preconceito de lado e pensar nos problemas à esquerda. Quais problemas os computadores podem resolver que nós não conseguimos? E como, quando juntamos essa capacidade à inteligência humana, podemos combinar as duas para fazer mais do que cada uma delas é capaz de fazer sozinha?

Como exemplo deste último, em 2005, o site de xadrez Playchess.com organizou o que chamou de torneio de xadrez estilo livre, ou seja, um torneio em que humanos e computadores podiam participar juntos como equipes híbridas. Em outras palavras, o torneio permitiu que a inteligência humana se unisse à inteligência orientada por dados, na forma de computadores que jogam xadrez, que se baseiam em enormes bancos de dados de aberturas e finais de jogo e que analisam inúmeras combinações possíveis de movimentos no meio do jogo. Um dos participantes do torneio foi a equipe por trás da série Hydra de computadores de xadrez. O Hydra, na época o computador de xadrez mais forte do mundo, nunca havia perdido uma partida em partidas regulares para nenhum jogador de xadrez humano e, em diversas ocasiões, derrotou facilmente os principais grandes mestres.

A equipe Hydra inscreveu dois de seus computadores, um jogando inteiramente sozinho e o outro com alguma assistência humana. Também se inscreveram no torneio várias equipes que emparelharam grandes mestres com computadores de xadrez fortes. Sozinhos, nem os grandes mestres nem seus computadores conseguiram igualar os Hydras. Mas as equipes conjuntas humano-computador derrotaram os Hydras. Não só nenhum dos Hydras venceu o torneio, como também nenhum deles chegou às quartas de final. Os grandes mestres conseguiram derrotar os Hydras porque sabiam quando confiar em seus computadores e quando confiar em seu próprio julgamento. Ainda mais interessante, o vencedor do torneio foi uma equipe chamada ZackS, composta por dois jogadores de baixo nível.

jogadores amadores classificados usando três computadores prontos para uso e software padrão para jogar xadrez. Eles não apenas superaram os Hydras, como também superaram os grandes mestres com seus potentes computadores para jogar xadrez.
Os operadores humanos do ZackS demonstraram uma habilidade extraordinária ao usar a inteligência baseada em dados de seus algoritmos computacionais para ampliar sua habilidade no xadrez. Como um dos observadores do torneio, Garry Kasparov, comentou posteriormente: "Um humano + máquina + processo melhor, fraco, foi superior a um computador forte sozinho e, ainda mais notavelmente, superior a um humano + máquina + processo inferior, forte."

A inteligência orientada por dados tem objetivos mais amplos do que a inteligência artificial. Em grande parte, a inteligência artificial assume tarefas nas quais os seres humanos são bons e visa imitar ou melhorar o desempenho humano. Pense em programas de computador para jogar jogos humanos como damas, xadrez e go, ou esforços para treinar computadores a entender a fala humana. A inteligência orientada por dados pode ser aplicada a essas tarefas tradicionalmente humanas — ela pode entender a fala humana ou jogar xadrez —, mas onde ela realmente se destaca é na resolução de diferentes tipos de problemas, problemas que envolvem habilidades complementares à inteligência humana, problemas como as buscas de Swanson na literatura de pesquisa médica ou a mineração de dados do SDSS por Boroson e Lauer em busca de pares de buracos negros em órbita. Uma inteligência orientada por dados completa seria capaz de jogar damas, xadrez ou go, mas não os jogaria por diversão. Ela jogaria jogos com um escopo cuja complexidade estaria inteiramente além da compreensão humana.

O termo "inteligência" é frequentemente usado para significar algum tipo de capacidade intelectual generalizada. A inteligência orientada por dados é mais direcionada por natureza, com diferentes tipos de inteligência orientada por dados usados para resolver diferentes tipos de problemas. Veremos um exemplo explícito na próxima seção, que analisa os algoritmos que os biólogos usam para sequenciar genomas. Um conjunto bastante diferente de algoritmos é usado para realizar buscas em serviços como o Medline. Para cada problema, um tipo diferente de inteligência orientada por dados é necessário. Uma consequência é que a inteligência orientada por dados em algum domínio de problema pode começar bastante estúpida, mas gradualmente se torna mais inteligente à medida que desenvolvemos métodos aprimorados. Por exemplo, quando Swanson fez seu trabalho sobre enxaqueca e magnésio, ferramentas de busca como o Medline usavam ideias relativamente simples. Os mecanismos de busca de hoje usam ideias muito mais sofisticadas, e os mecanismos de busca de amanhã, sem dúvida, serão muito melhores ainda. De fato, assim como a inteligência orientada por dados ajuda empresas como o Google a obter lucro, essas empresas investem dinheiro no desenvolvimento de técnicas ainda melhores, resultando em um círculo virtuoso de aprimoramento.

Como a inteligência orientada por dados se relaciona com a inteligência coletiva? Na verdade, essa não é exatamente a pergunta certa para a nossa discussão. Estamos interessados na inteligência orientada por dados como forma de ampliar nossa própria inteligência, e, portanto, uma pergunta mais adequada seria: como a inteligência orientada por dados se relaciona com as ferramentas que estudamos na Parte 1, as ferramentas que ***amplificam*** a inteligência coletiva? Como vimos, essas ferramentas funcionam reestruturando a atenção dos especialistas para que ela seja alocada de forma mais eficaz. Portanto, não há relação direta entre ferramentas que amplificam a inteligência coletiva e a inteligência orientada por dados. Mas as duas podem ser usadas de forma comp Por exemplo, vimos como ferramentas baseadas em dados, como o Medline, oferecem novas maneiras de encontrar significados ocultos no conhecimento coletivo de grandes grupos de pessoas, como a comunidade biomédica. E ferramentas baseadas em dados, como o Google, podem ser usadas para ampliar nossa inteligência coletiva, ajudando-nos a encontrar as informações e as pessoas às quais devemos prestar atenção. Por outro lado, o Google usa nossa inteligência coletiva para construir seu serviço, minerando conteúdo na web e usando a estrutura de links da web para descobrir quais páginas são mais importantes. Portanto, embora a inteligência baseada em dados e a inteligência coletiva sejam diferentes, elas podem ser usadas para se reforçar mutuamente.

Este não é um livro didático sobre inteligência orientada por dados, e não descreverei as centenas de algoritmos inteligentes em uso ou em desenvolvimento. Para nós, a inteligência orientada por dados é fundamentalmente importante como um conceito que unifica exemplos como o Google Flu Trends, a Grande Muralha de Sloan e a descoberta de Swanson sobre a enxaqueca e o magnésio. Subjacentes a todos esses exemplos estão algoritmos inteligentes que extraem significado de dados que, de outra forma, estariam além da compreensão humana. A inteligência orientada por dados é, em certo sentido, complementar à rede de dados: embora a inteligência orientada por dados possa ser aplicada a qualquer fonte de dados, ela atinge seu potencial máximo quando aplicada às fontes de dados mais ricas possíveis, e a rede de dados é a fonte de dados mais rica que podemos imaginar. A inteligência orientada por dados é o que nos permitirá pegar todo o conhecimento do mundo e dar significado a ele.

Inteligência baseada em dados em biologia

Para tornar a inteligência orientada por dados mais concreta, deixe-me descrever em detalhes um exemplo da biologia de como ela funciona. O exemplo mostra como podemos usar algoritmos inteligentes e o acervo de informações científicas para fazer algo notável: encontrar o genoma de um ser humano. Para entender o exemplo, primeiro precisamos relembrar um pouco sobre genética. Como você sabe, dentro de cada célula do seu corpo existem muitas fitas da molécula de DNA. Essas fitas de DNA carregam informações, e as informações que elas carregam são o design para você.

Para entender como o DNA carrega essa informação, lembre-se da estrutura de dupla hélice do DNA. As hélices são bonitas e memoráveis, mas a informação não é armazenada nas hélices, por si só. Em vez disso, ela é armazenada entre as hélices. A cada poucos nanômetros, conforme você sobe na dupla hélice, há um par de moléculas unindo os dois lados da hélice, chamado de par de bases. É um par de pequenas minimoléculas especiais que se ligam umas às outras e às estruturas da dupla hélice. Existem quatro tipos de moléculas de base, chamadas adenina, citosina, guanina e timina. Seus nomes geralmente são abreviados para A, C, G e T. O A se liga ao T e o C ao G, então os pares possíveis são AT e CG. Você pode, portanto, descrever a informação em uma única fita de DNA por meio de uma longa sequência de letras — digamos, CGTCAAGG. — representando as bases ligadas a um lado da hélice (o outro lado terá bases complementares, GCAGTTCC...). Essa sequência é uma descrição da sua arquitetura básica. Como exatamente ela especifica essa arquitetura ainda é apenas parcialmente compreendido, mas tudo o que sabemos sugere que a sequência de pares de bases do DNA é o modelo para o nosso projeto.

Como descobrimos a sequência de DNA de uma pessoa? Na verdade, se começarmos com um fragmento de DNA com apenas algumas centenas de pares de bases, ele pode ser sequenciado diretamente usando a química tradicional e inteligente — essencialmente, um cientista, em seu laboratório, misturando cuidadosamente os produtos químicos. Mas se a fita de DNA for muito maior do que isso, o problema do sequenciamento se torna mais complexo. Uma fita típica de DNA humano contém várias centenas de milhões de pares de bases, longa demais para ser sequenciada diretamente. Mas existe uma maneira inteligente de combinar o sequenciamento direto de fitas curtas de DNA com inteligência baseada em dados para descobrir a sequência completa do DNA.

Para entender como funciona, imagine que eu lhe dei um exemplar do primeiro livro de Harry Potter, ***Harry Potter e a Pedra Filosofal.*** Mas, em vez de lhe dar um exemplar comum, peguei uma tesoura e cortei o livro em pequenos fragmentos. Por exemplo, a abertura do livro poderia ser cortada nestes fragmentos:

“Sr. e Sra. Dursley, do número quatro, Priv”;

“et Drive, temos orgulho em dizer que eles nós”;

“está perfeitamente normal, muito obrigado.”

E assim por diante. Simplifiquei um pouco as coisas aqui, mostrando os fragmentos na mesma ordem em que aparecem no início do livro. Mas quero que você imagine que os entreguei na ordem errada, todos embaralhados. Ao mesmo tempo, imagine que eu também lhe dei um segundo exemplar do livro, também dividido em pequenos fragmentos, mas de uma forma diferente:

“Sr. e Sra. Dursley, de num”;

“ber quatro, Privet Drive, estávamos orgulhosos de”;

“dizer que eles eram perfeitamente normais”.

Embora os fragmentos nos dois casos sejam diferentes, há bastante sobreposição, e você pode usar essas sobreposições para descobrir quais fragmentos combinam. Observe, por exemplo, que o fragmento "Sr. e Sra. Dursley, do número quatro, Priv" se sobrepõe a "Sr. e Sra. Dursley, do número quatro, Priv". Dursley, do número "e" número quatro, Rua dos Alfeneiros, orgulhavam-se de". Isto sugere colar os dois últimos fragmentos juntos, para obter "Sr. e Sra. Dursley, do número quatro da Rua dos Alfeneiros, orgulhavam-se de fazê-lo. Continuando com muito cuidado dessa maneira, você poderia reconstruir sequências bastante longas do livro. Você só ficaria preso se, por acaso, a sobreposição entre dois fragmentos fosse tão curta que tornasse difícil dizer que eles realmente eram fragmentos sobrepostos do mesmo texto. Mas se eu lhe desse uma terceira (e uma quarta...) cópia do livro cortada aleatoriamente dessa maneira, as chances de todas as sobreposições serem curtas em qualquer ponto cairiam drasticamente, e você poderia muito bem ser capaz de reconstruir o livro inteiro.

O sequenciamento do genoma humano (e de outras formas de vida complexas) funciona de maneira semelhante. Embora não possamos sequenciar diretamente longas fitas de DNA, podemos fazer muitas cópias dessas fitas, cortá-las em locais aleatórios e sequenciar diretamente os fragmentos. Tudo isso pode ser feito usando a química tradicional, um cientista em seu laboratório, etc. Em seguida, usamos nossos computadores para descobrir onde os diferentes fragmentos se sobrepõem.

e juntar tudo de novo. (Aliás, ignorei algumas sutilezas, como a repetição de certas sequências de DNA ao longo do genoma humano, o que dificulta a remontagem da sequência completa de DNA. Essas sutilezas podem ser abordadas com outros truques, mas agora você entendeu a ideia geral.)

Agora, imagine que queremos sequenciar o DNA de alguém hoje.
Talvez seja para um teste de paternidade. Ou talvez seja parte de uma investigação criminal. Não importa qual seja o motivo. Acontece que podemos simplificar o procedimento acima para sequenciamento de DNA, usando os fatos de que (1) um genoma humano de referência já é conhecido e (2) graças ao mapa de haplótipos, sabemos onde no genoma as pessoas podem diferir e onde, ao que parece, somos sempre os mesmos. Para entender como o processo simplificado funciona, imagine agora que você possui uma cópia ***completa*** de ***Harry Potter e a Pedra Filosofal.*** Então, você recebe uma cópia recortada de um livro semelhante, mas que foi modificado em alguns locais. Na verdade, na vida real, o livro realmente foi alterado entre seu lançamento inicial no Reino Unido e seu lançamento nos Estados Unidos.

Uma mudança em especial se destaca: a palavra ***"filósofo"*** no título foi alterada para ***"feiticeiro",*** de modo que o título se tornou ***Harry Potter e a Pedra*** Filosofal . Ao longo de todo o livro, "filósofo" foi substituído por "feiticeiro" — presumivelmente, a editora acreditava que o livro teria maior apelo nos Estados Unidos dessa forma. É bastante óbvio que ter o texto completo do livro original para consultar tornaria muito mais fácil decifrar o texto do livro modificado. Em vez de ter que descobrir laboriosamente quais fragmentos combinavam com quais, você sempre poderia descobrir de qual parte do livro o fragmento que você está examinando é. De forma semelhante, o sequenciamento de um genoma humano pode ser feito de forma mais rápida e fácil consultando constantemente o genoma de referência e o mapa de haplótipos.

A propósito, embora o exemplo de Harry Potter seja fantasioso, não posso deixar de mencionar que uma técnica muito semelhante ***foi*** usada pelo autor Chuck Hansen para escrever seu livro " ***US Nuclear Weapons: The Secret History" (Armas Nucleares dos EUA: A História Secreta).*** Hansen baseou sua história em dezenas de milhares de documentos desclassificados que haviam sido sanitizados por meio do corte físico de informações sigilosas. Ele descobriu que cópias diferentes do mesmo documento às vezes eram sanitizadas de maneiras diferentes e, comparando versões diferentes, ele às vezes conseguia reconstruir as informações excluídas!

Os algoritmos que descrevi para sequenciamento genômico são bons exemplos de inteligência orientada por dados. Em nenhum sentido esses algoritmos são especialmente inteligentes. Eles não fazem muito além de simples padrões.

Correspondência e rearranjo. Mas, combinando esses algoritmos simples com um enorme poder de processamento de dados, podemos resolver um problema que um ser humano sem ajuda não consegue resolver. Além disso, combinando a inteligência orientada por dados com os dados abertos do genoma humano e do HapMap, podemos simplificar o problema do sequenciamento genético. Esse é o tipo de coisa que veremos em uma escala muito maior quando a inteligência orientada por dados for combinada com a rede de dados.

Construindo a Web de Dados

Hoje, a rede de dados está em seus primórdios. A maioria dos dados ainda está bloqueada. À medida que os dados são compartilhados, diversas tecnologias diferentes estão sendo utilizadas para esse compartilhamento. Os conjuntos de dados abertos disponíveis permanecem, em sua maioria, desconectados uns dos outros, ainda vivendo em silos separados
Em suma, o estado atual da rede de dados é confuso, caótico e incompleto. Tudo bem: os primeiros dias de uma nova tecnologia costumam ser confusos. Pense em quão confusa e caótica foi a história inicial da aviação, nas décadas de 1890 e início de 1900, antes dos irmãos Wright voarem pela primeira vez.
Dezenas de pessoas buscavam suas próprias ideias sobre a melhor maneira de construir máquinas voadoras mais pesadas que o ar. Foi dessa confusão de ideias que os primeiros aviões emergiram lentamente. De forma semelhante, hoje milhares de pessoas e organizações têm suas próprias ideias sobre a melhor maneira de construir a rede de dados. Todas apontam aproximadamente na mesma direção, mas há muitas diferenças nos detalhes. Talvez o esforço mais conhecido venha da academia, onde muitos pesquisadores estão desenvolvendo uma abordagem chamada web semântica. No mundo dos negócios, a situação é mais fluida, à medida que as empresas experimentam diversas maneiras de compartilhar dados. Devido a essas diversas abordagens, há discussões acaloradas sobre a melhor maneira de construir a rede de dados, muitas vezes realizadas com grande convicção e certeza. Mas a rede de dados ainda está em seus primórdios e é muito cedo para dizer qual abordagem terá sucesso. Por essas razões, uso o termo "rede de dados" de forma bastante ampla para me referir a ***todos*** os dados abertos, considerados em conjunto. É um pouco exagerado, já que muitos desses dados não estão devidamente vinculados ou são difíceis de encontrar online. Mas essa ligação está chegando, então tomei algumas liberdades.

Se não sabemos qual tecnologia será usada para construir a rede de dados, como podemos ter certeza de que ela crescerá e prosperará? Podemos, porque um número grande e crescente de pessoas deseja compartilhar seus dados e conectá-los a outras fontes. Vimos um pouco de como isso está acontecendo na ciência. Isso também se aplica a muitas empresas e governos, alguns dos quais estão disponibilizando pelo menos alguns dados. O site Twitter, por exemplo, disponibiliza alguns de seus dados abertamente, o que levou à criação de serviços de terceiros, como o TwitPic, que facilita o compartilhamento de fotos no Twitter, e o Tweetdeck, que oferece uma maneira simplificada de usar o Twitter. Outro exemplo: no dia seguinte à posse do presidente dos EUA, Barack Obama, ele emitiu um memorando sobre "Transparência e Governo Aberto". Esse memorando levou à criação de um site chamado data.gov, onde o governo dos EUA compartilha mais de 1.200 conjuntos de dados abertos sobre assuntos que vão do uso de energia a acidentes de aviação e licenças de televisão.

Exemplos como esses estão impulsionando o desenvolvimento de tecnologias para compartilhar dados entre o maior número possível de usuários. Qualquer tecnologia que obtiver ampla adoção se tornará, por padrão, a rede de dados. É por isso que não precisamos saber qual visão tecnológica da rede de dados prevalecerá para concluir que ela é inevitável.

Talvez os passos mais impressionantes em direção à rede de dados até o momento tenham sido dados na biologia. Biólogos estão selecionando pedaços do mundo biológico e mapeando-os, construindo um mapa unificado de toda a biologia. Discutimos alguns desses pedaços — o genoma humano, o mapa de haplótipos e o conectoma humano em formação. Mas há muitos outros. Existem bancos de dados online que descrevem o mundo biológico em um nível muito pequeno, por exemplo, mapeando a estrutura e a função das proteínas e as muitas interações possíveis entre elas (o "interatoma"). Existem bancos de dados online que descrevem o mundo biológico em larga escala, mapeando aspectos como padrões de migração animal e até mesmo catálogos que tentam mapear todas as espécies do mundo. E existem bancos de dados online em todos os níveis intermediários, uma infinidade de recursos para a descrição do mundo biológico. A lista de bancos de dados biológicos da Wikipédia tinha mais de 100 entradas em abril de 2011. Esses bancos de dados podem ser potencialmente vinculados para refletir as conexões em sistemas biológicos: informações genéticas são vinculadas a informações sobre proteínas, que são vinculadas a informações sobre interações entre proteínas, que são vinculadas a informações sobre metabolismo e assim por diante, tudo construindo um mapa unificado da biologia.

Estão sendo desenvolvidos serviços para minerar essa nascente rede de dados biológicos, uma espécie de Google da biologia, capaz de responder rapidamente a perguntas complexas sobre a vida. Imagine um mundo do futuro onde a parte biológica da rede de dados floresceu. Imagine ter o genoma de recém-nascidos imediatamente sequenciado e, em seguida, correlacionado com um gigantesco banco de dados de registros de saúde pública para determinar não apenas a quais doenças eles são especialmente vulneráveis — um velho clichê da ficção científica —, mas também quais fatores ambientais podem influenciar sua suscetibilidade a doenças. "Seu filho tem 80% de chance de desenvolver doenças cardíacas aos 40 anos se for sedentário aos 20 e 30 anos. Mas com três horas de exercício moderado por semana, essa probabilidade cai para 15%." À medida que os problemas se manifestam, medicamentos especiais podem ser criados, com seu design adaptado especificamente à composição genética individual e ao histórico médico anterior.

Hoje, a rede de dados biológicos é apenas um protótipo. A vida tem uma complexidade tremenda em muitos níveis diferentes, e estamos apenas começando a mapear o mundo biológico. Apenas estabelecer as categorias conceituais básicas já é desafiador. Veja a noção de gene. Até recentemente, os alunos aprendiam que um gene é parte do DNA que codifica uma proteína. Isso parece bastante simples. Mas, na verdade, o que os cientistas querem dizer com gene está mudando, à medida que passamos a entender melhor a relação entre DNA e proteínas. A percepção inicial de que os genes codificam proteínas é incompleta. Agora sabemos que a mesma sequência de DNA pode, às vezes, ser transcrita de maneiras diferentes, em proteínas diferentes. Ao mesmo tempo, uma única proteína pode ser formada pela transcrição de DNA de várias partes desconectadas do genoma, às vezes até mesmo de material genético em cromossomos diferentes. Essas são apenas duas das muitas maneiras pelas quais nossa noção de genes está mudando atualmente. De forma mais geral, à medida que nossa compreensão da biologia melhora, muitos conceitos fundamentais estão sendo redefinidos. E quando esse tipo de redefinição acontece, pode ter implicações profundas para a maneira como representamos o conhecimento. É fácil imaginar em algum momento no futuro a necessidade de reestruturar radicalmente nossos bancos de dados de conhecimento, à medida que aprendemos que nossos antigos esquemas conceituais estão errados e precisam ser atualizados.

O que a Web de Dados Significará para a Ciência

À medida que a rede de dados floresce, ela transformará a ciência de duas maneiras. A primeira maneira será aumentar drasticamente o número e a variedade de questões científicas que podemos responder. Já vimos como o SDSS permitiu que milhares de novas questões em astronomia fossem respondidas. Quanto mais fontes de dados estiverem disponíveis e quanto mais ricamente interligadas estiverem, mais drástico será o efeito. Pense na forma como os dados de pesquisa do Google e os dados sobre a gripe do CDC foram combinados. Com qualquer um dos conjuntos de dados isoladamente, é difícil responder à pergunta "Onde está a gripe, neste momento?". Mas quando se tem ambos os conjuntos de dados, é possível responder a essa pergunta. O resultado tem uma qualidade mágica e gratuita: combine dois conjuntos de dados e não só poderá responder a todas as perguntas originalmente respondidas por esses conjuntos de dados, como também poderá responder a novas perguntas surpreendentes que emergem das relações entre eles. À medida que a rede de dados cresce, também cresce o número e a variedade de perguntas que podem ser feitas. Em certo sentido, as perguntas que você pode responder são, na verdade, uma propriedade emergente de sistemas complexos de conhecimento: o número de perguntas que você pode responder cresce muito mais rápido do que o seu conhecimento. E a rede de dados aspira a conter todo o conhecimento do mundo.

A segunda maneira pela qual a rede de dados transformará a ciência é mudando a natureza da própria explicação. Historicamente, na ciência, prezamos explicações simples. Muitas de nossas maiores teorias têm uma qualidade de "coelho saído da cartola", explicando muitos fenômenos aparentemente diferentes por meio de uma única ideia central. Por exemplo, a teoria da evolução por seleção natural de Darwin tem uma ideia simples em seu cerne, mas é uma estrutura surpreendentemente poderosa para a compreensão da evolução da vida. Como outro exemplo, a teoria geral da relatividade de Einstein foi lindamente resumida em uma única frase pelo físico John Wheeler: "O espaço-tempo diz à matéria como se mover; a matéria diz ao espaço-tempo como se curvar". Essa ideia simples, quando expressa matematicamente, explica todos os fenômenos gravitacionais, desde o voo de uma bola arremessada, ao movimento dos planetas, até a origem do universo. É um milagre de explicação, e muitos cientistas (inclusive eu) experimentam uma epifania quando a entendemos pela primeira vez.

Mas alguns fenômenos não têm explicações simples. Pense no problema de traduzir do espanhol para o inglês. Essas línguas contêm uma grande dose de complexidade acidental, resultado de todas as contingências em sua gênese histórica. Para fazer traduções de alta qualidade, não temos escolha a não ser lidar com toda essa complexidade. Na vida cotidiana, os tradutores fazem isso, em parte, por meio de um vasto conhecimento sobre os detalhes da língua.

línguas, e em parte por meio de uma intuição difícil de descrever, construída ao longo de anos de exposição a ambas as línguas. Qualquer explicação realmente precisa sobre como traduzir do espanhol para o inglês será necessariamente bastante complexa e certamente não terá a simplicidade da teoria da evolução ou da teoria geral da relatividade.

Até recentemente, a complexidade das explicações científicas que usamos era limitada pelas limitações de nossas próprias mentes. Hoje, isso está mudando, à medida que aprendemos a usar computadores para construir e, em seguida, trabalhar com modelos extremamente complexos. Para explicar a mudança, deixe-me dar um exemplo do campo da tradução de linguagem de máquina. Por volta de 1950, pesquisadores começaram a construir sistemas computadorizados cujo objetivo era traduzir automaticamente de uma língua para outra. Infelizmente, os primeiros sistemas não eram muito bons. Eles tentaram fazer a tradução usando modelos inteligentes e relativamente simples, baseados nas regras gramaticais e outras regras da linguagem. Isso parece uma boa ideia, mas, apesar de muito esforço, nunca funcionou muito bem. Acontece que as línguas humanas contêm complexidade demais para serem capturadas em regras tão simples.

Na década de 1990, pesquisadores em tradução automática começaram a tentar uma abordagem nova e radicalmente diferente. Eles descartaram as regras convencionais de gramática e linguagem e, em vez disso, começaram seu trabalho reunindo um enorme corpus de textos e traduções — pense, por exemplo, em todos os documentos das Nações Unidas. A ideia deles era usar inteligência orientada por dados para analisar esses documentos ***em massa,*** tentando inferir um modelo de tradução. Por exemplo, ao analisar o corpus, o programa poderia notar que frases em espanhol contendo a palavra "hola" frequentemente contêm a palavra "hello" na tradução para o inglês. A partir disso, o programa estimaria uma alta probabilidade de que a palavra "hola" resulte na palavra "hello" no texto traduzido, enquanto a probabilidade para palavras em inglês não relacionadas a "hola" ("tiger", "couch" e "January", por exemplo) seria muito menor. O programa também examinaria o corpus para descobrir como as palavras se movimentavam na frase, observando, por exemplo, que "hola" e "hello" tendem a estar nas mesmas partes da frase, enquanto outras palavras se movimentam mais. Repetindo isso para cada par de palavras nas línguas espanhola e inglesa, o programa deles gradualmente construiu um modelo estatístico de tradução — um modelo imensamente complexo, mas ainda assim armazenável em um computador moderno. Não descreverei os modelos que eles usaram em detalhes aqui, mas o exemplo do "olá-alô" dá uma ideia. Depois de analisarem o corpus e construírem seu modelo estatístico, eles o usaram para traduzir novos textos. Para traduzir uma frase em espanhol, a ideia era encontrar

a frase em inglês que, de acordo com o modelo, tinha a maior probabilidade. Essa frase de alta probabilidade seria gerada como a tradução.

Francamente, quando ouvi falar de tradução automática estatística pela primeira vez, não achei que soasse muito promissora. Fiquei tão surpreso com a ideia que pensei que devia estar entendendo algo errado. Esses modelos não só não entendem o significado de "hola" ou "hello", como também não entendem as coisas mais básicas da linguagem, como a distinção entre substantivos e verbos. E, ao que parece, meu ceticismo se justifica: a abordagem não funciona muito bem — se o corpus inicial usado para inferir o modelo contiver apenas alguns milhões de palavras. Mas se o corpus tiver bilhões de palavras, a abordagem começa a funcionar muito bem. Hoje, é assim que os melhores sistemas de tradução automática funcionam.

Se você já fez uma pesquisa no Google que retornou um resultado em um idioma estrangeiro, notará que o Google oferece a opção de "traduzir esta página". Essas traduções não são feitas por seres humanos ou por algoritmos especiais criados manualmente com conhecimento detalhado dos idiomas envolvidos.
Em vez disso, o Google utiliza um modelo estatístico incrivelmente detalhado de como traduzir. Está longe de ser perfeito, mas hoje é o melhor sistema de tradução automática disponível. Logo após o lançamento do seu serviço de tradução, o Google venceu facilmente uma competição internacional de tradução automática inglês-árabe e inglês-chinês. O que é realmente notável é que ninguém na equipe do Google Tradutor falava chinês ou árabe. Eles não precisavam. O sistema conseguia traduzir sozinho.

Esses modelos de tradução são, em certo sentido, teorias ou explicações sobre como traduzir. Mas, enquanto a teoria da evolução de Darwin pode ser resumida em algumas frases e a teoria geral da relatividade de Einstein pode ser expressa em uma única equação, essas teorias de tradução são expressas em modelos com bilhões de parâmetros. Você pode objetar que tal modelo estatístico não se parece muito com uma explicação científica convencional, e você estaria certo: não é uma explicação no sentido convencional. Mas talvez devesse ser considerado, em vez disso, como um novo tipo de explicação. Normalmente, julgamos as explicações em parte por sua capacidade de prever novos fenômenos. No caso da tradução, isso significa traduzir com precisão frases nunca antes vistas. E até agora, pelo menos, os modelos estatísticos de tradução fazem um trabalho melhor nisso do que qualquer teoria convencional da linguagem. É revelador que um modelo que nem sequer compreende a distinção substantivo-verbo possa superar nossos melhores modelos linguísticos. No mínimo, deveríamos levar a sério a ideia de que esses modelos estatísticos expressam verdades não encontradas em modelos mais convencionais.

explicações sobre a tradução de línguas. Será que os modelos estatísticos contêm mais verdade do que nossas teorias convencionais da linguagem, com suas noções de verbo, substantivo e adjetivo, sujeitos e objetos, e assim por diante? Ou talvez os modelos contenham um tipo diferente de verdade, em parte complementar e em parte sobreposta às teorias convencionais da linguagem? Talvez pudéssemos desenvolver uma teoria da linguagem melhor combinando os melhores insights da abordagem convencional e da abordagem baseada em modelagem estatística em uma explicação única e unificada?

Infelizmente, ainda não sabemos como criar teorias tão unificadas.
Mas é estimulante especular que substantivos e verbos, sujeitos e objetos, e toda a outra parafernália da linguagem são, na verdade, propriedades emergentes cuja existência pode ser deduzida a partir de modelos estatísticos da linguagem. Hoje, ainda não sabemos como dar esse salto dedutivo, mas isso não significa que não seja possível.

Que status devemos dar a explicações complexas desse tipo? À medida que a rede de dados for construída, será cada vez mais fácil para as pessoas construírem tais explicações, e acabaremos com modelos estatísticos de todos os tipos de fenômenos complexos. Precisaremos aprender a analisar modelos complexos, como os modelos de linguagem, e extrair conceitos emergentes, como verbos e substantivos. E precisaremos aprender a lidar com o fato de que, às vezes, esses conceitos emergentes serão apenas aproximados. Em suma, precisaremos desenvolver mais e melhores ferramentas para extrair significado desses modelos complexos.

Dito isso, ainda parece intuitivo que explicações simples contenham mais verdade do que explicações complexas. Esse preconceito contra explicações complexas na ciência está tão arraigado que até tem um nome: nós o chamamos de navalha de Occam. A ideia é que, se temos duas explicações alternativas para o mesmo fenômeno, devemos preferir a explicação mais simples. Essa crença também se reflete de outras maneiras. Quando chegamos a uma explicação única e simples que explica uma ampla variedade de fenômenos aparentemente díspares, somos inclinados a pensar que ela é verdadeira. Gritamos "Eureka", descobrimos, quando algo que parecia complexo acaba tendo uma explicação simples. Pense na incrível descoberta de Newton de que suas leis da gravitação explicam tanto como os objetos caem na Terra quanto o movimento dos planetas ao redor do Sol. Antes da descoberta de Newton, esses fenômenos pareciam completamente separados um do outro: como é notável que as mesmas leis expliquem ambos!

Nossa confiança na verdade de explicações simples é tão grande que, quando descobrimos violações aparentes de tal explicação, podemos fazer todo o possível para salvá-la. Na década de 1970, a astrônoma Vera Rubin

descobriu que estrelas nas regiões mais externas da nossa galáxia, a Via Láctea, estão girando em torno do centro da galáxia muito mais rápido do que esperaríamos com base na nossa melhor teoria da gravidade, a teoria geral da relatividade. Mas, em vez de abandonar a relatividade geral, a maioria dos astrônomos prefere postular a existência de matéria escura invisível permeando a galáxia. Se a distribuição da matéria escura estiver correta, a relatividade geral pode explicar corretamente a velocidade das estrelas nas bordas externas da galáxia. Em comparação com a popularidade da matéria escura, novas teorias da gravidade têm sido desenvolvidas por relativamente poucos astrônomos.

Até agora, fiz pouca distinção entre explicações convencionais e modelos complexos. Essa confusão despreocupada dos dois talvez tenha incomodado alguns leitores. Muitas pessoas acreditam que existe uma distinção clara e inequívoca entre uma explicação e um modelo: as explicações contêm algum elemento de verdade, enquanto os modelos são meras muletas convenientes, úteis para iluminar algum fenômeno, mas que, em última análise, não expressam a verdade. Esse ponto de vista tem um apelo intuitivo, mas na história da ciência a distinção entre modelos e explicações é tênue a ponto de inexistir. Ideias que começam como "meros" modelos frequentemente contêm a semente de verdades que surpreendem até mesmo seus criadores. Em 1900, o físico Max Planck tentava entender como a cor e a intensidade da luz emitida por um objeto dependem de sua temperatura. Por exemplo, carvões em brasa inicialmente brilham em vermelho, mas, à medida que aquecem, mudam de cor e eventualmente brilham em azul. Descobrir a relação entre temperatura e cor foi um enigma, porque as melhores teorias físicas da época davam duas respostas diferentes, ambas contraditas por experimentos! Planck tentou muitas ideias para resolver o problema, chegando finalmente a um modelo no qual fez a suposição ad hoc de que a energia associada à luz deve estar em pacotes quantizados, ou seja, deve ser um múltiplo de alguma unidade básica. Essa foi uma suposição arbitrária, e o próprio Planck disse mais tarde: "Eu realmente não pensei muito sobre isso" — foi apenas um truque que o levou ao resultado que ele queria.

De fato, descobriu-se que a ideia contida no modelo de Planck foi, em última análise, a semente de uma das grandes descobertas da ciência, a teoria da mecânica quântica. Então, devemos considerar as ideias de Planck meramente como um modelo ou como uma explicação? Na época, parecia um modelo, mas esse modelo continha uma verdade mais profunda do que qualquer uma das teorias da época. Em qualquer contabilidade razoável, as ideias de Planck são tanto um modelo quanto uma explicação: modelos e explicações são ambos parte do mesmo continuum. E assim, à medida que as ferramentas online aumentam nossa capacidade de constru

extrair significado de modelos complexos, eles também mudarão a natureza da explicação científica.

CAPÍTULO 7

Democratizando a Ciência

Em 7 de agosto de 2007, uma professora holandesa de 25 anos chamada Hanny van Arkel estava navegando na internet quando encontrou o site do Galaxy Zoo. Como você deve se lembrar do capítulo de abertura, o Galaxy Zoo recruta voluntários para ajudar a classificar imagens de galáxias. Os voluntários veem fotografias de galáxias — muitas vezes, galáxias que nenhum ser humano jamais viu — e são solicitados a responder perguntas como "Esta é uma galáxia espiral ou elíptica?" ou "Se esta é uma espiral, os braços giram no sentido horário ou anti-horário?". É uma espécie de censo cosmológico, o maior já realizado, com mais de 200.000 voluntários até agora classificando mais de 150 milhões de imagens de galáxias. Quando encontrou o Galaxy Zoo, van Arkel ficou imediatamente fascinada e começou a classificar galáxias em seu tempo livre. Poucos dias depois de entrar, ela notou uma estranha mancha azul pairando logo abaixo de uma das galáxias. O que ela viu está reproduzido na próxima página, em preto e branco, com uma seta apontando para a mancha.

Intrigada, em 13 de agosto ela postou uma nota no Galaxy Zoo online fórum, perguntando se alguém sabia o que poderia ser a mancha azul.

Ninguém sabia.

Testes foram feitos. A mancha azul não era algum tipo de defeito na fotografia, era real. Observações foram feitas em outros telescópios para obter informações mais detalhadas, incluindo observações com o poderoso telescópio William Herschel, nas Ilhas Canárias. Essas observações mostraram que a mancha azul estava aproximadamente à mesma distância da Terra que a galáxia que pairava sobre ela, o que significava que a mancha era enorme, com dezenas de milhares de anos-luz de diâmetro. Mais especialistas foram chamados, nenhum dos quais jamais havia visto algo parecido.

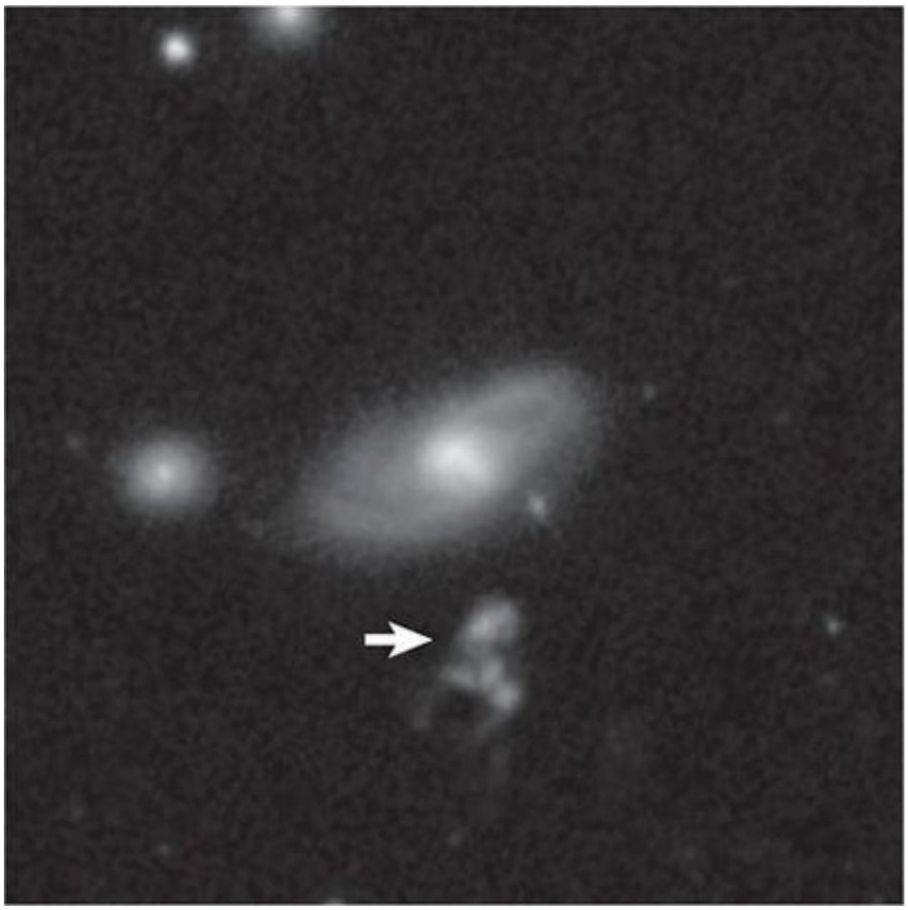


Figura 7.1. Uma reprodução em preto e branco mostrando a estranha mancha observada pela primeira vez por Hanny van Arkel. Na imagem colorida original, a mancha era de um azul impressionante e contrastava intensamente com a galáxia acima. Crédito: Sloan Digital Sky Survey.

O mistério aumentou. Mais e mais pessoas começaram a especular sobre o que a mancha azul poderia ser. O objeto foi apelidado de Hanny's Voorwerp, em homenagem ao descobridor e à palavra holandesa para objeto.

Lentamente, uma explicação para o voorwerp tomou forma, uma explicação que o conectava aos objetos incrivelmente brilhantes conhecidos como quasares. Para entender essa explicação, primeiro precisamos voltar um pouco e falar sobre quasares. Como você deve saber, os quasares estão entre os objetos mais estranhos e misteriosos do universo. Eles são incrivelmente brilhantes: um quasar do tamanho do nosso sistema solar pode brilhar tanto quanto um trilhão de sóis, ofuscando em muito o brilho de uma galáxia gigante como a nossa Via Láctea. Felizmente para nós, os quasares mais próximos estão a centenas de milhões de anos-luz de distância — se um quasar se ativasse a poucos anos-luz de distância, ele fritaria a Terra.

Quando os quasares foram descobertos pela primeira vez, em 1963, era um mistério como objetos tão pequenos conseguiam brilhar tanto. Foi preciso

astrônomos e astrofísicos levaram muitos anos para entender e concordar sobre o que está acontecendo, mas na década de 1980 era amplamente aceito que os quasares são alimentados por buracos negros do tamanho do sistema solar no centro das galáxias. Esses buracos negros devoram a matéria circundante — estrelas, poeira, tudo o que você quiser — enquanto a outra matéria gira em torno do buraco negro, sem chegar a cair, mas acelerada a uma velocidade próxima à da luz. Essa enorme aceleração produz vastas quantidades de energia, parte da qual é emitida como luz. É essa luz que vemos na Terra como o quasar. Mas, embora essa imagem rudimentar dos quasares seja amplamente aceita, muitas questões fundamentais permanecem sem resposta.

Com essa compreensão dos quasares em mente, voltemos ao voorwerp. Enquanto as pessoas no Galaxy Zoo se perguntavam o que o voorwerp poderia ser, consideraram muitas explicações possíveis e gradualmente chegaram a uma explicação simples que parecia se encaixar em todos os fatos: o voorwerp é um espelho de quasar. A ideia é que, há cerca de 100.000 anos, a galáxia próxima ao voorwerp continha um quasar. Esse quasar se desligou desde então, por razões desconhecidas, e não o vemos mais. Mas enquanto o quasar ainda brilhava, a luz do quasar estava aquecendo o gás dentro de uma galáxia anã próxima, fazendo-a brilhar. É esse gás brilhante que agora vemos como uma mancha azul, e é por isso que podemos pensar no voorwerp como um espelho de quasar. Na verdade, é uma enorme coleção de espelhos, distribuídos por uma vasta região do espaço, ecoando a luz do quasar em muitos momentos diferentes de sua história. Claro, estou usando o termo "espelho" de forma imprecisa aqui, já que a luz do voorwerp não é luz refletida, mas sim o brilho de gás aquecido. É uma espécie de eco de luz do quasar.

Nem todos os astrônomos e astrofísicos consideram a explicação do espelho do quasar convincente. Para alguns, parece um pouco conveniente demais que o quasar tenha se desligado. Outro grupo propôs uma explicação alternativa para o voorwerp, envolvendo um tipo diferente de fonte na galáxia próxima, uma fonte que também é alimentada por um buraco negro, mas que não é um quasar. Essa suposta fonte é chamada de núcleo galáctico ativo (AGN).
Trata-se de um buraco negro supermassivo que emite o que chamamos de jato, um cone estreito de plasma com dezenas de milhares de anos-luz de comprimento. Por acaso, o jato está direcionado na direção do voorwerp. O jato aquece o gás no voorwerp e o faz brilhar. Portanto, nesta explicação, o voorwerp não é um espelho de quasar, mas sim um espelho de AGN (novamente, falando de forma mais ampla)!

Enquanto escrevo, astrônomos e astrofísicos ainda estão tentando descobrir qual explicação é a correta. Mas, independentemente de qual explicação seja...

Correto, ou mesmo que seja necessária alguma explicação, o voorwerp é fascinante. Suponha, por exemplo, que seja realmente um espelho de quasar. Como vimos, isso significa que o voorwerp é uma enorme coleção de espelhos, ecoando a luz do quasar em muitos momentos diferentes ao longo de sua vida. Isso significa que a luz do voorwerp é um pouco como uma biografia do quasar, e examinando-o atentamente, podemos aprender muito: como o quasar viveu, como morreu e talvez até como nasceu.
Isso torna o voorwerp tremendamente importante como forma de estudar o ciclo de vida dos quasares. Da mesma forma, se o voorwerp estiver realmente brilhando devido a um jato de um AGN, estudá-lo será uma ótima maneira de aprender mais sobre AGNs. Em ambos os casos, os astrônomos estão entusiasmados com as possibilidades e planejam investigações subsequentes com o objetivo de obter uma imagem mais detalhada do voorwerp. O tempo de observação foi obtido em alguns dos telescópios mais requisitados do mundo, incluindo o Hubble e outros telescópios espaciais. A partir dessas e de outras observações, aprenderemos mais sobre o voorwerp e, talvez, também sobre quasares ou núcleos galácticos ativos. A história do voorwerp está apenas começando.

Redefinindo a relação da ciência com a sociedade

Tomamos como certo que a ciência é, em grande parte, feita por cientistas. Parte do que torna o Voorwerp de Hanny empolgante é que ele viola essa suposição. Que notável que uma professora de 25 anos tenha descoberto esta grande e bela nuvem de gás! Que inesperado que um amador pudesse fazer uma descoberta que pudesse mudar nossa compreensão de quasares ou núcleos galácticos ativos! Quando a descoberta do voorwerp foi anunciada, foi notícia na mídia em todo o mundo, recebendo cobertura na CNN e na BBC, na ***The Economist*** e em muitos outros grandes veículos de comunicação. Embora eu estivesse encantado por Hanny van Arkel e a equipe do Galaxy Zoo, como escritor, meu primeiro sentimento com toda essa publicidade foi de uma certa decepção egoísta, pensando que eu precisaria remover o voorwerp do meu livro e substituí-lo por um exemplo mais atual. Mas, depois de pensar mais, decidi deixar o voorwerp: a própria divulgação na mídia ilustra o quão fortemente tomamos como certo que a ciência é feita por cientistas e o quão fascinados nós

são exceções a esta regra. A manchete da CNN diz tudo: "Astrônomo de poltrona descobre 'fantasma cósmico' único". Que choque e surpresa que um leigo pudesse fazer uma descoberta astrofísica significativa!

O Galaxy Zoo e o voorwerp fazem parte de uma história maior sobre como as ferramentas online estão gradualmente mudando a relação entre ciência e sociedade. Uma das áreas mais férteis onde isso está acontecendo é a ciência cidadã, com projetos como o Galaxy Zoo recrutando voluntários online para ajudar a fazer descobertas científicas. Na primeira metade deste capítulo, examinaremos a ciência cidadã em profundidade, observando como ela muda quem pode ser cientista e como permite que novos tipos de problemas científicos sejam atacados. Mas a ciência cidadã não é a única maneira pela qual as ferramentas online estão mudando a relação entre ciência e sociedade. Na segunda metade do capítulo, examinaremos outras novas instituições de ponte possibilitadas por ferramentas online e consideraremos como tais instituições podem mudar o papel da ciência no debate público e na tomada de decisões. Essa discussão talvez pareça tangencial ao tema principal do livro, uma vez que não se relaciona diretamente com a forma como os cientistas fazem descobertas. Mas, a longo prazo, essas mudanças sociais podem alterar significativamente o contexto em que a ciência é feita, e vale a pena explorá-las com alguma profundidade. Primeiro, vamos voltar a examinar o Galaxy Zoo em mais detalhes.

Galaxy Zoo revisitado

> Posso dizer honestamente que o Galaxy Zoo é a melhor coisa que já fiz.
> · · Não sei bem o que é, mas o Galaxy Zoo mexe com as pessoas. As contribuições, tanto criativas quanto acadêmicas, que as pessoas fizeram para o fórum são tão impressionantes quanto a visão de qualquer espiral, e nunca deixam de me comover.
>
> **—Alice Sheppard, moderadora voluntária do Galaxy Zoo**

O Galaxy Zoo começou em 2007, com dois cientistas da Universidade de Oxford, Kevin Schawinski e Chris Lintott. Como parte de seu trabalho de doutorado, Schawinski observava fotos de galáxias. As galáxias vêm em muitos formatos e tamanhos, mas a maioria das galáxias são espirais, como a nossa Via Láctea, ou então galáxias elípticas, bolas aproximadamente esféricas de estrelas e gás. A sabedoria convencional, em 2007, sustentava que a maioria das estrelas em galáxias elípticas são estrelas muito antigas, chegando a atingir 10 bilhões de anos de idade. Quando as estrelas envelhecem, elas frequentemente mudam de cor e tamanho, transformando-se em gigantes vermelhas, com o resultado de que muitas galáxias elípticas têm uma cor avermelhada.

coloração quando comparada com galáxias espirais, que são mais jovens e contêm muitas estrelas azuis recém-formadas.

Schawinski suspeitava que a sabedoria convencional estava errada, que algumas galáxias elípticas talvez não fossem tão antigas assim, e que poderia haver muita formação estelar acontecendo dentro delas. Para testar sua suspeita, Schawinski passou uma semana examinando fotos de 50.000 galáxias do Sloan Digital Sky Survey (SDSS), procurando ver quais galáxias eram elípticas e quais eram espirais. Como mencionei no capítulo inicial, distinguir galáxias elípticas de espirais é algo que os humanos ainda fazem melhor do que os computadores. Assim que terminou a classificação, Schawinski usou um programa de computador para analisar cada galáxia elíptica, para ver o quão vermelha ou azul ela era. Como ele suspeitava, os resultados sugeriram que a sabedoria convencional estava errada, que a formação estelar ***estava*** acontecendo em algumas elípticas. Infelizmente, o efeito foi fraco, e ele precisou analisar uma amostra muito maior de galáxias para realmente defini-lo. Felizmente, como discutimos no capítulo anterior, o SDSS havia disponibilizado publicamente imagens de 930.000 galáxias. Era um recurso promissor, mas desafiador. A classificação das primeiras 50.000 galáxias envolveu um esforço heroico de uma semana por parte de Schawinski — classificar 50.000 galáxias em sete dias úteis de 12 horas exige a classificação de uma imagem a cada seis segundos! Mesmo nesse ritmo tremendo, levaria muitos meses para classificar 930.000 galáxias. E não havia como Schawinski manter esse ritmo. Mesmo que dedicasse a maior parte do seu tempo de trabalho à classificação, levaria anos.

Em um dia de março de 2007, Schawinski foi ao Royal Oak, um pub em Oxford, acompanhado de Chris Lintott, um cientista de pós-doutorado recém-chegado à universidade. Tomando uma cerveja, eles consideraram uma ideia maluca para classificar as fotos do SDSS. Em vez de fazerem o trabalho de classificação eles mesmos, talvez pudessem criar um site que convidasse o público em geral a ajudar. Eles recrutaram alguns amigos que trabalhavam como desenvolvedores web para ajudar a construir o site e, em 11 de julho de 2007, o site do Galaxy Zoo foi lançado com um anúncio no programa ***Today*** da BBC Radio 4 .

A resposta ao anúncio do Galaxy Zoo superou as expectativas, sobrecarregando e derrubando rapidamente o novo site. Durante as seis horas seguintes, os tratadores que administravam o site trabalharam freneticamente para colocá-lo de volta no ar. Quando o site finalmente reapareceu, os voluntários rapidamente começaram a se inscrever e, ao final do primeiro dia, mais de 70.000 classificações de galáxias eram feitas a cada hora — mais do que Schawinski havia conseguido em sua semana heroica. Cada galáxia foi examinada.

independentemente por muitos voluntários, permitindo que os tratadores do zoológico identificassem e descartassem automaticamente classificações incorretas. Isso tornou os resultados comparáveis à classificação cuidadosa feita por astrônomos profissionais. Embora a taxa de classificação de galáxias tenha diminuído gradualmente de seu pico de 70.000 por hora, a primeira classificação de galáxias do Galaxy Zoo foi concluída em apenas alguns meses. Isso forneceu a Schawinski os dados necessários para concluir seu projeto. Veredito: sim, a sabedoria convencional sobre espirais versus elípticas estava errada, e algumas elípticas realmente contêm muitas estrelas recém-formadas.

O Galaxy Zoo começou com as perguntas de Schawinksi, mas com o tempo o site se expandiu para abordar uma gama muito mais ampla de questões. Muitas descobertas foram feitas por acaso, quando algum Zooíta (como os participantes se autodenominam) notou algo incomum em uma foto, como na descoberta da voorwerp por Hanny van Arkel. Um segundo exemplo, mais complexo, de descoberta por acaso é a história das galáxias "ervilha-verde". Essa história ilustra o potencial da ciência cidadã ainda melhor do que a voorwerp, e por isso a recontarei aqui. A propósito, meu relato se baseia em um artigo maravilhoso escrito por uma das Zooítas, Alice Sheppard, que você pode encontrar referenciado nas "Notas" no final do livro.

Em 28 de julho de 2007, duas semanas após a inauguração do zoológico, um usuário do fórum Galaxy Zoo, chamado Nightblizzard, postou a imagem de uma galáxia verde difusa, observando que era incomum que galáxias fossem verdes. Algumas semanas depois, em 11 de agosto de 2007, outra pessoa postou a imagem de uma estranha galáxia verde. Ela era excepcionalmente brilhante, e o usuário, chamado Pat, perguntou se a galáxia poderia ser um quasar. Ninguém tinha certeza.

No dia seguinte, 12 de agosto, um terceiro usuário, o onipresente Hanny van Arkel, encontrou outra das estranhas galáxias verdes. Van Arkel apelidou a galáxia de "ervilha verde" e a publicou no fórum com uma mensagem intitulada "Dê uma chance às ervilhas!". Outros usuários do Zoo acharam isso hilário e começaram a desenterrar ervilhas, adicionando-as à "sopa de ervilhas" que tomava forma no fórum. Por vários meses, o tópico de discussão cresceu. No início, eram principalmente pessoas adicionando objetos ou fazendo piadas sobre ervilhas ("parem as ervilhas"). Mas as pessoas também faziam perguntas reflexivas. O que exatamente eram as ervilhas? Por que ninguém tinha ouvido falar delas antes? Um usuário comentou: "Eles falam sobre estrelas, galáxias, nebulosas, planetas etc. nos cursos de astronomia, mas nunca mencionam as ervilhas. Deve ser um grande segredo entre astrônomos profissionais. Eles provavelmente querem todas as ervilhas para comerem."

No início, a coleta de ervilhas era apenas um hobby divertido para os zooítas. Mas, à medida que a coleta de ervilhas aumentava, o mistério em torno delas também aumentava. Algumas se revelaram estrelas ou nebulosas comuns. Mas algumas das galáxias verdes ainda se destacavam como incomuns. Os zooítas descobriram — descreverei como em breve — que algumas das galáxias ervilhas eram cercadas por gás oxigênio ionizado incrivelmente quente. Isso era incomum para uma galáxia. O que eram essas pequenas galáxias verdes e altamente luminosas, cercadas por oxigênio ionizado e quente? E por que ninguém nunca tinha ouvido falar delas antes?

Permitam-me fazer uma pausa aqui para explicar como os zoóítas descobriram que as ervilhas estavam rodeadas por oxigênio quente e ionizado. É uma ciência interessante e ilustra o quão sérios alguns zoóítas estavam se tornando. Obviamente, eles não conseguiam determinar a presença de oxigênio visitando uma das galáxias. Em vez disso, descobriram isso aprendendo uma técnica chamada análise espectral. Não precisamos entrar em detalhes sobre como a análise espectral funciona, mas a ideia básica é bem simples. Ela se baseia no que chamamos de espectro de uma galáxia. O que o espectro mostra é como a luz de uma galáxia se divide em cores diferentes — digamos, um pouco de vermelho, muito verde e um toque de azul.

Na verdade, o espectro pode até mostrar (por exemplo) que a luz é uma mistura de vários tons de verde ligeiramente diferentes, exatamente quais tons são esses e suas respectivas proporções. Portanto, o espectro é uma maneira muito detalhada e precisa de decompor imagens de galáxias em suas diferentes cores.

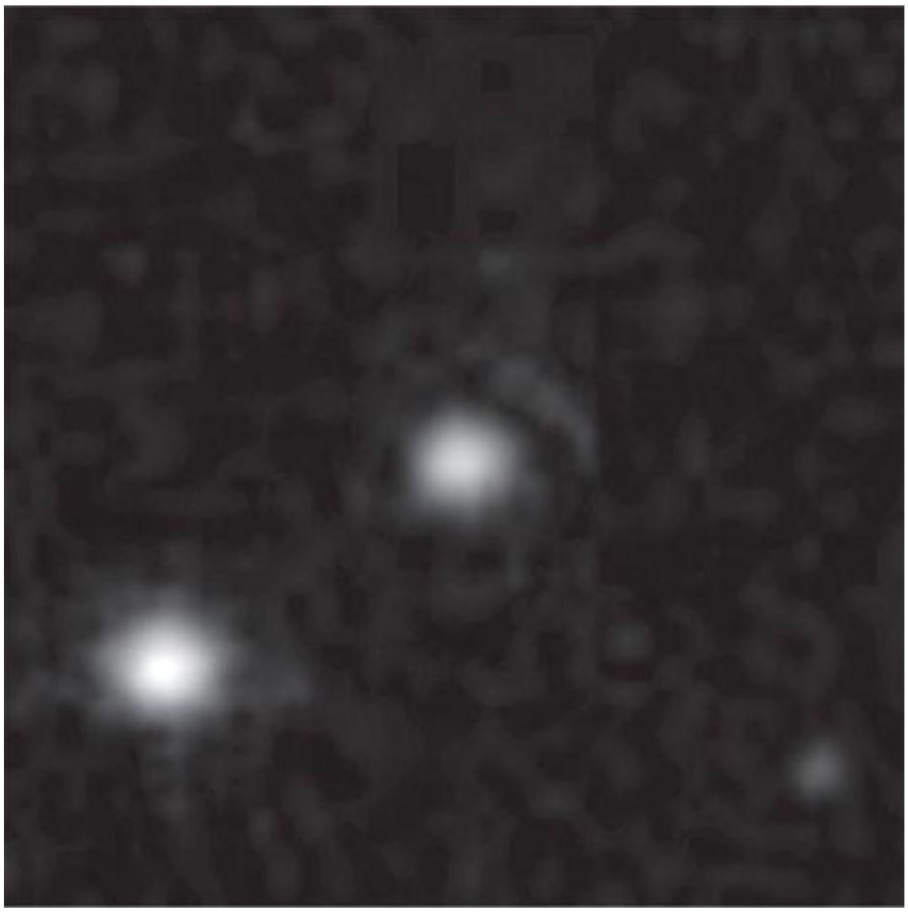


Figura 7.2. A primeira das galáxias ervilha-verde, descoberta por Nightblizzard, membro do fórum Galaxy Zoo, em julho de 2007. A ervilha-verde está no centro. Como todas as ervilhas, ela parece bastante indefinida e, se você não conhece galáxias, é tentador pensar que se trata apenas de mais uma galáxia elíptica, ou talvez uma estrela. Mas muitos dos zoólogos se tornaram especialistas em analisar imagens de galáxias, e não demorou muito para que percebessem que as ervilhas eram incomuns. Crédito: Sloan Digital Sky Survey.

A razão pela qual o espectro de uma galáxia é importante é porque ele permite aos astrônomos descobrir do que a galáxia é feita. Isso pode parecer surpreendente, mas, novamente, a ideia é bem simples: quando você aquece um material, digamos, sódio, ele tende a brilhar com uma mistura particular de cores.
É por isso que os postes de luz de sódio brilham com uma cor amarelo-alaranjada muito particular. Acontece que cada material — não apenas o sódio, mas também o oxigênio, o hidrogênio, o carbono e qualquer outro — tem seu próprio espectro único, ou seja, brilha com uma mistura característica de cores. O espectro de um material é, portanto, um pouco como uma assinatura, e observando atentamente essas assinaturas no espectro de uma galáxia, é possível descobrir do que ela é feita. É uma das descobertas mais notáveis da ciência: observando atentamente a cor de objetos distantes, podemos deduzir

descobrir do que são feitas e até mesmo quão quentes são, já que aquecer um material altera ligeiramente seu espectro característico. O SDSS disponibilizou espectros de alta qualidade para todas as galáxias do Galaxy Zoo, e foi observando atentamente o espectro das ervilhas que os zooítas descobriram que algumas delas estavam cercadas por gás oxigênio ionizado e quente.

(Não resisto a uma digressão para mencionar o fato maravilhoso de que a substância hélio foi descoberta usando análise espectral! Em 1868, os astrônomos Pierre Jules César Janssen e Joseph Norman Lockyer observaram independentemente que o espectro do Sol tinha características diferentes de qualquer substância já vista na Terra. Eles deduziram, corretamente, que estavam vendo o primeiro sinal de uma nova substância química. Mas foi somente quase 30 anos depois que um químico chamado William Ramsay descobriu o hélio na Terra.)

Chega de análise espectral; voltando ao Galaxy Zoo e ao mistério que as cercava. A essa altura — 12 de dezembro de 2007 — o tratador Kevin Schawinski já estava intrigado com essas estranhas galáxias. Ele decidiu observar as ervilhas mais de perto. Realizou alguns testes e rapidamente confirmou que se tratava de um novo tipo de galáxia.

Você poderia pensar que os astrônomos profissionais agora se mudariam e assumiriam o projeto. Afinal, os amadores do Galaxy Zoo tinham acabado de descobrir uma classe inteiramente nova de galáxias! Mas os profissionais, incluindo Schawinski, estavam ocupados com outras coisas, incluindo a Voorwerp de Hanny, e não assumiram o projeto imediatamente. Em vez disso, o que aconteceu em seguida foi uma notável obra científica conduzida pelos amadores. O tom foi dado por um zooísta chamado Rick Nowell. Nowell revisou todas as imagens de ervilhas que haviam sido postadas no fórum do Galaxy Zoo e identificou sistematicamente 39 objetos que pareciam ser o novo tipo de galáxia. Inspirados pela lista de Nowell, outras pessoas começaram a fazer suas próprias listas e a debater quais critérios deveriam ser usados para distinguir esse novo tipo de galáxia de objetos de aparência semelhante, como estrelas verdes. O tom do projeto começou a mudar, concentrando-se em chegar ao fundo do mistério da ervilha. As pessoas encontraram galáxias vermelhas com características semelhantes às ervilhas verdes, mas mais distantes. Cada vez mais, a discussão se concentrou nas propriedades detalhadas dos espectros das galáxias, e vários dos Zooítas se tornaram bastante adeptos da análise espectral — o tipo de especialização geralmente atribuída aos astrônomos profissionais.

A discussão de ideias nesse estágio foi surpreendente.
Gostaria de fazer um relato detalhado, mas levaria muito tempo até mesmo para resumir aqui — este não é um livro sobre como descobrir e compreender um novo tipo de galáxia! Mas o que foi especialmente notável

O que mais chamou a atenção na discussão foi o seu estilo. É o tipo de discussão que qualquer cientista reconhece. Descobertas científicas geralmente começam com um pouco de mistério, vagas suspeitas e algumas ideias malfeitas — assim como a vaga suspeita inicial de que as ervilhas poderiam ser um novo tipo de galáxia. Essa suspeita inicial é gradualmente refinada. Novas ideias são introduzidas, testadas, aprimoradas e, às vezes, descartadas. Os participantes ficam obcecados, à medida que suas suspeitas lentamente se transformam em fatos concretos e detalhados. Este é o processo de pesquisa, familiar a qualquer cientista pesquisador, e é exatamente o que se vê na discussão sobre as ervilhas no Galaxy Zoo. É assustadoramente reminiscente das discussões no Projeto Polímata. Os zoólogos podem ser amadores — eles sabem muito menos sobre astronomia do que muitos polímatas sobre matemática, e há mais leviandade na discussão sobre o Galaxy Zoo —, mas, por trás dessas diferenças, há o mesmo senso fértil de ideias crescendo e sendo refinadas, de uma convicção de que há algo aqui para ser conhecido e uma determinação para chegar ao fundo disso. Os zooítas não têm as credenciais de alguns polímatas. Mas são cientistas.

À medida que os zoólogos desenvolveram critérios mais precisos para caracterizar as galáxias-ervilha, eles também se tornaram mais sofisticados na busca por imagens candidatas. Eles não estavam mais apenas examinando as imagens do Galaxy Zoo manualmente. Em vez disso, recorreram aos dados originais do SDSS e desenvolveram consultas sofisticadas ao banco de dados que buscavam automaticamente em todo o conjunto de dados do SDSS por galáxias que se encaixassem em seus critérios. Essas candidatas foram então examinadas de perto por voluntários, e uma lista de cerca de 200 candidatas foi elaborada, parecendo provavelmente ser o novo tipo de galáxia-ervilha.

Os profissionais assistiram a toda essa discussão com interesse e, no início de julho de 2008, Schawinski, agora um cientista de pós-doutorado na Universidade de Yale, e uma aluna de Yale chamada Carolin Cardamone decidiram intensificar seu envolvimento. Em colaboração com os Zooítas, Cardamone e Schawinski iniciaram análises espectrais detalhadas das ervilhas usando um software de computador sofisticado. Nos nove meses seguintes, eles concluíram o trabalho iniciado pelos Zooítas. A imagem das ervilhas que surgiu mostrou que elas eram, de fato, um novo tipo de galáxia. Elas eram ultracompactas, com menos de 10% da massa da nossa Via Láctea, e formavam estrelas muito rapidamente — enquanto a Via Láctea produz apenas uma ou duas novas estrelas por ano, as ervilhas produzem cerca de 40 novas estrelas por ano, apesar de serem muito menores. E as galáxias eram extremamente brilhantes para seu tamanho.

As ervilhas e o voorwerp são apenas duas das muitas descobertas feitas pelo Galaxy Zoo. Outro projeto do Galaxy Zoo era buscar imagens de galáxias em fusão (veja a imagem na próxima página). Fusões são eventos transformadores para as galáxias e, portanto, entendê-las é de grande interesse para astrônomos e astrofísicos. Nossa Via Láctea está atualmente se fundindo com várias pequenas galáxias anãs e prevê-se que um dia se fundirá com a gigante galáxia de Andrômeda, atualmente a dois milhões de anos-luz de distância. Infelizmente, apesar de sua importância, galáxias em fusão não são tão fáceis de encontrar e, como resultado, a maioria dos estudos de fusões usa amostras contendo apenas algumas dezenas de galáxias em fusão. O projeto de fusão do Galaxy Zoo rapidamente encontrou 3.000 galáxias em fusão, um baú do tesouro de fusões para estudos futuros. Outros objetos que os Zooítas têm caçado incluem lentes gravitacionais (objetos cuja gravidade deforma e focaliza a luz de objetos mais distantes) e galáxias pareadas (galáxias que parecem estar uma em cima da outra, mas onde uma galáxia está, na verdade, muito mais próxima que a outra). Existe até um projeto voorwerp, e os Zooítas já caçaram com sucesso vários outros voorwerps.

No total, o Galaxy Zoo foi usado para escrever 22 artigos científicos, sobre uma ampla variedade de tópicos, e muitos outros artigos estão a caminho. As descobertas às vezes são fortuitas, como no caso do voorwerp, e às vezes baseadas em análises sistemáticas, como no projeto de fusões.

Às vezes, a serendipidade é acompanhada por uma extensa análise sistemática, como no estudo das ervilhas. Os projetos subsequentes Galaxy Zoo 2 e Galaxy Zoo: Hubble foram lançados e estão fornecendo informações ainda mais detalhadas sobre algumas das galáxias observadas pelo SDSS e também pelo Telescópio Espacial Hubble. Outros novos projetos da equipe que iniciou o Galaxy Zoo incluem o Moon Zoo, que visa compreender melhor as crateras da Lua, e o Projeto Solar Storm Watch, que visa detectar explosões no Sol. Um dos astrônomos envolvidos no Galaxy Zoo 2, Bob Nichol, da Universidade de Portsmouth, comparou o Galaxy Zoo à astronomia cotidiana desta forma:

Figura 7.3. Duas galáxias espirais em fusão (conhecidas conjuntamente como UGC 8335). Créditos: NASA, ESA, Hubble Heritage (STScI/AURA) - Colaboração ESA/Hubble e A. Evans (Universidade da Virgínia, Charlottesville/NRAO/Universidade Stony Brook).

> [No meu trabalho diário] Posso perguntar "quantas galáxias têm uma barra no meio" e, normalmente, embarcaria em uma jornada que duraria a vida toda para responder a essa pergunta fundamental. Posso até recrutar algum pobre estudante de pós-graduação para analisar 50.000 galáxias e responder à pergunta (como fizeram com o Kevin!).
> Mas agora, dois dias após o lançamento [do Galaxy Zoo 2], já temos os dados para responder a essa questão e é um pouco rápido demais para um veterano como eu. . . . A Internet é claramente a tecnologia revolucionária desta geração de astrônomos. . . . O Galaxy Zoo é uma demonstração incrível de quão poderosa esta nova ferramenta pode ser [quando] usada para abordar novas questões.

Como um computador, o Galaxy Zoo pode encontrar padrões em grandes conjuntos de dados, conjuntos de dados muito além da compreensão de qualquer indivíduo. Mas o Galaxy Zoo pode ir além dos computadores, porque também pode aplicar inteligência humana na análise, o tipo de inteligência que reconhece que a galáxia voorwerp ou uma ervilha verde é fora do comum e merece investigação mais aprofundada. O Galaxy Zoo é, portanto, um híbrido, capaz de fazer análises profundas de grandes conjuntos de dados que são impossíveis de qualquer outra forma. É uma nova maneira de transformar dados em conhecimento. Repetidamente, os tratadores do zoológico encontram novos astrônomos que dizem que seu trabalho poderia ser auxiliado pelo Galaxy Zoo, e mais de vinte astrônomos estão agora usando o Galaxy Zoo como uma forma de estudar uma ampla gama de questões astronômicas. O Galaxy Zoo está rapidamente se tornando uma plataforma de uso geral que conecta astrônomos profissionais a membros interessados do público em geral, para que possam fazer ciência juntos.

Quando amadores rivalizam com profissionais

Não é só na astronomia que a ciência cidadã é útil. Um dos grandes problemas em aberto na biologia é entender como o código genético dá origem à forma de um organismo. É claro que todos nós já ouvimos muitas vezes que o DNA é o "projeto da vida". Mas, embora o slogan seja familiar — afinal, é o destino dos grandes slogans se tornarem clichês —, isso não significa que alguém ainda entenda em detalhes como o DNA dá origem à vida.

Suponhamos que biólogos nunca tivessem visto a tromba de um elefante. Seriam capazes de examinar o DNA de um elefante e, de alguma forma, ver a tromba ali — ou seja, prever a existência da tromba com base apenas na sequência de pares de bases do código genético do elefante? Hoje, a resposta a essa pergunta é não: como o DNA determina a forma de um organismo é um dos mistérios da biologia.

Para ajudar a resolver esse mistério, um projeto de ciência cidadã chamado Foldit está recrutando voluntários online para jogar um jogo de computador que os desafia a descobrir como o DNA dá origem às moléculas chamadas proteínas. Esse desafio pode parecer muito distante de deduzir a existência da tromba do elefante — ***é*** muito distante —, mas é um passo crucial ao longo do caminho, porque as proteínas realizam muitos dos processos mais importantes em nossos corpos. Além de seu interesse científico intrínseco, o Foldit também é interessante como uma demonstração da grande complexidade do trabalho que pode ser feito.

ser feito por voluntários. No Galaxy Zoo, os participantes geralmente realizam tarefas simples, como classificar uma galáxia como espiral ou elíptica. No Foldit, os jogadores são solicitados a realizar tarefas que desafiariam um doutorado em bioquímica. E, como veremos, os melhores jogadores do Foldit estão realizando essas tarefas extraordinariamente bem.

Antes de discutirmos o Foldit em detalhes, vamos falar um pouco sobre proteínas em geral. Biólogos são obcecados por proteínas, e com razão: elas são moléculas que fazem tudo, desde digerir nossos alimentos até contrair nossos músculos. Um bom exemplo de proteína é a molécula de hemoglobina. A hemoglobina é um dos principais componentes do nosso sangue: é a molécula que nosso corpo usa para transportar oxigênio dos pulmões para o resto do corpo. Outra classe importante de proteínas são os anticorpos do nosso sistema imunológico. Cada anticorpo tem seu próprio formato especial que permite que ele se fixe em vírus e outros invasores em nosso corpo, marcando-os para serem atacados pelo nosso sistema imunológico.

Atualmente, compreendemos apenas parcialmente como o DNA dá origem a proteínas como a hemoglobina. O que sabemos é que certas seções do nosso DNA codificam proteínas, o que significa que descrevem uma proteína específica. Por exemplo, há uma seção codificadora de proteína para a hemoglobina em algum lugar do seu DNA. Essa região é uma longa sequência de bases de DNA, que começa assim: CACTCTTCTGGT... Acontece que é útil dividir essa sequência de bases em tripletos, que são chamados códons: CAC TCT TCT GGT. . . . A maneira como as proteínas são formadas é que cada códon na seção de codificação de proteína do seu DNA é transcrito em uma molécula correspondente na proteína chamada aminoácido. Assim, por exemplo, o primeiro códon para hemoglobina, CAC, é transcrito em um aminoácido conhecido como histidina. Não vou explicar exatamente o que é histidina, ou o que ela faz — para nós, não importa muito. O que importa é que em todos os lugares onde o códon CAC aparece na sequência de DNA da hemoglobina (ou de qualquer outra proteína), ele é transcrito para histidina. De forma semelhante, o segundo códon, TCT, é transcrito para o aminoácido serina. E assim por diante. A proteína resultante é uma cadeia contendo todos esses aminoácidos — então a hemoglobina é uma cadeia contendo histidina, serina e assim por diante.

Ok, até aqui, tudo bem: o DNA pode ser usado como uma receita para gerar proteínas. As proteínas, no entanto, diferem do DNA por terem cada uma sua forma especial, diferente da estrutura completamente regular do DNA. Essa forma é extremamente importante. Por exemplo, como mencionei antes, os anticorpos em nosso sistema imunológico são proteínas, e a forma de um anticorpo determina quais vírus ele pode atacar. O que acontece é que, à medida que a informação no DNA é transcrita para formar os aminoácidos

Na proteína, a proteína "se dobra" em seu formato. Como esse dobramento ocorre ainda é apenas parcialmente compreendido, mas existem algumas regras básicas que devem lhe dar uma ideia do que está acontecendo. Alguns aminoácidos gostam de estar perto da água — eles são chamados de ***hidrofílicos,*** das raízes gregas "hidro" e "filia", para água e amor, respectivamente. Como as proteínas dentro de uma célula são cercadas por água, a proteína tenderá a se dobrar para que os aminoácidos hidrofílicos fiquem do lado de fora, perto da água. Histidina e serina são exemplos de aminoácidos hidrofílicos. Em contraste, os aminoácidos ***hidrofóbicos*** — aminoácidos que não gostam de água — acabam agrupados dentro da proteína. Às vezes, essas tendências entram em conflito: aminoácidos vizinhos na proteína podem ser alternadamente hidrofóbicos e hidrofílicos, com o resultado de que a proteína pode acabar se dobrando em um formato muito complexo.

Há um truque incrivelmente inteligente aqui que a natureza está usando. O DNA é um arranjo de informações completamente regular, o que o torna fácil de copiar e relativamente simples de transcrever em aminoácidos. Mas a competição entre hidrofilia, hidrofobia e outras forças significa que a proteína pode se dobrar para formar formas complexas. Ao mudar o DNA, podemos mudar os aminoácidos na proteína, o que, por sua vez, faz com que o formato da proteína mude. O que é inteligente nisso é que nos leva da informação regularmente organizada no DNA, que é facilmente copiada, para as muitas formas possíveis da proteína.

A priori, formas não parecem tão fáceis de copiar. É como se você pudesse traçar a planta de uma casa, e a versão traçada, de alguma forma, surgisse como uma pequena maquete de casa. A conexão DNA-proteína é a maneira da Natureza facilitar a tarefa aparentemente impossível de copiar formas complexas.

Mas há um problema com essa história interessante. Só porque conhecemos a sequência de DNA de uma proteína não significa que podemos prever facilmente qual formato a proteína tem, ou o que ela fará. Na verdade, hoje temos apenas uma compreensão muito incompleta de como as proteínas se dobram. Estruturas completas — as formas exatas — são conhecidas para apenas 60.000 proteínas, apesar de conhecermos as sequências de DNA de milhões de proteínas. A maioria dessas estruturas completas foi encontrada usando uma técnica chamada difração ***de raios X*** — basicamente, projetar raios X em uma proteína e descobrir sua forma observando atentamente a sombra de raios X que ela projeta. É um trabalho lento, caro e meticuloso, e as técnicas estão melhorando gradualmente. O que realmente gostaríamos é de uma maneira rápida e confiável de prever a forma a partir da descrição genética. Se pudéssemos fazer isso, eliminando a lenta e cara etapa de difração de raios X, passaríamos de

conhecendo o formato de 60.000 proteínas até conhecer o formato de milhões. Ainda mais significativo, tal método seria uma ferramenta tremendamente poderosa para nos ajudar a projetar proteínas com os formatos desejados. Isso nos ajudaria, por exemplo, a desenvolver novos anticorpos para combater doenças.

Para resolver o problema do enovelamento de proteínas, bioquímicos recorreram a computadores na tentativa de prever o formato da proteína a partir da descrição genética. Para fazer suas previsões, eles usam a ideia de que uma proteína eventualmente se enovelará em sua forma de menor energia, assim como uma bola rola para o fundo de um vale entre duas colinas. Tudo o que é necessário é um bom método para encontrar a forma de menor energia de uma proteína. Isso parece promissor, mas na prática é difícil pesquisar em todas as formas possíveis, procurando a forma com a menor energia. A dificuldade é o número de formas diferentes em que uma proteína pode potencialmente se enovelar. As proteínas normalmente têm centenas ou até milhares de aminoácidos. Determinar a estrutura significa conhecer a posição e a orientação exatas de cada um desses aminoácidos. Com tantos aminoácidos envolvidos, o número de formas possíveis é astronômico, demais para pesquisar mesmo em um computador muito potente. Um enorme esforço tem sido feito para encontrar algoritmos inteligentes que possam ser usados para restringir o número de configurações que devem ser examinadas, e os algoritmos estão ficando muito bons. Mas ainda há um longo caminho a percorrer antes que possamos usar computadores para prever com segurança o formato das proteínas.

Em 2007, um bioquímico chamado David Baker e um pesquisador de computação gráfica chamado Zoran Popovic, ambos da Universidade de Washington, em Seattle, tiveram uma ideia para uma maneira melhor de resolver o problema. A ideia de Baker e Popovio era criar um jogo de computador que mostrasse uma proteína ao jogador e lhe desse controles para mudar a forma, girando a proteína, movendo aminoácidos e assim por diante. Alguns dos controles incorporados ao jogo são semelhantes às ferramentas usadas por bioquímicos profissionais. Quanto menor a energia da forma que o jogador cria, maior sua pontuação e, portanto, as formas com maior pontuação são boas candidatas para a forma real da proteína. Baker e Popovic esperavam que esta pudesse ser uma abordagem melhor para o enovelamento de proteínas do que as abordagens convencionais, combinando técnicas computacionais de última geração com a persistência e as habilidades dos jogadores de computador em correspondência de padrões e resolução de problemas 3D.

Fiquei cético quando ouvi falar do Foldit pela primeira vez. Parecia aqueles jogos educativos de computador sem graça que eu via na escola quando era criança, na década de 1980. Mas baixei o jogo e passei horas jogando por vários dias. Naquele momento, a desculpa "Estou fazendo pesquisa para o meu livro" surgiu.

estava rapidamente se tornando um eufemismo para "esta é uma ótima maneira de procrastinar a escrita do meu livro", e me forcei a parar. Até agora, mais de 75.000 pessoas se inscreveram. As pessoas jogam o jogo porque é bom. Ele tem a qualidade envolvente e viciante que todos os bons jogos de computador têm: uma tarefa desafiadora, mas não impossível, feedback instantâneo sobre o seu desempenho e a sensação de que você está sempre a um passo da melhoria. É a mesma qualidade viciante que vimos anteriormente na competição MathWorks e que também é sentida por muitos participantes do Galaxy Zoo. Além disso, como o Galaxy Zoo, o Foldit é profundamente significativo para muitos dos jogadores. Einstein certa vez explicou por que estava mais interessado em ciência do que em política, dizendo: "Equações são mais importantes para mim, porque política é para o presente, mas uma equação é algo para a eternidade". Cada vez que você classifica uma galáxia ou encontra uma maneira melhor de dobrar uma proteína, você está fazendo uma pequena, mas real contribuição para o conhecimento humano. Para muitos participantes, Foldit e Galaxy Zoo não são prazeres culposos, como jogar World of Warcraft ou outros jogos online. Em vez disso, são uma forma de contribuir com algo importante para a sociedade. Um dos principais jogadores de Foldit, Aotearoa, descreve-o como "o jogo mais desafiador, emocionante, estimulante, intenso e viciante que já joguei" e comenta que ele oferece uma maneira para as pessoas "oferecerem algo proativo para resolver alguns dos quebra-cabeças mais complexos do mundo/sociedade, em vez de perder tempo jogando um 'jogo' que não oferece as mesmas 'recompensas' que dobrar proteínas, desta forma!"

Além da motivação individual para jogar, o Foldit também incentiva a resolução coletiva de problemas pelos jogadores. Há um fórum de discussão online e um wiki, onde os jogadores compartilham notícias e discutem suas estratégias para o enovelamento de proteínas. O jogo incorpora uma linguagem de programação simples que os jogadores podem usar para criar scripts — programas curtos — que automatizam tarefas do jogo. Um script típico pode implementar uma estratégia para melhorar um enovelamento ou identificar qual parte do formato atual da proteína mais precisa de melhorias. Centenas desses scripts foram compartilhados publicamente — uma abordagem de código aberto para o enovelamento de proteínas. Muitos jogadores trabalham em grupos, compartilhando suas ideias sobre as melhores maneiras de enovelamento. Todo esse trabalho é amplamente informado pela pontuação do jogo, que, como na competição MathWorks, concentra a atenção dos participantes onde ela será mais útil: quando um dos jogadores com maior pontuação compartilha uma dica de estratégia ou um script, os outros jogadores prestam atenção. Os próprios atores são muito variados, desde um autodenominado "caipira educado" de Dallas, Texas, até um historiador de teatro de Dakota do Sul, e uma avó de três filhos com ensino médio.

Quão bons são os jogadores do Foldit em enovelar proteínas? A cada dois anos, desde 1994, há uma competição mundial de bioquímicos usando computadores para prever estruturas de proteínas. A competição, chamada CASP (Avaliação Crítica de Técnicas para Predição de Estrutura de Proteínas), é muito importante para os cientistas que trabalham com predição de estrutura de proteínas. Antes do início da competição, os organizadores do CASP abordam algumas das instalações que determinam a estrutura de proteínas usando a abordagem tradicional de difração de raios X e perguntam quais estruturas de proteínas eles esperam completar nos próximos meses. Eles então usam essas proteínas como quebra-cabeças no CASP. Começando com a sequência de aminoácidos que compõem a proteína, os competidores do CASP são solicitados a prever a estrutura. Ao final da competição, as equipes são classificadas de acordo com o quão perto elas chegam da estrutura real.

Jogadores do Foldit competiram nas competições CASP 2008 e 2010. Eles tiveram um desempenho extremamente bom, terminando perto ou no topo em muitos dos desafios do CASP. O desenvolvedor do Foldit, Zoran Popoviÿ, resumiu os resultados da competição de 2008 dizendo que "os jogadores do Foldit estão no mesmo nível, mas não melhores do que os especialistas em dobramento de proteínas ao tentar resolver o mesmo problema com todas as ferramentas disponíveis. Parece também que o Foldit superou todos os envios de servidor totalmente automatizados". Assim, uma equipe de amadores pode ser competitiva com alguns dos melhores bioquímicos do mundo, equipada com computadores de última geração. Popoviÿ me disse que seu "objetivo final é mostrar que os especialistas são inequivocamente inferiores à população em geral com este problema... um PhD em bioquímica não se autoseleciona para raciocínio espacial. A previsão de estruturas é toda sobre resolução de problemas 3D e muito pouco sobre bioquímica". De fato, mesmo especialistas em previsão de estruturas de proteínas geralmente gastam apenas uma pequena fração de seu tempo trabalhando diretamente E embora possuam uma expertise que os amadores não possuem, grande parte desse conhecimento está incorporada na mecânica do jogo. Isso nivela o campo de jogo o suficiente para que a disparidade restante em expertise possa ser superada pelo maior comprometimento de tempo dos jogadores de Foldit. É uma simbiose: os profissionais desenvolvem a compreensão sistemática que fundamenta a mecânica do jogo, e os amadores, então, fornecem a arte dedicada necessária para aproveitar ao máximo essa compreensão sistemática.

Ciência Cidadã Hoje

A ciência cidadã não é uma invenção da era da internet. Muitos dos primeiros cientistas eram amadores, muitas vezes praticando a ciência como hobby, em paralelo a uma profissão mais lucrativa, como a astrologia. Mas mesmo após a profissionalização da ciência, os amadores continuaram a dominar algumas áreas da ciência. Por exemplo, muitos dos caçadores de cometas mais bem-sucedidos da história foram astrônomos amadores, pessoas como John Caister Bennett, um funcionário público da cidade sul-africana de Pretória, que descobriu um dos cometas mais espetaculares do século XX, o grande cometa de 1968, o Cometa Bennett.

Embora a ciência cidadã não seja novidade, as ferramentas online estão permitindo que muito mais pessoas participem — pense nos mais de 200.000 participantes do Galaxy Zoo e nos mais de 75.000 participantes do Foldit — e também expandindo o alcance do trabalho científico que essas pessoas podem realizar. Para ser um caçador de cometas na década de 1960, era preciso comprar ou construir um telescópio, aprender a usá-lo e, então, passar muitas e muitas horas observando o céu. As barreiras para entrar e para continuar contribuindo eram altas. Em contraste, você pode começar no Galaxy Zoo ou no Foldit em questão de minutos. É até possível classificar galáxias no seu smartphone. Além de eliminar as barreiras de entrada, as ferramentas online também permitem treinamentos interativos sofisticados e reúnem os participantes em comunidades onde podem aprender uns com os outros e apoiar o trabalho uns dos outros. Como resultado, estamos vendo um grande florescimento da ciência cidadã.

Como exemplo desse florescimento, a caça aos cometas foi transformada pela internet. Em 1995, a Agência Espacial Europeia e a NASA lançaram uma nave espacial chamada SOHO, projetada para tirar fotos excepcionalmente boas do Sol e de sua vizinhança imediata. (SOHO significa Observatório Solar e Heliosférico.) Acontece que perto do Sol é um ótimo lugar para procurar cometas, em parte porque os cometas são muito bem iluminados lá e em parte porque suas caudas são alongadas pelo vento solar. Normalmente, esses cometas não apareceriam em fotos devido ao brilho do Sol, mas um dos instrumentos no SOHO foi especialmente projetado para bloquear a luz do corpo principal do Sol, permitindo que ele tire fotos da coroa solar — a "atmosfera" de plasma logo acima da superfície solar. A equipe do SOHO decidiu compartilhar suas imagens da coroa abertamente na internet, e muitos caçadores amadores de cometas começaram a vasculhar as fotos em busca de cometas. O mais bem-sucedido é um alemão

O astrônomo amador Rainer Kracht, que passa horas por semana observando com extremo cuidado as imagens do SOHO, tornou-se o caçador de cometas mais bem-sucedido da história, tendo descoberto até agora mais de 250 cometas, quase um em cada 15 cometas já descobertos.

Outro exemplo de ciência cidadã é o Projeto eBird, administrado pelo Laboratório de Ornitologia da Universidade Cornell. O eBird solicita que observadores amadores de pássaros enviem informações sobre as aves que observam para um site online: que espécie de ave eles viram, quando a viram e onde a viram. Ao combinar todas as observações enviadas, o eBird pode construir uma compreensão das populações de aves do mundo. Este é outro caso em que a ciência cidadã está se baseando em uma tradição anterior, desta vez uma tradição de colaboração entre observadores amadores de pássaros e ornitólogos profissionais. Mas o Projeto eBird está possibilitando essa colaboração em uma escala sem precedentes, com participantes relatando até agora mais de 30 milhões de observações de aves. Cerca de 2.500 observadores de pássaros são colaboradores frequentes do site, fazendo 50 ou mais contribuições, e dezenas de milhares de pessoas usam o site regularmente. Os dados coletados podem ser usados, por exemplo, para gerar mapas de distribuição mostrando a densidade de algumas espécies específicas de aves em diferentes locais. À medida que o eBird coleta mais dados (começou em 2002), esses mapas de distribuição se tornarão cada vez mais úteis para rastrear o impacto sobre as aves de efeitos como mudanças climáticas, mudanças na população humana próxima e outros fatores ambientais.

Outro exemplo de ciência cidadã vem do estudo dos dinossauros. A maioria das pesquisas sobre dinossauros concentra-se em apenas um ou alguns fósseis. Em setembro de 2009, os paleontólogos Andy Farke, Mathew Wedel e Mike Taylor tiveram a ideia de criar um grande banco de dados contendo informações sobre muitos dinossauros, combinando os resultados de centenas ou até milhares de artigos científicos. A esperança deles era que o banco de dados pudesse então ser explorado para responder a muitas novas perguntas. Mas, em vez de construir o banco de dados sozinhos, eles decidiram aproveitar o conhecimento distribuído e o esforço de uma comunidade mais ampla de pessoas.
Eles iniciaram o Open Dinosaur Project, recrutando pessoas do mundo todo para, digamos, desenterrar artigos sobre dinossauros. Enquanto escrevo, eles estão se concentrando em medições de membros de dinossauros. Se um voluntário encontra um artigo estudando, digamos, um espécime ***de Stegosaurus*** com um fêmur direito de 1.242 milímetros de comprimento, ele registra esses dados no banco de dados. O projeto, portanto, criou uma lista de medições de 1.659 espécimes distintos de dinossauros, com contribuições de 46 pessoas, muitas delas amadoras.
A esperança deles é que isso lhes permita responder a perguntas sobre (por exemplo)

A evolução da locomoção dos dinossauros. O Projeto Dinossauro Aberto ainda está em seus primórdios e, embora os dados estejam sendo coletados rapidamente, é muito cedo para dizer sua utilidade. Mas este é mais um exemplo de como uma comunidade composta por cientistas amadores e profissionais pode fazer mais do que qualquer um dos grupos conseguiria sozinho.

A partir desses e de exemplos anteriores, vemos diversas maneiras distintas pelas quais os cientistas cidadãos estão contribuindo para a ciência. A ciência cidadã pode ser uma maneira poderosa tanto de coletar quanto de analisar enormes conjuntos de dados. Nesses conjuntos de dados, os cientistas cidadãos podem explorar o incomum e o inesperado, descobertas como a do voorwerp e a das ervilhas, descobertas que seriam difíceis de programar um computador para detectar. A ciência cidadã, portanto, complementa as ferramentas de inteligência orientada por dados descritas no capítulo anterior.

Cientistas cidadãos também podem trabalhar para estender simbioticamente a capacidade dessas ferramentas, como demonstrado pela habilidade dos jogadores do Foldit em usar as ferramentas de predição da estrutura de proteínas. Em outra reviravolta nessa ideia, os tratadores do zoológico usaram recentemente as classificações de galáxias dos zoóitas para treinar um algoritmo computacional para distinguir entre galáxias espirais e elípticas. Os resultados preliminares são promissores, com o algoritmo alcançando 90% de concordância com as classificações humanas. Esse resultado é interessante em parte porque futuros levantamentos do céu com instrumentos como o Large Synoptic Survey Telescope (LSST, descrito na página 107) produzirão muito mais dados do que até mesmo a enorme multidão de voluntários do Galaxy Zoo pode esperar analisar. Talvez os resultados do LSST sejam compreendidos primeiro pedindo a amadores que analisem uma pequena parte dos dados e, em seguida, usando algoritmos computacionais para aprender com as análises dos amadores, com os computadores completando a classificação de todo o conjunto de dados. Possibilidades como essas estão criando um enorme florescimento de projetos de ciência cidadã, com pessoas comuns participando de pesquisas científicas de maneiras inimagináveis uma geração atrás.

Quanto a ciência cidadã mudará a ciência?

Exemplos como Galaxy Zoo, Foldit e o projeto Open Dinosaur são interessantes e divertidos. Mas a ciência é vasta e, embora a ciência cidadã seja

Embora provavelmente cresça rapidamente nos próximos anos e décadas, isso não significa que se tornará uma parte dominante de como a ciência é feita. Embora projetos como o Galaxy Zoo sejam importantes, não é óbvio se são curiosidades ou prenúncios de uma mudança mais ampla na ciência. A ciência cidadã algum dia terá um impacto amplo e decisivo na forma como a ciência é feita? Ou está destinada a ser útil principalmente em alguns setores específicos da ciência? Não sei a resposta para essas perguntas. Estamos apenas começando a explorar as maneiras pelas quais as ferramentas online podem expandir o impacto da ciência cidadã. A situação é bem diferente das mudanças descritas no último capítulo. Lá, como vimos, novas ferramentas poderosas para encontrar significado no conhecimento já estão revolucionando muitas áreas da ciência. Até o momento, as perspectivas para a ciência cidadã são mais incertas. Mas, embora não possamos saber com certeza quão importante a ciência cidadã será, podemos pelo menos pensar um pouco mais sobre seu potencial, onde ela pode ser aplicada e quais podem ser suas limitações.

Parte desse potencial reside na criação de comunidades estimulantes e de apoio à ciência cidadã. Antes da internet, a maioria dos cientistas cidadãos trabalhava em grande parte por conta própria, isolados do incentivo e das críticas dos colegas. Hoje, isso está mudando. Nos fóruns do Galaxy Zoo, você vê uma comunidade onde as pessoas se ajudam mutuamente, um ambiente de apoio no qual podem aprender e crescer como astrônomos, um lugar onde as pessoas podem fazer perguntas e outras pessoas responderão de forma amigável. Considere, por exemplo, a maneira como os zooítas se ajudaram mutuamente em sua busca para entender as galáxias da ervilha-verde.

Eles criticavam e aprimoravam repetidamente as ideias uns dos outros sobre o que tornava as ervilhas verdes únicas, incentivando uns aos outros e compartilhando informações sobre problemas como a melhor forma de analisar o espectro de uma galáxia ou como fazer consultas em bancos de dados para encontrar ervilhas verdes automaticamente nos dados do SDSS. Quando você está em uma comunidade como essa, você recebe um feedback constante que diz, na verdade: "Ei, isso é importante, isso é o que realmente ***importa".***
Pense na maneira como as crianças jogam futebol ou beisebol nas ruas e parques — elas jogam incansavelmente, hora após hora, dia após dia, gradualmente se aprimorando como parte de uma comunidade que exige o máximo delas e torna a conquista do máximo uma alegria. Todas as comunidades mais criativas fazem o mesmo.

Esse novo tipo de construção de comunidade é importante, mas os projetos de ciência cidadã atuais têm muito a melhorar. Galaxy Zoo, Foldit e a maioria dos outros projetos de ciência cidadã ainda não contam com o tipo de estrutura de apoio para desenvolvimento e mentoria disponível para cientistas profissionais, apoio que os ajude a adquirir a ampla base de conhecimento necessária para muitos tipos de trabalho científico.

Será interessante observar como os projetos de ciência cidadã evoluem. Veremos ambientes de aprendizagem cada vez mais eficazes, um lugar onde amadores possam aprender à medida que avançam, adquirindo gradualmente mais expertise? Veremos o surgimento de sistemas de mentoria, proporcionando às pessoas uma forma estruturada de aprendizagem? Imagine comunidades online construídas em torno de séries de seminários e conferências virtuais, sessões de perguntas e respostas online e grupos de discussão. Essas e outras ideias podem ser usadas para criar uma comunidade online desafiadora e gratificante que apoie a ciência cidadã.

Os maiores projetos de ciência cidadã recrutaram um grande número de pessoas — o Galaxy Zoo tem mais de 200.000 participantes — e você pode se perguntar se ainda há muito espaço para a ciência cidadã crescer. Ou será que o interesse do público por ciência cidadã já se esgotou?

Há uma maneira interessante de pensar sobre essas questões, inspirada em uma análise das perguntas análogas para a Wikipédia feita pelo autor Clay Shirky, da Universidade de Nova York. Para começar, vamos fazer uma estimativa aproximada do esforço total envolvido em um projeto como o Galaxy Zoo. Até agora, os zooítas realizaram aproximadamente 150 milhões de classificações de galáxias. Se cada classificação levar, digamos, 12 segundos, isso equivale a 500 mil horas de trabalho. É como ter 250 funcionários trabalhando em tempo integral por um ano! Embora seja uma quantidade impressionante de trabalho, na escala da sociedade como um todo, é uma gota no oceano. Em média, os americanos assistem a cinco horas de televisão por dia, o que, ao longo de um ano, significa que os americanos assistem a mais de 500 bilhões de horas de televisão. Isso representa um milhão de projetos do Galaxy Zoo!

Vamos analisar uma atividade mais próxima do Galaxy Zoo em escala. O clube de futebol inglês Manchester United tem capacidade para 76.000 pessoas em seu estádio, o Old Trafford. Os jogos duram duas horas, com paradas, então os espectadores de uma partida gastam cerca de 150.000 horas no total, quase um terço do tempo que os Zooites gastaram classificando galáxias! Em outras palavras, imagine que você lotou o estádio do Manchester United e, em vez de assistir a um jogo de futebol, pediu às pessoas que classificassem galáxias por algumas horas. Se você fizesse isso três vezes, aproximadamente igualaria o esforço investido no Galaxy Zoo. É claro que o Galaxy Zoo está em cartaz há três anos, enquanto escrevo, enquanto o Manchester United joga dezenas de jogos em casa por ano. Portanto, os Zooites estão um ou dois níveis abaixo da devoção demonstrada pelos torcedores do Manchester United em seus jogos em casa. Uma comparação mais próxima seria com um clube de futebol muito menor, como o Bristol Rovers, que atrai alguns milhares de torcedores para cada jogo em casa. Há muito espaço para a ciência cidadã crescer!

Shirky cunhou a expressão "excedente cognitivo" para descrever o tempo e a energia disponíveis em nossa sociedade — todo o tempo que temos coletivamente quando não estamos lidando com as obrigações básicas da vida, como ganhar a vida ou alimentar nossa família. É o tempo que dedicamos a atividades de lazer, como assistir televisão, sair com amigos ou relaxar com um hobby. Geralmente, essas são atividades que fazemos individualmente ou em pequenos grupos.

O que as ferramentas online fazem é facilitar a coordenação de projetos criativos complexos em um grupo grande. Sempre foi possível reunir um grande grupo de pessoas e torcer em um jogo de futebol. Mas é muito mais difícil reunir um grande grupo de pessoas para trabalhar em prol de um objetivo criativo complexo. Uma maneira é pagar todas essas pessoas para se unirem e formarem uma hierarquia organizada em gerentes e subordinados. Chamamos isso de empresa, organização sem fins lucrativos ou governo. Mas, sem dinheiro, historicamente tem sido difícil manter projetos criativos tão complexos em conjunto. As ferramentas online facilitam muito essa coordenação complexa, mesmo sem dinheiro como motivador. Como Shirky poeticamente coloca:

> Estamos acostumados a um mundo onde pequenas coisas acontecem por amor e grandes coisas acontecem por dinheiro. O amor motiva as pessoas a fazer um bolo e o dinheiro motiva as pessoas a criar uma enciclopédia. Agora, porém, podemos fazer grandes coisas por amor.

Projetos como o Galaxy Zoo e o Foldit estão fazendo exatamente isso, usando o excedente cognitivo da nossa sociedade para resolver problemas científicos.

Quanto do excedente cognitivo da nossa sociedade será usado para fazer ciência cidadã? Hoje, não é possível responder a essa pergunta. A ciência cidadã está nos primeiros dias de uma grande expansão possibilitada por ferramentas online.

A extensão final da expansão dependerá da imaginação dos cientistas em encontrar novas maneiras inteligentes de se conectar com leigos, maneiras que os inspirem e os ajudem a fazer contribuições que considerem significativas. Você pode ter um vislumbre disso na história de uma das participantes mais prolíficas do Galaxy Zoo, uma mulher chamada Aida Berges.

Berges é uma dona de casa de 53 anos, mãe de dois filhos, natural da República Dominicana e atualmente morando em Porto Rico. Ela classifica centenas de galáxias todas as semanas, totalizando mais de 40.000 galáxias até agora. Ela trabalhou na busca por ervilhas verdes, por voorwerps, por galáxias em fusão e em muitos outros projetos. Ela descobriu duas estrelas de hipervelocidade, estrelas que se movem tão rápido que estão, na verdade, deixando nossa galáxia. Menos de vinte estrelas desse tipo foram descobertas, no total. Sra. Berges

juntou-se ao Galaxy Zoo depois de ler sobre ele online e disse sobre a experiência que “minha vida mudou para sempre... foi como voltar para casa meu.

Os cínicos dirão que a maioria das pessoas não é inteligente ou interessada o suficiente para contribuir com a ciência. Acredito que projetos como o Galaxy Zoo e o Foldit mostram que esses cínicos estão errados. A maioria das pessoas é inteligente o suficiente para contribuir com a ciência, e muitas delas estão interessadas. Tudo o que falta são ferramentas que as ajudem a se conectar com a comunidade científica de forma que lhes permitam fazer essa contribuição.

Hoje, podemos construir essas ferramentas.

Mudando o papel da ciência na sociedade

Depois que Jonas Salk anunciou sua vacina contra a poliomielite em 1955, ela foi rapidamente difundida nos países ricos e desenvolvidos, e as taxas de poliomielite despencaram. Mas nos países em desenvolvimento a história foi diferente. Em 1988, cerca de 350.000 pessoas no mundo em desenvolvimento foram infectadas com poliomielite. Naquele ano, a Organização Mundial da Saúde (OMS) decidiu lançar uma iniciativa global para erradicar a doença. Eles fizeram progressos rápidos e, em 2003, havia apenas 784 novos casos em todo o mundo, a maioria concentrada em apenas alguns países. O país mais atingido foi a Nigéria, onde quase metade (355) dos novos casos ocorreram. A OMS decidiu lançar um grande programa de vacinação na Nigéria, mas a iniciativa foi bloqueada por líderes políticos e religiosos em três estados do norte da Nigéria — Kano, Zamfara e Kaduna — com uma população total de 18 milhões de pessoas.

Líderes nesses estados alertaram que as vacinas poderiam estar contaminadas com agentes causadores do HIV/AIDS e infertilidade, e aconselharam os pais a não permitirem que seus filhos fossem vacinados. O governo de Kano descreveu sua oposição à vacinação como "o menor de dois males: sacrificar duas, três, quatro, cinco ou até mesmo dez crianças à poliomielite [em vez de] permitir que centenas de milhares ou possivelmente milhões de meninas provavelmente se tornem inférteis". O líder do poderoso Conselho Supremo da Sharia do Estado de Kano disse que as vacinas contra a poliomielite foram "corrompidas e contaminadas por malfeitores da América e seus aliados ocidentais". As vacinações foram suspensas em Kano, e um novo surto de poliomielite ocorreu, espalhando-se para

oito países vizinhos, causando paralisia em 1.500 crianças.

A vacinação contra a poliomielite está longe de ser a única questão em que a boa ciência não leva necessariamente a bons resultados de saúde pública. No Reino Unido, o uso da vacina contra sarampo, caxumba e rubéola caiu drasticamente no início dos anos 2000, após um artigo de 1998 na prestigiosa revista médica ***The Lancet*** sugerir que a vacina poderia causar autismo em crianças. (A metodologia do artigo era falha e foi posteriormente retratada pela revista e pela maioria dos autores). A suposta ligação entre vacina e autismo tornou-se um tópico de grande controvérsia pública no Reino Unido, com o primeiro-ministro Tony Blair apoiando publicamente a vacina, mas recusando-se a confirmar se seu filho Leo havia sido vacinado. A taxa de vacinação caiu de 92% para 80%. Isso pode parecer uma pequena queda, mas o número de casos de sarampo aumentou drasticamente, aumentando dezessete vezes em apenas alguns anos. Para entender por que o aumento do sarampo foi tão drástico — e, portanto, por que uma queda nas taxas de vacinação é tão importante — observe que a fração de pessoas ***não*** vacinadas aumentou de 8% para 20%. Em termos gerais, isso significava que uma pessoa infectada com sarampo estaria exposta a duas vezes e meia mais pessoas suscetíveis do que antes. E se alguma dessas pessoas contraísse sarampo, estaria, por sua vez, exposta a duas vezes e meia mais pessoas suscetíveis do que antes. E assim por diante. É por isso que mesmo uma pequena queda nas taxas de vacinação pode causar um grande aumento na incidência da doença.

Apesar dos fiascos com vacinas, nossa sociedade frequentemente faz um bom trabalho convertendo ciência em bem social. Mercados e empreendedorismo, por exemplo, são instituições poderosas que frequentemente ajudam a transformar a ciência em bens que melhoram nossas vidas. Pense em um desenvolvimento como os lasers. Quando os lasers foram inventados, muitas pessoas os consideravam brinquedos com pouca utilidade aparente. Mas empreendedores descobriram maneiras engenhosas de usar lasers para fazer tudo, desde reproduzir filmes (DVDs) até corrigir a visão por meio de cirurgia ocular a laser. Como sociedade, somos muito, muito bons em usar a ciência para desenvolver novos produtos para entrega ao mercado.

Mas, embora sejamos bons em levar ciência ao mercado, temos um histórico mais misto quando se trata de levar ciência por meio de políticas públicas.
Em um mercado, todos podem decidir por si mesmos se querem usar um produto. Se a cirurgia ocular a laser te deixa enjoado, ninguém está te obrigando a fazer. Mas as decisões políticas são frequentemente decisões coletivas, como se a vacinação infantil deveria ser obrigatória. Tais decisões não podem ser tomadas individualmente, como em um mercado, mas exigem amplo consenso para serem eficazes. E quando cientistas descobrem algo com impacto político drástico...

Se, por exemplo, as emissões humanas de dióxido de carbono estão levando ao aquecimento global, elas são tratadas, em muitos aspectos, como apenas mais um grupo de interesse tentando pressionar o governo. Mas a ciência não é apenas um grupo de interesse. É uma forma de entender o mundo. Idealmente, nossas instituições de governança incorporariam nas políticas públicas o conhecimento adquirido pela ciência — por mais imperfeito, incerto e provisório que seja — da melhor forma possível. Mas nas democracias de hoje, não é isso que acontece. Este é o problema da ciência na democracia.

Não tenho soluções para o problema das vacinas ou, de forma mais ampla, para o problema da ciência na democracia. Estou descrevendo esses problemas porque são exemplos concretos de falhas críticas no papel que a ciência desempenha atualmente em nossa sociedade. Qualquer solução para esses e outros problemas semelhantes exigirá grandes mudanças no papel da ciência na sociedade. Na maioria das vezes, essas mudanças ocorrem muito lentamente e, por isso, é tentador tomar esse papel como garantido, encará-lo como um estado natural das coisas. Mas, na verdade, o estado atual das coisas não é nada natural: o papel da ciência tem sido radicalmente diferente em diferentes sociedades e em diferentes épocas — basta pensar, por exemplo, em todas as sociedades nas quais o pensamento científico foi totalmente suprimido. Historicamente, grandes mudanças no papel da ciência têm sido frequentemente impulsionadas por novas tecnologias e pelas novas instituições que elas possib Pense no papel da imprensa como facilitadora do Renascimento, da Reforma e do Iluminismo. Podemos mudar o papel da ciência na sociedade se mudarmos as respostas institucionais que damos a perguntas fundamentais como "Quem financia a ciência?" ou "Como a ciência é incorporada às políticas governamentais?" ou mesmo "Quem pode ser um cientista?"

Como exemplo concreto da forma como as instituições impactam o papel da ciência, voltemos ao sistema de mercado. A importância do mercado para o papel da ciência é vividamente ilustrada pelo que aconteceu quando o mercado foi suprimido na União Soviética. Embora a União Soviética tivesse um dos melhores sistemas de pesquisa científica do mundo, sem um sistema de mercado, era praticamente incapaz de disponibilizar inovações científicas aos seus cidadãos. Outro exemplo do poder das instituições é a forma como a introdução da escolaridade obrigatória aumentou a alfabetização científica geral. Embora seja senso comum em muitos círculos reclamar dos padrões de alfabetização científica, pelos padrões históricos, vivemos em uma era iluminada. Tanto o mercado quanto as escolas atuam como instituições de ponte, conectando a ciência à sociedade de uma forma que traz muitos benefícios sociais. Como último exemplo, desta vez negativo, considere a supressão da ciência pela Igreja Cristã primitiva. Isso durou mais de um milênio, desde o reinado do imperador cristão Justiniano.

do fechamento da Academia em Atenas em 529 d.C. até o julgamento e prisão domiciliar de Galileu em 1633 d.C.

Ao mudar as instituições da nossa sociedade, podemos mudar radicalmente o papel da ciência na sociedade e, talvez, abordar alguns dos problemas mais significativos da nossa sociedade. Para isso, serão necessárias a imaginação e a vontade de inventar novos mecanismos institucionais que possam abordar problemas como o problema das vacinas ou o problema da ciência na democracia. Pode parecer irrealista mudar as nossas instituições desta forma. Na maioria das vezes, as instituições mudam muito lentamente. Mas, hoje em dia, não é assim na maioria das vezes. Ferramentas online são máquinas geradoras de instituições. Exemplos como Galaxy Zoo, Wikipédia e Linux demonstram o quanto se tornou mais fácil criar novas instituições, e até mesmo criar tipos radicalmente novos de instituições. Ao mesmo tempo, as ferramentas online estão a transformar as instituições existentes na nossa sociedade — considere o colapso das empresas tradicionais de música e jornais nos últimos dez anos e a ascensão gradual de novos modelos em seu lugar. E assim, estamos num ponto muito interessante da história, um ponto em que se tornou muito mais fácil criar novas instituições e reinventar as instituições existentes. Isso não significa que possamos resolver facilmente problemas como o problema das vacinas. O que isso significa é que temos a oportunidade de reimaginar e, até certo ponto, recriar o papel da ciência na sociedade. Já estamos começando a ver isso acontecer, com projetos de ciência cidadã como o Galaxy Zoo e o Foldit mostrando como ferramentas online podem ser usadas para mudar algo fundamental: quem pode ser cientista. No restante deste capítulo, exploraremos outras maneiras pelas quais as ferramentas online mudam o papel da ciência na sociedade, incluindo maneiras pelas quais elas melhoram o acesso público tanto aos resultados da ciência quanto aos próprios cientistas.

Acesso aberto

Imagine que você é uma mulher que foi ao médico para uma mamografia de rotina e seu médico voltou com uma notícia surpreendente e terrível: você tem câncer de mama em estágio inicial. Chocada, você vai para casa e começa a planejar seu ataque à doença. Você decide que a primeira coisa a fazer é se informar melhor. Você pesquisa online e descobre muitas informações úteis em sites como o

cancer.gov, administrado pelo Instituto Nacional do Câncer dos EUA. Mas, depois de um tempo, todas as informações introdutórias que você encontra na web se tornam repetitivas. Você quer conhecimento mais atualizado sobre as pesquisas atuais mais promissoras. Um amigo menciona que o Google tem um mecanismo de busca especial — chamado Google Acadêmico — que ajudará você a pesquisar na literatura científica os melhores e mais recentes artigos sobre câncer de mama. Você acessa o site, pesquisa "câncer de mama" e descobre milhares de artigos. Excelente! Melhor ainda, o Google Acadêmico ordena os resultados de acordo com a melhor estimativa do Google quanto à sua importância. Você baixa o artigo que o Google classifica como o melhor resultado e descobre que precisa pagar 50 dólares pelo download. "Deixa pra lá", você pensa, "voltarei a esse artigo mais tarde. Mas quando você olha o segundo artigo, descobre que custa 15 dólares para baixar. No terceiro artigo, a editora quer cobrar de você também, mas é recatada sobre o preço, pedindo que você se registre no site deles primeiro. Conforme você continua folheando os resultados, o padrão de taxas continua, e sua euforia inicial se transforma em descrença raivosa. "Certamente", você raciocina, "com dezenas de bilhões de dólares do dinheiro do contribuinte gastos a cada ano em pesquisa científica, deveríamos pelo menos poder ler os resultados dessa pesquisa?!Ý Agora, o câncer de mama é uma doença séria e você fica tentado a engolir sua raiva e pagar as taxas. Mas existem milhares de artigos. Não há como você pagar nem mesmo por uma pequena fração deles.

A publicação científica tradicional baseia-se num modelo de pagamento por acesso. Em muitos aspetos, funciona de forma muito semelhante ao mercado das revistas, e há menos diferença do que se possa imaginar entre uma revista científica líder como a ***Physical Review Letters*** e revistas como ***a Time*** e ***a People.*** Tal como as revistas, as revistas científicas são coletâneas de artigos, mas em vez de discutir notícias, política e celebridades, os artigos descrevem descobertas científicas. As revistas podem não ter capas e anúncios chamativos, nem encontrará a maioria delas em exposição nas bancas de jornal locais, mas tanto as revistas como os jornais ganham muito dinheiro cobrando aos leitores. Uma assinatura anual de uma revista pode custar centenas, milhares ou mesmo dezenas de milhares de dólares. E, como acabámos de ver, as revistas complementam essas taxas cobrando pelo acesso único a artigos na web, normalmente entre 10 e 50 dólares.

Este modelo de negócio baseado em assinaturas tem sido utilizado por editoras científicas há centenas de anos. É um modelo que tem servido bem tanto à ciência quanto à sociedade. Mas a internet possibilita a transição para um novo modelo de ***acesso aberto*** a artigos científicos, onde esses artigos podem ser baixados gratuitamente. Isso faz parte da mudança que vimos no último capítulo.

Com todo o conhecimento científico mundial gradualmente se tornando acessível online. Uma ressalva a essa história, porém, é que, atualmente, grande parte do conhecimento só é acessível ***se você for um cientista.*** Em particular, cientistas frequentemente trabalham em universidades que têm assinaturas em massa de milhares de periódicos científicos. Um cientista pode baixar gratuitamente quantos artigos sobre câncer de mama ou qualquer outro assunto desejar, enquanto outras pessoas são impedidas de acessar por causa das taxas. É como se houvesse um muro dividindo a humanidade. De um lado do muro estão mais de 99% dos seres humanos que já viveram. E do outro lado do muro está o conhecimento científico mundial. O ***movimento de acesso aberto*** está tentando derrubar esse muro.

Assim como a ciência cidadã está mudando quem pode ser um cientista, o movimento de acesso aberto está mudando quem tem acesso aos resultados da ciência.

Um dos sucessos notáveis do movimento de acesso aberto é um site popular conhecido como preprint de física arXiv (pronuncia-se "archive"). Um "preprint" é um artigo científico, geralmente em fase final de rascunho, pronto para ser considerado por um periódico científico para publicação, mas ainda não publicado em um periódico. Você pode acessar o arXiv agora mesmo e encontrará centenas de milhares de preprints atualizados de físicos do mundo todo, todos disponíveis para download gratuito. Quer saber o que Stephen Hawking anda pensando atualmente? Acesse o arXiv, pesquise "Hawking" e você poderá ler seu artigo mais recente — não algo que ele escreveu alguns anos ou décadas atrás, mas o artigo que ele concluiu ontem, na semana passada ou no mês passado. Quer saber as últimas novidades sobre a busca por partículas fundamentais da natureza no Grande Colisor de Hádrons (LHC)? Acesse o arXiv, pesquise "LHC" e você encontrará uma pilha de artigos de tirar o fôlego. Se você se diverte surpreendendo as pessoas, isso pode render uma conversa inusitada em um coquetel: "Então, você viu as últimas novidades sobre a busca do LHC pela partícula de Higgs? Acontece que..." Claro, nem tudo é fácil de ler. Muitos artigos são escritos por físicos para físicos e podem ser extremamente técnicos. Mas mesmo os artigos mais técnicos costumam ter detalhes intrigantes que são acessíveis a leigos.

O site arXiv funciona assim. Quando um físico conclui seu artigo mais recente, ele acessa o site do arXiv e o envia. Os moderadores do arXiv fazem uma verificação rápida para remover submissões inadequadas — você não verá anúncios de Viagra nem muitos artigos claramente malucos. Poucas horas depois, o artigo aparece no site, onde pode ser baixado e lido por qualquer pessoa no mundo. Muitos físicos enviam seus artigos para o arXiv assim que são concluídos, e muito antes de serem publicados em um periódico científico convencional. Mais da metade de todos os artigos em física aparecem no arXiv, e em algumas subáreas da física essa fração é...

Quase 100%. Muitos físicos começam seu dia de trabalho consultando o arXiv para ver o que apareceu durante a noite. Ele revolucionou a física, acelerando a velocidade com que as descobertas científicas podem ser compartilhadas. Ao mesmo tempo, o arXiv tornou grande parte do conhecimento da humanidade sobre física livremente acessível a qualquer pessoa com conexão à internet. Independentemente de você ter ou não interesse pessoal em física, é de grande benefício para a sociedade que esse conhecimento esteja disponível gratuitamente para empreendedores e engenheiros, jornalistas e estudantes, e para muitos outros que podem se beneficiar, mas que antes estavam excluídos.

O arXiv é um dos grandes sucessos do movimento de acesso aberto. Mas na maioria dos campos da ciência, como medicina, ciência climática e meio ambiente, o conhecimento científico da humanidade ainda é acessível, em grande parte, apenas aos cientistas e a qualquer pessoa que possa pagar pelo acesso. Por isso, e inspiradas em parte pelo sucesso do arXiv, diversas organizações estão criando soluções de acesso aberto para outras áreas além da física. Um exemplo é a Biblioteca Pública de Ciências, ou PLoS. Fundada em 2000, a PLoS se assemelha, em muitos aspectos, a uma editora de periódicos tradicional do que ao arXiv. Mas, em vez de cobrar dos leitores pelo acesso aos artigos, a PLoS cobra dos autores pela publicação de seus artigos.

Essa taxa financia as operações da PLoS, possibilitando que seus artigos sejam disponibilizados gratuitamente na internet. Usando esse modelo, a PLoS rapidamente construiu periódicos considerados entre os melhores em suas áreas, como a PLoS Biology e a PLoS Medicine.

O arXiv e o PLoS são apenas dois dos muitos esforços que visam tornar o acesso aberto à literatura científica a norma. Muitos outros projetos de acesso aberto foram lançados. Esses projetos vêm ganhando força e, em 2008, o Congresso dos EUA sancionou a Política de Acesso Público dos Institutos Nacionais de Saúde (NIH). A política do NIH exige que qualquer pessoa financiada pelo NIH envie seus artigos concluídos para um arquivo de acesso aberto dentro de 12 meses após a publicação em um periódico convencional. Com um orçamento de mais de 30 bilhões de dólares por ano, o NIH é a maior agência de fomento científico do mundo e, portanto, essa política está aumentando rapidamente a quantidade de pesquisas de acesso aberto. Muitas outras agências de fomento e universidades ao redor do mundo estão implementando políticas de acesso aberto semelhantes. Por exemplo, todos os Conselhos de Pesquisa do Reino Unido agora têm políticas semelhantes às do NIH, exigindo que os pesquisadores disponibilizem seus artigos abertamente. Embora muitas pesquisas científicas ainda permaneçam bloqueadas por paywalls de editoras, podemos estar à beira de uma grande mudança em direção ao acesso aberto como a norma, não a exceção. Se isso acontecer, as pessoas nas próximas déca

Houve uma época em que não tínhamos acesso universal à ciência. Será uma mudança institucional semelhante à introdução do mercado.

O benefício mais óbvio do acesso aberto generalizado é para os cidadãos: não há mais restrições à capacidade de pessoas com doenças de baixarem as pesquisas mais recentes! Mas, a longo prazo, um benefício ainda maior do acesso aberto será permitir a criação de outras instituições que conectem a ciência ao restante da sociedade. Já estamos começando a ver isso acontecer. Por exemplo, sites de notícias online gerados por usuários, como Digg e Slashdot, rotineiramente vinculam as pesquisas mais recentes no arXiv, na PLoS e em outras fontes de acesso aberto. Esses sites de notícias permitem que pessoas comuns decidam coletivamente quais são as notícias e oferecem um espaço onde podem discuti-las. Frequentemente, o que as pessoas escolhem discutir inclui os artigos mais recentes do arXiv sobre temas como cosmologia e teletransporte quântico, ou os artigos mais recentes da PLoS sobre temas como genética e biologia evolutiva. Quando as pessoas nos sites de notícias publicam links para periódicos pagos, como ***Nature*** e ***Science,*** muitas vezes surgem reclamações, e os usuários às vezes apontam cópias online piratas como uma alternativa. (Isso não é algo que eu aprovo, mas acontece!) De forma semelhante, sites de notícias on-line produzidos profissionalmente, como o ScienceNews, oferecem sua perspectiva sobre as pesquisas mais recentes.

Eles abrangem notícias de acesso aberto e fechado, mas as notícias de acesso aberto geralmente recebem mais atenção simplesmente porque as pessoas podem clicar para ver a pesquisa original. Esses sites oferecem uma janela para a comunidade científica, complementando e expandindo recursos como o arXiv e a PLoS. É claro que o efeito dessas mudanças às vezes é misto. Muitos artigos jornalísticos foram escritos sobre artigos de mérito científico duvidoso que apareceram no arXiv e em outros recursos de acesso aberto. Mas, na medida em que os cientistas têm evidências de boa-fé a seu favor, o acesso aberto é uma plataforma poderosa para a construção de novas instituições para o aprimoramento da sociedade.

As reações das editoras científicas tradicionais, que oferecem serviços pagos por acesso, ao acesso aberto têm variado. Algumas iniciaram seus próprios experimentos com acesso aberto. Mas muitas, incluindo algumas das maiores editoras, sentem-se ameaçadas pelo movimento do acesso aberto. Para elas, arquivos e periódicos de acesso aberto não são concorrentes comuns. Em vez disso, têm o potencial de mudar radicalmente o modelo de negócios da publicação científica. As editoras tradicionais enfrentam uma escolha difícil. Devem adotar o modelo de acesso aberto da PLoS e de periódicos que o seguem? Ou devem permanecer como estão? Devem ir ainda mais longe e lutar contra o acesso aberto, por exemplo, fazendo lobby contra políticas como a do NIH para acesso aberto?

Política? É uma escolha difícil de tomar, pois, se optarem pelo acesso aberto, é possível que isso reduza significativamente as receitas dos periódicos. A menos que essas empresas desenvolvam novas fontes de receita, seus funcionários perderão empregos e os acionistas perderão dinheiro. É difícil encarar isso depois de décadas, e às vezes séculos, de trabalho árduo construindo negócios que serviram bem à sociedade. Mas o melhor interesse da sociedade se afastou daquele antigo modelo de negócios. Não é de se admirar que muitas editoras tradicionais se sintam ameaçadas. A tecnologia disponível pode ter mudado, mas isso não significa que os modelos de negócios mudaram.

Em termos financeiros, há muito em jogo aqui: a publicação científica é um grande negócio. Isso pode ser uma surpresa para você. Certamente, quando se trata de profissões de alto nível, poucas pessoas pensam na publicação de periódicos científicos. CEOs de editoras científicas não costumam aparecer na capa da ***Forbes*** ou ***da Business Week,*** ao lado de magnatas do software ou operadores de fundos de hedge. Mas talvez devessem, porque a publicação científica é incrivelmente lucrativa. A maior editora de periódicos científicos do mundo é a empresa Elsevier. Em 2009, a Elsevier obteve um lucro de 1,1 bilhão de dólares americanos, mais de um terço de sua receita total de 3,2 bilhões de dólares. Como parcela da receita, esse é o tipo de lucro desfrutado por empresas como Google, Microsoft e algumas outras. A Elsevier é tão lucrativa que sua controladora, o Reed Elsevier Group, vendeu recentemente outra grande parte de seus negócios, a editora educacional Harcourt, por quase cinco bilhões de dólares, para ajudar a financiar a expansão do negócio de publicação de periódicos da Elsevier. E embora a Elsevier seja a maior editora científica, muitas outras também se saem incrivelmente bem. Até mesmo algumas sociedades científicas sem fins lucrativos lucram muito publicando periódicos para seus membros, com os lucros subsidiando outras atividades da sociedade. Por exemplo, em 2004, a Sociedade Americana de Química obteve um lucro de cerca de 40 milhões de dólares americanos com seus periódicos e bancos de dados online, com uma receita de 340 milhões de dólares. Isso é muito menos do que a Elsevier, mas lembre-se: esta é uma sociedade sem fins lucrativos!

Com tanto em jogo, não é surpresa que algumas editoras tradicionais de periódicos científicos tenham começado a fazer lobby agressivamente contra o acesso aberto. De acordo com uma reportagem publicada pela ***Nature*** em 2007, uma importante associação comercial de editoras contratou o consultor de relações públicas Eric Dezenhall, um profissional de alto valor, para ajudá-las a enfrentar o movimento do acesso aberto. Dezenhall ganhou a reputação de “pit bull” do mundo das relações públicas, com clientes como Jeffrey Skilling, o ex-chefe da Enron em desgraça, e a ExxonMobil, que contratou a empresa de Dezenhall para ajudar

eles enfrentam o Greenpeace. Dezenhall aconselhou os editores a se concentrarem em mensagens simples como "Acesso público é igual a censura governamental" e sugeriu que tentassem "pintar um quadro de como seria o mundo sem artigos revisados por pares". (Ambas as noções são falsas: acesso aberto não envolve censura, nem significa abrir mão da revisão por pares.) Quando questionado sobre a decisão de contratar Dezenhall, um vice-presidente da associação de editores respondeu: "É comum contratar uma empresa de relações públicas quando você está sob cerco". Pouco depois de receber o conselho de Dezenhall, a associação de editores lançou uma organização chamada PRISM, a Parceria para a Integridade da Pesquisa em Ciência e Medicina. O PRISM iniciou uma iniciativa publicitária argumentando contra políticas de acesso aberto, como a política do NIH, alegando que o acesso aberto ameaçaria "a viabilidade econômica dos periódicos e do sistema independente de revisão por pares" e potencialmente introduziria "viés seletivo no registro científico".

A história Dezenhall-PRISM é apenas uma das muitas escaramuças na batalha entre algumas editoras científicas tradicionais e o movimento de acesso aberto. Por um lado, temos uma situação em que o acesso aberto representa uma ameaça aos lucros e, em última análise, aos empregos tanto das editoras científicas tradicionais quanto das sociedades científicas sem fins lucrativos.
Mas, em contrapartida, há uma oportunidade maravilhosa: como mostram exemplos como o arXiv, a PLoS e a política de acesso aberto do NIH, agora é possível tornar todo o conhecimento científico disponível gratuitamente para toda a humanidade.
E isso trará benefícios surpreendentes, benefícios grandes demais para serem recusados apenas para preservar alguns negócios bem-sucedidos. Como acontece com tanta frequência com a introdução de novas tecnologias, estamos ponderando um grande bem para a sociedade em detrimento de alguns. As editoras tradicionais que lutam contra o acesso aberto devem ter nossa simpatia, mas não nosso apoio.

Blog de ciência

Em abril de 2008, o autor Simon Singh escreveu um artigo no jornal ***Guardian ,*** onde criticou a Associação Britânica de Quiropraxia
(BCA) por afirmar que “os seus membros podem ajudar a tratar crianças com cólicas, problemas de sono e alimentação, infecções de ouvido frequentes, asma e

choro prolongado, mesmo sem a mínima evidência. Esta organização é a face respeitável da profissão quiroprática e, ainda assim, promove alegremente tratamentos falsos." A BCA respondeu processando Singh sob as leis de difamação do Reino Unido, alegando que a eficácia de seus tratamentos era apoiada por uma "abundância de evidências". O caso recebeu grande atenção pública no Reino Unido e, quatorze meses após o artigo de Singh, a BCA divulgou um documento de sete páginas descrevendo evidências da eficácia dos tratamentos quiropráticos.

O que aconteceu em seguida foi inesperado. Quase imediatamente, as evidências divulgadas pela BCA foram investigadas e desmontadas por um grupo ad hoc de blogueiros científicos, agindo por iniciativa própria. Veja como os eventos foram descritos em um artigo no ***The Lawyer*** escrito por Robert Dougans, advogado que representou Singh no caso, e David Allen Green, blogueiro que cobria o caso:

> Em menos de um dia, a credibilidade dessas evidências — e, de fato, a da BCA por elogiá-las — foi destruída. Cerca de uma dúzia de cientistas-blogueiros, incluindo um membro da Royal Society, conseguiram rastrear e avaliar cada um dos artigos científicos citados pela BCA e demonstraram, sem sombra de dúvida, que esses artigos não apoiavam a posição da BCA. Foi um exercício de blog impressionante e devastador, e quando foi formalmente repetido pelo British Medical Journal algumas semanas depois, foi quase uma reflexão tardia. As evidências técnicas de um reclamante em um caso controverso foram simplesmente demolidas — e vistas como demolidas —, mas não pelos meios convencionais de provas periciais contrárias e interrogatórios forenses dispendiosos, mas por blogueiros especialistas.
>
> E não há razão para que blogueiros tão especializados não façam o mesmo em um caso semelhante.

Dougans e Green chamaram o processo de "litígio wiki" e comentaram que sua importância para o caso ia muito além da mera demolição das provas apresentadas pelo BCA. Eles afirmaram que o blog influenciou substancialmente a cobertura do caso na grande mídia e também "forneceu a Singh opiniões variadas e bem fundamentadas em cada etapa, além daquelas de sua equipe jurídica e dos entusiastas da campanha. Singh certamente levou essas opiniões em consideração em sua tomada de decisão". É uma demonstração notável de como um grupo de blogueiros pode causar mudanças na sociedade.

Assim como o acesso aberto e a ciência cidadã, os blogs científicos são uma instituição que está mudando o papel da ciência na sociedade. Não vou falar sobre tudo isso.

Muitos detalhes sobre blogs científicos aqui. O motivo é que, desde que os blogs (em todas as suas formas, não apenas os científicos) começaram na década de 1990, tem havido muita comoção sobre eles — perdi a conta de quantos artigos de revistas e jornais vi dizendo: "Os blogs estão revolucionando a política!"; "E o jornalismo!"; "Não, não estão!"; e assim por diante. Não quero abordar esse assunto já tão conhecido novamente. Mas quero descrever alguns exemplos que dão uma ideia de como os blogs científicos podem estabelecer um novo tipo de relacionamento entre a comunidade científica e a comunidade em geral, complementando e ampliando ideias como o acesso aberto.

Um aspecto notável dos blogs científicos mais lidos é sua popularidade. Pharyngula, um blog administrado pelo biólogo Paul Myers, da Universidade de Minnesota, recebe mais de 100.000 visitas por dia, comparável à circulação de um importante jornal diário em um grande centro metropolitano, como o ***Des Moines Register*** ou o ***Salt Lake Tribune.*** Isso não é ruim para alguém que escreve em seu tempo livre — e muito mais atenção do que todos os jornalistas da grande mídia impressa, exceto os mais famosos, recebem regularmente.

Pharyngula é o blog de ciência mais popular, mas muitos outros blogs de ciência têm milhares ou dezenas de milhares de leitores regulares. Meu voto para o melhor blog do mundo é o blog de Terence Tao, um matemático vencedor da Medalha Fields, radicado na UCLA. (Conhecemos Tao brevemente antes, como um dos participantes do Projeto Polímata.) O blog de Tao contém centenas de posts. Alguns dos posts são leves ("Mecânica quântica e Tomb Raider"), mas a maioria contém matemática altamente técnica. Só para dar uma ideia, os posts incluem "Consequências finitárias do problema do subespaço invariante" e "O princípio da transferência e equações lineares em primos". Embora os títulos pareçam assustadores para quem não é matemático, para matemáticos esses posts são exposições notavelmente claras e perspicazes de tópicos difíceis, muitas vezes contendo muitos insights originais e ponderados. Apesar de sua natureza técnica, mais de 10.000 pessoas leem o blog de Tao. A seção de comentários revela que, embora muitas dessas pessoas sejam matemáticas profissionais, muitas também são estudantes, às vezes em locais remotos. Alguns dos comentaristas têm pouca formação em matemática: são apenas pessoas interessadas que desejam aprender mais e que gostam de ser expostas diretamente ao pensamento de um dos principais cientistas do mundo.

O que devemos fazer com os blogs científicos? Eles vão transformar o mundo? Em sua forma atual, não creio. Em vez disso, a maneira de pensar sobre os blogs científicos é como um prenúncio do que é possível. Ciência

Os blogs mostram, em sua forma incipiente, o que pode acontecer quando você remove as barreiras que separam os cientistas do restante da comunidade e possibilita um fluxo de informações verdadeiramente bidirecional. Um amigo meu que teve a sorte de estudar na Universidade de Princeton me disse certa vez que a melhor coisa de estudar em Princeton não eram as aulas, nem mesmo os colegas que ele conhecia. Em vez disso, era conhecer alguns dos professores extraordinariamente talentosos e perceber que eles eram apenas pessoas — pessoas que às vezes se irritavam com coisas triviais, ou que faziam piadas bobas, ou que cometiam erros estúpidos, ou que enfrentaram grandes desafios na vida e que, de alguma forma, apesar de suas falhas e desafios, muito ocasionalmente conseguiam fazer algo extraordinário.

"Se eles conseguem, eu também consigo" foi a lição mais importante que meu amigo aprendeu.

O importante, então, é que os blogs possibilitem que qualquer pessoa com conexão à internet tenha uma visão informal e rápida da mente de muitos cientistas do mundo. Você pode acessar o blog de Terence Tao e acompanhá-lo em sua luta para ampliar nossa compreensão de algumas das ideias mais profundas da matemática. Não é apenas o conteúdo científico que importa, é a cultura que é revelada, uma maneira particular de ver o mundo. Essa visão de mundo pode assumir muitas formas. No blog do físico experimental Chad Orzel, você pode ler suas explicações excêntricas de física para seu cachorro ou suas discussões sobre explosões em laboratório. O conteúdo varia bastante, mas, à medida que você lê, um padrão começa a tomar forma: você começa a entender pelo menos um pouco sobre como um físico experimental vê o mundo: o que ele acha engraçado, o que ele acha importante, o que ele acha irritante. Você pode não necessariamente concordar com essa visão de mundo, ou compreendê-la completamente, mas ela é interessante e transformadora mesmo assim. A exposição a essa visão de mundo sempre foi possível para quem mora em uma das capitais intelectuais do mundo, como Boston, Cambridge e Paris. Muitos leitores de blogs, sem dúvida, moram nesses centros intelectuais. Mas você também vê rotineiramente comentários no blog de pessoas que moram fora desses centros. Cresci em uma cidade grande (Brisbane), na Austrália. Comparado à maioria da população mundial, tive uma juventude intelectual privilegiada. E, no entanto, a primeira vez na minha vida que ouvi um cientista falando informalmente foi quando eu tinha 16 anos. Mudou minha vida.

Agora, qualquer pessoa com conexão à internet pode acessar a internet e ter uma ideia de como os cientistas pensam e veem o mundo, e talvez até participar da conversa. Quantas vidas isso mudará?

Imaginando novas instituições

Instituições como ciência cidadã, acesso aberto e blogs científicos estão mudando o papel da ciência em nossa sociedade. Hoje, essas instituições são pequenas, mas estão crescendo rapidamente. Embora eventos como o caso Singh e a descoberta do voorwerp por Hanny sejam significativos, seu impacto é minúsculo quando comparado às maiores instituições da sociedade, como a escolaridade obrigatória. Mas a maioria das instituições grandes e importantes começa pequena e inconsequente — pense nas origens humildes do sistema escolar ou do governo democrático. O que importa não é o tamanho absoluto de uma instituição, mas sim seu potencial de crescimento. Instituições são o que acontece quando as pessoas são inspiradas por uma ideia comum, tão inspiradas que coordenam suas ações em busca dessa ideia. Ferramentas online facilitam muito a criação de instituições, amplificando ideias mais rapidamente do que nunca e ajudando a coordenar ações.

Por exemplo, o Galaxy Zoo começou em 2007 com dois caras em um bar, trabalhando com um orçamento de ousadia e imaginação. Três anos depois, envolveu 25 astrônomos profissionais e 200.000 amadores. Ele se expandiu para incluir projetos como o Moon Zoo e o Projeto Solar Storm Watch. Quanto maior será em dez anos? Suponha que o Galaxy Zoo decida solicitar sistematicamente propostas da comunidade astronômica para a análise de conjuntos de dados. Não é um salto muito grande imaginar o Galaxy Zoo se tornando uma instituição crucial para todo o campo da astronomia, e talvez para outros campos também. Que outras novas instituições teremos a ousadia e a imaginação para sonhar?

Que outras novas respostas encontraremos para questões fundamentais sobre o papel da ciência na sociedade?

Reduzindo a Lacuna da Ingenuidade

O lugar mais isolado do mundo é a Ilha de Páscoa. É uma pequena ilha no sudeste do Pacífico, com apenas 25 quilômetros de largura, 3.500 quilômetros a oeste do Chile e 2.100 quilômetros a leste das Ilhas Pitcairn. A ilha foi originalmente colonizada por

Os ilhéus polinésios e sua cultura prosperaram por centenas de anos, com a população crescendo para algo entre 10.000 e 30.000 pessoas. Mas, à medida que a população crescia, os ilhéus consumiam cada vez mais os recursos da ilha e, em algum momento entre os anos 1500 e 1600, sua sociedade entrou em colapso. Quando o descobridor europeu da Ilha de Páscoa, o explorador holandês Jacob Roggeveen, chegou em 1722, encontrou uma ilha desprovida de recursos naturais. Não havia uma única árvore com mais de três metros de altura em qualquer lugar da ilha. Hoje, analisando o pólen da ilha, sabemos que a Ilha de Páscoa era antigamente uma floresta subtropical, com pelo menos 21 espécies de árvores, algumas delas atingindo até 30 metros de altura.
Roggeveen também não encontrou uma única espécie de ave terrestre. Hoje sabemos que pelo menos seis espécies de aves terrestres viviam na ilha. À medida que os pascoenses destruíam seus estoques de alimentos e madeira, começaram a passar fome, e a população despencou, diminuindo talvez 90%. A cultura da Ilha de Páscoa decaiu para a guerra e, eventualmente, para o canibalismo.

O autor Thomas Homer-Dixon cunhou a expressão "lacuna de engenhosidade" para descrever a diferença de dificuldade entre os problemas enfrentados por uma sociedade e sua capacidade de resolvê-los. O que aconteceu com os habitantes da Ilha de Páscoa foi que eles foram dominados pela lacuna de engenhosidade que sua sociedade enfrentava, incapazes de encontrar soluções para os problemas que haviam criado. Essa lacuna de engenhosidade causou o colapso de sua civilização.

A sociedade global moderna enfrenta sua própria lacuna de engenhosidade. Temos problemas como o HIV/AIDS, que reduz a expectativa média de vida nos países africanos mais afetados em 6,5 anos, de 54,8 para 48,3 anos. Temos o problema das armas nucleares, com a Índia e o Paquistão, ambos dotados de armas nucleares, disputando a Caxemira, e as duas novas superpotências mundiais, China e Índia, competindo pela supremacia na Ásia. À medida que a proliferação nuclear continua, o número de conflitos nucleares plausíveis está aumentando rapidamente. Enfrentamos potenciais escassez de petróleo e água e a possibilidade de bioterrorismo no futuro. E, claro, há a ameaça existencial mais conhecida de nosso tempo: as mudanças climáticas causadas pelo homem. Muitos desses são problemas que compreendemos cientificamente. Mas só porque entendemos os problemas e suas soluções em um nível factual não significa que possamos reunir a capacidade coletiva para agir. Falta-nos a engenhosidade institucional necessária para transformar nosso conhecimento em soluções reais. Hoje, as ferramentas online nos oferecem a oportunidade de criar novas instituições para mudar e redefinir a relação entre ciência e sociedade. Espero que essa oportunidade nos ajude a criar uma sociedade mais resiliente e, nas palavras memoráveis de Hassan Masum e Mark Tovey, a preencher a lacuna da engenhosidade.

CAPÍTULO 8

O desafio de fazer ciência ao ar livre

No final de 1609, Galileu Galilei apontou um de seus telescópios recém-construídos para o céu noturno e começou a fazer uma das séries de descobertas mais impressionantes da história da ciência. A primeira grande descoberta de Galileu, feita em janeiro de 1610, foi a das quatro maiores luas de Júpiter. Hoje, essa descoberta talvez pareça banal, mas causou a maior mudança em nossa concepção do universo desde os tempos antigos. A descoberta se tornou uma sensação, e Galileu foi festejado em toda a Europa. Também lhe rendeu o patrocínio de um dos homens mais ricos da Europa, o Grão-Duque da Toscana, Cosimo de Médici.

Com a fama e o patrocínio, veio a pressão para repetir seu sucesso, e Galileu queria mais descobertas que se equiparassem às luas de Júpiter. Ele não teve que esperar muito. Pouco antes do amanhecer da manhã de 25 de julho de 1610, Galileu apontou seu telescópio para Saturno e observou que não se tratava de um único disco redondo, como se pensava até então. Em vez disso, ao lado do disco principal de Saturno, ele viu duas pequenas saliências, uma de cada lado do disco principal, fazendo parecer que Saturno consistia não apenas de um corpo, mas sim de três. Essas duas saliências em cada lado do disco principal foram o primeiro indício dos anéis de Saturno. Infelizmente para Galileu, seu telescópio não era bom o suficiente para resolver os anéis com clareza. Isso teria que esperar pelo cientista holandês Christiaan Huygens, em 1655. Ainda assim, esta foi outra descoberta importante, e Galileu é frequentemente creditado, juntamente com Huygens, como o descobridor dos anéis.

Ansioso para reivindicar o crédito por sua nova descoberta, Galileu imediatamente enviou cartas a vários de seus colegas, incluindo seu grande colega e rival, o astrônomo Johannes Kepler. A carta de Galileu a Kepler (e seus outros colegas) era peculiar. Em vez de explicar diretamente o que havia visto, Galileu explicou que descreveria sua última descoberta na forma de um anagrama:

smaismrmilmepoetaleumibunenugttauiras

Ao enviar este anagrama, Galileu evitou revelar os detalhes de sua descoberta, mas ao mesmo tempo garantiu que, se outra pessoa — como Kepler — posteriormente fizesse a mesma descoberta, Galileu poderia revelá-lo e reivindicar o crédito. Isso lhe deu tempo para que ele próprio pudesse desenvolver a descoberta. Ao mesmo tempo, Galileu também escreveu aos seus patronos, os Médici. Mas nessa carta, ansioso para mantê-los felizes, Galileu revelou todos os detalhes de sua descoberta, pedindo aos Médici que a mantivessem em segredo por enquanto. Essa situação durou pouco mais de três meses, até que, a pedido do patrono de Kepler, o Sacro Imperador Romano Rodolfo II, Galileu cedeu e revelou que o anagrama era o latim "Altissimum planetam tergeminum observavi", significando, grosso modo, que ele havia observado que o mais alto dos planetas (Saturno) tinha a forma de três.

Há uma conclusão divertida nessa história. Após a descoberta das quatro luas de Júpiter por Galileu, Kepler desenvolveu uma teoria de que Marte devia ter duas luas, com base no fato de que a Terra tinha uma lua, Júpiter tinha quatro e Marte era o planeta entre a Terra e Júpiter. Quando Kepler recebeu o anagrama de Galileu sobre Saturno, trabalhou arduamente para decifrá-lo e, finalmente, decodificou-o como "Salve umbistineum geminatum Martia proles", que significa, aproximadamente, "Sejam saudados, dois botões, filhos de Marte". Aha, pensou Kepler, Galileu devia ter visto as duas luas de Marte! Kepler não tinha certeza, porém, porque uma letra do anagrama de Galileu não foi utilizada. Infelizmente para Kepler, a descoberta das duas luas de Marte teve que esperar até 1877, quando telescópios muito mais potentes estavam disponíveis.

A Primeira Revolução da Ciência Aberta

Galileu não foi o único grande cientista da época a usar anagramas para anunciar descobertas. Newton, Huygens e Hooke usaram anagramas ou cifras para propósitos semelhantes. De fato, muitos cientistas da época relutavam em divulgar suas descobertas de qualquer forma. A infame controvérsia Newton-Leibniz sobre quem inventou o cálculo ocorreu em parte porque Newton alegou tê-lo inventado nas décadas de 1660 e 1670, mas só publicou um relato completo de suas descobertas em 1693.

Enquanto isso, Leibniz desenvolveu e publicou sua própria versão do cálculo. Imagine a biologia moderna se a publicação dos pares de bases do genoma humano tivesse sido adiada por 30 anos, ou se os pares de bases tivessem sido anunciados como um anagrama ("AACCGGGT...", digamos, em vez de "CGTCAAGG...")?

Por que Galileu, Newton e outros cientistas pioneiros eram tão reservados? De fato, uma cultura reservada de descobertas era uma resposta natural às condições da época. Muitas vezes, havia pouco ganho pessoal para os cientistas em compartilhar descobertas, e muito a perder. No início de sua carreira, Galileu cometeu o erro de mostrar uma bússola militar que havia inventado a um jovem chamado Baldassare Capra. Baldassare posteriormente reivindicou a descoberta como sua e acusou Galilei de plágio. Galileu levou anos de esforço e despesas consideráveis para recuperar o crédito por sua descoberta, sem mencionar sua reputação. Não é de se admirar que ele fosse tão reservado quanto à questão de Saturno ser "triformado".

Esse comportamento secreto parece peculiar aos nossos olhos modernos. Hoje, quando os cientistas fazem uma descoberta, eles compartilham seus resultados o mais rápida e amplamente possível, publicando-os em uma revista científica.
Para avanços realmente significativos, os cientistas envolvidos podem escrever um artigo e submetê-lo a um periódico em questão de dias. Alguns periódicos científicos oferecem serviços de publicação rápida para artigos importantes, prometendo publicá-los em poucas semanas após a submissão. É claro que a razão pela qual os cientistas de hoje estão tão ansiosos para compartilhar seus resultados é que seu sustento depende disso: quando um cientista se candidata a um emprego, a parte mais importante da candidatura é seu histórico de artigos científicos publicados. A frase "publicar ou perecer" tornou-se um clichê na ciência moderna porque expressa sucintamente um fato fundamental da vida científica.
Cientistas modernos consideram essa conexão entre publicação e sucesso profissional como algo natural, mas em 1610, quando Galileu fez sua série de grandes descobertas, tal conexão não existia. Não poderia existir, porque as primeiras revistas científicas só foram criadas 55 anos depois, em 1665.

O que causou essa mudança de uma cultura fechada e secreta de descobertas para a cultura moderna da ciência, onde os cientistas anseiam por publicar seus melhores resultados o mais rápido possível? O que aconteceu é que os grandes avanços científicos do século XVII motivaram patronos abastados a começar a subsidiar a ciência como profissão. Essa motivação veio, em parte, do benefício público proporcionado pelas descobertas científicas e, também, em parte, do prestígio conferido a líderes (como os Médici) pela associação com tais descobertas. Ambos os motivos seriam melhor atendidos se as descobertas fossem amplamente divulgadas por meio de um meio como o científico.

periódico. Como resultado, os patrocinadores exigiam uma mudança em direção a uma cultura científica na qual o compartilhamento de descobertas fosse recompensado com empregos e prestígio para o descobridor. Essa transformação estava apenas começando na época de Galileu, mas dois séculos após sua morte, a cultura havia mudado tanto que, quando o grande físico do século XIX Michael Faraday foi questionado sobre o segredo de seu sucesso, ele respondeu que poderia ser resumido em três palavras: "Trabalhar. Concluir. Publicar". Naquela época, uma descoberta não publicada em um periódico científico não estava verdadeiramente completa.

A transformação de uma cultura de descoberta fechada e secreta para a cultura mais aberta da ciência moderna foi um dos eventos mais importantes da história. Resultou na ampla adoção e crescimento do sistema de periódicos científicos. Esse sistema, modesto no início, floresceu em um rico corpo de conhecimento compartilhado para nossa civilização, uma memória coletiva de longo prazo que é a base de grande parte do progresso humano. Esse sistema de compartilhamento de conhecimento funcionou tremendamente bem e mudou apenas lentamente ao longo dos últimos 300 anos.

Hoje, como vimos, as ferramentas online apresentam uma nova oportunidade, uma oportunidade para criar uma memória de trabalho coletiva de curto prazo, um espaço comum conversacional para o rápido desenvolvimento colaborativo de ideias. Ao mesmo tempo, essas ferramentas nos dão a oportunidade de ampliar e enriquecer enormemente nossa memória coletiva de longo prazo. Essas são oportunidades tremendamente empolgantes e promissoras. Já vimos como dados abertos de projetos como o Sloan Digital Sky Survey estão lançando as bases para uma rede de dados que mudará a maneira como explicamos o mundo. E vimos como projetos como o Galaxy Zoo, o Foldit e o arXiv estão mudando a relação entre ciência e sociedade. Mas, embora esses exemplos sejam encorajadores, eles estão muito aquém do potencial da ciência em rede. Há um gargalo fundamental que deve ser superado para que esse potencial se concretize. Vislumbramos esse gargalo anteriormente, na relutância demonstrada por alguns cientistas em compartilhar seus dados e na falta de interesse inicial que os cientistas demonstraram pela Wikipédia. Infelizmente, esses não são exemplos isolados, mas sim sintomas de uma resistência mais profundamente enraizada que muitos cientistas têm em trabalhar online. Essa resistência está freando a ciência da mesma forma que a cultura secreta da descoberta inibiu a ciência no século XVII. Para entender a natureza dessa resistência, vamos analisar mais de perto alguns exemplos promissores, mas fracassados, de ferramentas online para cientistas.

Wikis de Ciência

Embora os cientistas relutassem em contribuir para a Wikipédia em seus primórdios, à medida que a Wikipédia cresceu, inspirou vários cientistas a introduzir wikis focados em descobertas científicas. Um exemplo de tal projeto é o qwiki (abreviação de "quantum wiki"), criado em 2005 por John Stockton, então doutorando no Instituto de Tecnologia da Califórnia (Caltech). Ao contrário da Wikipédia, que visa o público em geral, o qwiki era voltado para cientistas profissionais que trabalhavam na área de computação quântica. O objetivo de Stockton para o qwiki era fornecer uma referência única e centralizada descrevendo todas as pesquisas mais recentes em computação quântica e áreas relacionadas, uma espécie de superlivro-texto em rápida evolução e constantemente atualizado. Mas o qwiki tinha o potencial de ir muito além de um livro-texto: seria infinitamente extensível e modificável, capaz de transmitir material que variava de simples introduções de conceitos-chave até explicações detalhadas dos resultados de pesquisas mais recentes e indicações para problemas não resolvidos na fronteira da pesquisa. Poderia incluir animações e simulações interativas para ilustrar conceitos-chave da computação quântica, bem como materiais de origem para que outras pessoas pudessem aprimorar ainda mais essas animações e simulações. Poderia se tornar um locus para colaboração no estilo Polymath, com teóricos se reunindo para atacar os problemas teóricos mais profundos da computação quântica, em um novo tipo de ciência wiki. Ou experimentalistas poderiam se reunir para compartilhar as melhores práticas, todos os detalhes sutis e difíceis de descrever de experimentos que muitas vezes permanecem como conhecimento tácito, dificultando a reprodução dos resultados de um laboratório para outro. Mesmo que essa visão fosse apenas parcialmente concretizada, o impacto no campo da computação quântica seria extraordinário.

O lançamento do qwiki ocorreu em um workshop do qual participei por acaso, realizado no Caltech em 2005. ch causou bastante rebuliço. Em conversas durante os intervalos do workshop, ouvi algumas pessoas expressarem otimismo de que o qwiki poderia fazer pelo conhecimento especializado em computação quântica o que a Wikipédia e o Google fizeram pelo conhecimento geral. Infelizmente, esse otimismo não se traduziu em uma disposição dessas pessoas em contribuir. Em vez disso, elas esperavam que outra pessoa assumisse a liderança. Afinal, por que contribuir para o qwiki quando você poderia estar fazendo algo mais útil para sua própria carreira, como escrever um artigo ou obter uma bolsa? Por que compartilhar suas melhores e mais recentes ideias no qwiki, quando isso só ajudaria seus concorrentes? E por que contribuir para o qwiki quando ele ainda

seus estágios iniciais e ainda não estava claro se floresceria?
A única parte do qwiki que realmente prosperou foram as "Páginas de Pesquisadores", páginas personalizadas onde cientistas individuais podiam adicionar descrições de si mesmos e de seus trabalhos. Muitos cientistas se contentavam em dedicar uma ou duas horas (e, em alguns casos, mais) para desenvolver essas páginas personalizadas. Mas poucos estavam dispostos a gastar dez minutos adicionando material a outras partes do qwiki. Simplesmente não era uma prioridade. O resultado é que hoje, seis anos após seu lançamento, o qwiki fracassou. Apenas algumas páginas do qwiki são atualizadas com alguma regularidade. Spammers vagam pelo site, adicionando links para produtos duvidosos. Quase todo o conteúdo científico do site foi colocado lá pelo próprio Stockton, por pessoas que trabalhavam no mesmo laboratório ou pelo sucessor de Stockton como mantenedor do qwiki, o estudante de pós-graduação da Universidade de Stanford, Anthony Miller. Esse fracasso não se deveu a nenhuma falta de entusiasmo ou capacidade por parte de Stockton ou Miller. Eles trabalharam duro, adicionando grandes quantidades de material excelente ao qwiki e incentivando outros a ajudar. Infelizmente, embora muitos cientistas acreditassem que tal site tinha o potencial de ser um recurso tremendo, poucos estavam dispostos a contribuir com conteúdo.

A mentalidade por trás do fracasso do qwiki é semelhante à que descrevi no capítulo inicial deste livro, a mentalidade que torna os cientistas relutantes em compartilhar seus dados ou contribuir para a Wikipédia. Na raiz do problema está a intensidade monomaníaca que cientistas ambiciosos devem trazer para a busca por publicações e bolsas científicas.
Para os jovens cientistas, especialmente, essa é uma intensidade gerada pela competição acirrada por empregos científicos. Por exemplo, a cada ano, 1.300 pessoas obtêm doutorados em física em universidades dos EUA, mas apenas 300 vagas para professores de física são abertas. Ao mesmo tempo, muitos programas de doutorado incutem nos jovens cientistas a ideia de que "sucesso" significa conseguir uma posição de professor em uma universidade voltada para a pesquisa, e qualquer outra coisa é um fracasso. O resultado é um enorme impasse de cientistas tentando desesperadamente conseguir posições de professor. Como um jovem cientista, você não está apenas competindo com os outros 1.299 recém-doutores, você também está competindo com pessoas de anos anteriores que ainda estão tentando conseguir empregos de professor. Como resultado, muitos jovens cientistas experimentam grande e prolongada angústia por não conseguirem um emprego de professor. Mesmo em universidades de nível médio, uma vaga de emprego pode facilmente atrair mais de 100 candidatos. Em um ambiente tão competitivo, semanas de trabalho de mais de 80 horas são comuns, e o máximo de tempo possível é dedicado ao objetivo que resultará em uma vaga em uma universidade de ponta: um histórico impressionante de artigos científicos. Os artigos também trazem as bolsas de pesquisa e cartas de recomendação ne

ser contratado. Cientistas que já ocupam cargos efetivos continuam a precisar de apoio financeiro, o que exige uma forte ética de trabalho, ainda focada na produção de artigos. Diante de tudo isso, como um cientista poderia ter tempo para contribuir com iniciativas como o qwiki? Eles podem concordar, em princípio, que gostariam que o qwiki fosse bem-sucedido, mas, na prática, estão ocupados demais escrevendo artigos e propostas de financiamento para ter tempo de contribuir.

O qwiki é apenas um dos muitos wikis científicos que foram lançados. Esforços semelhantes foram feitos para desenvolver wikis para genética, teoria das cordas, química e muitos outros assuntos. Assim como o qwiki, muitos desses wikis científicos tinham grande potencial, e alguns geraram considerável repercussão e otimismo em suas áreas. Mas a maioria não conseguiu decolar, afundando devido à falta de tempo e motivação dos cientistas para contribuir.

Os wikis científicos que dão certo geralmente desempenham um papel de apoio a algum projeto mais convencional. Muitos laboratórios, por exemplo, mantêm wikis internos como forma de armazenar materiais de referência para seus experimentos. Outro wiki de sucesso vem do Projeto Polymath, que usa seu wiki como um local para destilar os insights mais valiosos da colaboração com o Polymath. O wiki do Polymath atraiu milhares de edições, por mais de 100 usuários, e nos horários de pico atrai dezenas de edições e milhares de visualizações de página por dia. (Observe que eu criei o wiki do Polymath e não sou um juiz independente de seu sucesso.)

Mais uma vez, porém, o wiki Polymath apoia um objetivo convencional: resolver um problema matemático e escrever um artigo. Em cada um desses casos, o wiki não foi um fim em si mesmo. A wikiciência, por mais promissora que seja, continua sendo um sonho.

Sites de comentários contribuídos por usuários para a ciência

Não são apenas os wikis científicos que estão fracassando. Diversas organizações criaram sites de comentários com contribuições de usuários, onde cientistas podem compartilhar suas opiniões sobre artigos científicos e, assim, ajudar outros cientistas a decidir quais artigos valem a pena ler e quais não valem o esforço. A ideia é semelhante a sites como o Amazon.com, que coletam avaliações de clientes sobre livros, aparelhos eletrônicos e outros produtos. Como qualquer pessoa que já usou o Amazon.com sabe, as avaliações podem ser muito úteis quando

decidir se compra ou não um produto. Talvez algo semelhante seja útil para cientistas?

O site de comentários com contribuições de usuários de maior visibilidade foi criado por uma das editoras científicas mais prestigiadas, ***a Nature.*** Em 2006, ***a Nature*** lançou um site onde cientistas podiam escrever comentários abertos sobre artigos submetidos à ***Nature.*** Apesar de muito esforço e publicidade, o teste não foi um sucesso. O relatório final que encerrou o teste explicou:

> Houve um nível significativo de interesse expresso na revisão aberta por pares. . . . Uma pequena maioria dos autores que participaram [do ensaio] recebeu comentários, mas geralmente muito poucos, apesar do tráfego significativo na internet. A maioria dos comentários não era tecnicamente substancial. O feedback sugere que há uma relutância acentuada entre os pesquisadores em oferecer comentários abertos.

Em outras palavras, enquanto muitas pessoas queriam ler comentários sobre artigos de outras pessoas, quase ninguém queria realmente escrever comentários.

O teste ***da Nature*** é apenas uma das muitas tentativas de criar sites de comentários científicos com contribuições de usuários. A física, em particular, já viu muitos desses sites, talvez por ter sido a primeira área a adotar amplamente a web como forma de distribuição de artigos científicos. A primeira tentativa foi o site Quick Reviews, que entrou no ar em 1997 e foi descontinuado por falta de uso em 1998. Um site semelhante, Physics Comments, foi criado alguns anos depois, mas sofreu o mesmo destino, sendo descontinuado em 2006. Um site ainda mais recente, Science Advisor, ainda está ativo, mas tem mais membros (1.240) do que revisões (1.119) no momento em que escrevo. Parece que muitos cientistas querem ler comentários em artigos científicos, mas muito poucos querem se voluntariar para escrever tais comentários.

Por que os sites de comentários com contribuições de usuários estão fracassando? Em princípio, a maioria dos cientistas concorda que seria extremamente útil se comentários criteriosos sobre artigos científicos estivessem amplamente disponíveis. Mas, se isso for verdade, parece um enigma que esses sites — muitos deles bem projetados e com bom suporte — fracassem, enquanto as seções de comentários em sites como o Amazon.com prosperam. O problema dos sites de comentários científicos é que, embora comentários criteriosos sobre artigos científicos sejam extremamente úteis para ***outros*** cientistas, isso não significa que seja do interesse de cada um escrever comentários. Imagine como as coisas parecem do ponto de vista de um cientista individual considerando comentar em tal site.

Por que escrever um comentário quando você poderia estar fazendo algo mais útil

para você individualmente, como escrever um artigo ou uma proposta de financiamento? Mesmo que você tenha escrito um comentário, provavelmente relutaria em criticar publicamente o artigo de outra pessoa. Afinal, a pessoa que você critica pode ser um parecerista anônimo em posição de sabotar seu próximo artigo ou pedido de financiamento.

O contraste entre o fracasso dos sites de comentários científicos criados por usuários e o sucesso das avaliações da Amazon.com é gritante. Para citar apenas um exemplo, você encontrará mais de 1.500 avaliações de produtos Pokémon na Amazon.com, mais do que o número total de avaliações em todos os sites de comentários científicos que descrevi acima. Você pode objetar que há mais pessoas comprando produtos Pokémon do que cientistas.

É verdade. Mas ainda existem mais de um milhão de cientistas profissionais no mundo, e esses cientistas passam grande parte de suas vidas profissionais formando opiniões sobre artigos escritos por outros, muito mais tempo do que até mesmo os pais mais entusiasmados conseguem dedicar a Pokémon. É uma situação ridícula: a cultura popular é aberta o suficiente para que as pessoas sintam o desejo de escrever resenhas sobre Pokémon, mas a cultura científica é tão fechada que os cientistas não compartilham publicamente suas opiniões sobre artigos científicos de forma análoga. Algumas pessoas acham esse contraste curioso ou divertido; acredito que isso signifique algo seriamente errado com a ciência.

O Desafio Moderno da Ciência Aberta

O fracasso dos wikis científicos e dos sites de comentários científicos com contribuições de usuários faz parte de um padrão muito maior. Projetos como o Projeto Polymath, o Galaxy Zoo e o Foldit foram todos muito bem-sucedidos, mas esse sucesso se deve em parte a um conservadorismo fundamental: todos eles visam, em última análise, produzir artigos científicos. Ferramentas como os wikis científicos e os sites de comentários com contribuições de usuários rompem com esse conservadorismo, uma vez que as contribuições para tais sites são fins em si mesmas e não resultam diretamente em artigos científicos. Infelizmente, o resultado é que cientistas com foco na carreira têm pouco incentivo para contribuir com tais sites e, em vez disso, concentram seus esforços em fazer o que é recompensado: escrever artigos. As grandes ideias para ampliar a inteligência coletiva que discutimos na parte 1 têm pouca chance de prosperar quando ideias incrementais como wikis científicos e sites de comentários com contribuições de usuários já estão além do aceitável. Muitas das ferramentas com potencial para...

Mudar e aprimorar drasticamente a forma como a ciência é feita são simplesmente inviáveis. Não é por acaso que tantos dos melhores exemplos de amplificação da inteligência coletiva na Parte 1 vieram de fora da ciência; com muita frequência, os cientistas estão atrasados, não liderando o desenvolvimento de novas ferramentas para a produção de conhecimento. E, embora tenhamos visto alguns projetos impressionantes voltados para a ciência, eles exploram apenas uma pequena fração do panorama de possibilidades. Estamos perdendo uma oportunidade gigantesca.

De fato, mesmo as possibilidades que estão sendo exploradas não estão prosperando como deveriam. Enquanto iniciativas como o Sloan Digital Sky Survey e o Projeto Genoma Humano estão abrindo seus dados para outros cientistas, os dados da grande maioria dos experimentos científicos permanecem fechados. Os cientistas normalmente têm pouco incentivo para divulgar seus dados e, por isso, os acumulam. Nas palavras das pesquisadoras médicas Elizabeth Pisani e Carla AbouZahr, na ciência é "publicar [artigos] ou perecer", não "publicar [dados] ou perecer". E enquanto isso continuar sendo verdade, grande parte do conhecimento científico mundial permanecerá bloqueado, impedindo que a rede de dados científicos atinja seu pleno potencial.

Igualmente preocupantes são os desincentivos para os cientistas desenvolverem novas ferramentas online. Enquanto eu escrevia este livro, um físico renomado me disse que Paul Ginsparg, o físico que criou o arXiv, havia "desperdiçado seu talento" para a física ao criar o arXiv, e que o que Ginsparg estava fazendo era como "coleta de lixo": era bom que alguém estivesse fazendo isso, mas abaixo de alguém com as habilidades de Ginsparg. Tenha em mente que essa espantosa estreiteza de espírito vinha de uma pessoa que usa o arXiv todos os dias. Ginsparg talvez tenha feito mais pela física (para não mencionar o resto da humanidade) do que qualquer outro físico de sua geração.

No entanto, sentimentos como esses são frequentemente expressados em particular pelos cientistas. Pessoas que criam ferramentas como o arXiv são descartadas como "meros" criadores de ferramentas, como se fosse de alguma forma indigno criar ferramentas que aceleram todo o processo de fazer ciência. Essa falta de consideração se estende ao nível institucional, onde frequentemente há pouco apoio para o desenvolvimento de novas ferramentas. Projetos como o Galaxy Zoo e o arXiv frequentemente começam com pouco ou nenhum financiamento, em parte porque sua primeira etapa envolve a criação de uma ferramenta, não a escrita de um artigo. Como ideias como a ciência cidadã e a web de dados podem atingir seu potencial em um ambiente onde o desenvolvimento de novas ferramentas é tão mal visto?

O padrão geral, então, é que a ciência em rede está sendo fortemente inibida por uma cultura científica fechada que valoriza principalmente contribuições na forma de artigos científicos. O conhecimento compartilhado em mídias não convencionais não é valorizado pelos cientistas, independentemente de seu valor científico intrínseco, e

Por isso, os cientistas relutam em trabalhar nesses meios. O potencial da ciência em rede — ideias como a rede de dados, a ciência cidadã e os mercados de colaboração — permanece, portanto, inexplorado. Para atingir todo o seu potencial, a ciência em rede deve ser ciência aberta.

A ironia em tudo isso é que o valor do compartilhamento aberto de informações científicas foi profundamente compreendido pelos fundadores da ciência moderna há séculos. Foi essa compreensão que levou ao sistema moderno de periódicos, um sistema que talvez seja o mais aberto para a transmissão de conhecimento que poderia ser construído com a mídia do século XVII. A adoção desse sistema foi alcançada subsidiando cientistas que publicavam suas descobertas em periódicos. Mas esse mesmo subsídio agora ***inibe*** a adoção de tecnologias mais eficazes, porque continua a incentivar os cientistas a compartilhar seus trabalhos em periódicos convencionais, enquanto há pouco ou nenhum incentivo para que usem ou desenvolvam ferramentas modernas. De fato, quando os cientistas de hoje resistem a compartilhar seus dados e ideias, eles estão inconscientemente ecoando o comportamento de Galileu, Newton e companhia, com seu sigilo e seus anagramas.

Pode ser uma resposta prática a preocupações pessoais imediatas, mas, a longo prazo, é a maneira errada de fazer ciência.

Para aproveitar ao máximo as ferramentas modernas de produção de conhecimento, precisamos criar uma cultura científica aberta, onde o máximo de informação possível seja transferido da cabeça e dos laboratórios das pessoas para a rede. Isso não significa apenas as informações convencionalmente compartilhadas em artigos científicos, mas ***todas*** as informações de valor científico, desde dados experimentais brutos e códigos de computador até todas as questões, ideias, conhecimento popular e especulações que atualmente estão trancadas na cabeça de cientistas individuais. Informações que não estejam na rede não podem ser úteis.

Em um mundo ideal, alcançaríamos um tipo de ***abertura extrema.*** Isso significa expressar todo o nosso conhecimento científico em formas que sejam não apenas legíveis por humanos, mas também por máquinas, como parte de uma rede de dados, para que os computadores possam nos ajudar a encontrar significado em nosso conhecimento coletivo. Significa abrir a comunidade científica para o resto da sociedade, em uma troca bidirecional de informações e ideias. Significa uma ética de compartilhamento, na qual todas as informações de valor científico são colocadas na rede. E significa permitir uma reutilização e modificação mais criativas do trabalho existente. Essa abertura extrema é a expressão máxima da ideia de que outros devem ser capazes de desenvolver e ampliar o trabalho de cientistas individuais, talvez de maneiras que eles próprios jamais teriam concebido. Na prática, será necessário haver alguns limites — pense em

Preocupações como a confidencialidade do paciente na pesquisa médica — e discutiremos esses limites no próximo capítulo. Mas mesmo dentro desses limites, a abertura que defendo representaria uma mudança cultural gigantesca na forma como a ciência é feita, uma segunda revolução da ciência aberta, estendendo e completando a primeira revolução da ciência aberta, dos séculos XVII e XVIII.
No próximo capítulo, discutiremos como essa cultura mais aberta pode ser alcançada.

Um aparte sobre comercialização e sigilo em Ciência

Neste capítulo, vimos como o forte compromisso dos cientistas com os artigos científicos como expressão máxima da descoberta científica está inibindo novas e melhores maneiras de fazer ciência. Mas, para alguns cientistas, há uma inibição adicional: a necessidade de sigilo, pois buscam patentes e derivações comerciais de seus trabalhos. Por exemplo, de 2001 a 2003, fiz parte de um grande centro de pesquisa que trabalhava com computação quântica. Embora o centro estivesse longe de produzir um produto comercial, seus líderes esperavam que um dia houvesse tais derivações. Quando os cientistas participavam de seminários de pesquisa no centro, eram (por um tempo) solicitados a assinar acordos de confidencialidade, prometendo não falar com outras pessoas sobre o conteúdo dos seminários. Muitos cientistas do centro documentavam meticulosamente seus trabalhos em cadernos, onde cada página era datada e assinada por funcionários do centro, para ajudar a estabelecer a prioridade em caso de pedidos de patente posteriores. Esse sigilo pode ajudar a levar ao sucesso comercial. Mas é impossível que tal cultura coexista com a atmosfera aberta e colaborativa vista, por exemplo, no Projeto Polymath, ou que é necessária para o sucesso do wiki.

Esse sigilo comercial é relativamente novo em nossas universidades, onde a maior parte da pesquisa básica é realizada. De fato, até bem recentemente, as universidades concentravam a maior parte de seus esforços de pesquisa científica em pesquisa básica, sem aplicação comercial imediata. Isso mudou nas últimas décadas, em grande parte devido a uma lei chamada Lei Bayh-Dole, aprovada pelo Congresso dos EUA em 1980. O que

O que a Lei Bayh-Dole fez foi conceder às universidades americanas (em vez do governo, como era o caso anteriormente) a propriedade de patentes e outras propriedades intelectuais produzidas com o auxílio de subsídios governamentais. Após a aprovação da Lei Bayh-Dole, muitas universidades começaram a ampliar seu foco para além da pesquisa básica, apoiando mais pesquisas aplicadas na esperança de lucrar com spin-offs comerciais. Simultaneamente, e pelo mesmo motivo, houve também um aumento nas patentes relacionadas à pesquisa básica conduzida nas universidades. Muitos outros países seguiram o exemplo dos EUA e aprovaram leis semelhantes à Lei Bayh-Dole, com efeito semelhante em sua cultura de pesquisa. O sucesso desses esforços é questionável — muitas universidades, na verdade, perdem dinheiro tentando comercializar suas pesquisas —, mas o interesse na comercialização e na propriedade intelectual, no entanto, tornou muitos cientistas mais reservados.

Esse sigilo comercialmente motivado representa uma grande mudança cultural em nossas universidades. Historicamente, antes da Lei Bayh-Dole e de leis similares, os resultados da ciência básica eram geralmente (eventualmente) divulgados abertamente, na forma de artigos, na crença de que uma melhor compreensão de como o mundo funciona beneficiaria a todos a longo prazo. Por exemplo, a pesquisa básica sobre eletricidade e magnetismo foi a base para invenções como motores, iluminação elétrica, rádio e televisão.

A pesquisa básica em mecânica quântica foi crucial para a indústria de semicondutores. É a ideia comum de que a maré alta faz todos os barcos flutuarem. E assim houve uma divisão bastante clara em nosso sistema de inovação. De um lado, estava o sistema de pesquisa básica, cujos resultados finais eram compartilhados publicamente como artigos científicos, sob a justificativa de que, a longo prazo, todos ganhariam. Do outro lado, estava a pesquisa aplicada, financiada privadamente, que visava ao desenvolvimento de produtos a curto prazo, e que muitas vezes era realizada em segredo. A Lei Bayh-Dole começou a desmembrar essa divisão, e hoje governos e agências de fomento veem cada vez mais a busca por patentes e outras propriedades intelectuais como um motivo importante para apoiar a pesquisa básica.

Essa mudança é um verdadeiro obstáculo ao compartilhamento aberto necessário para o desenvolvimento da ciência em rede. No entanto, devemos manter a dimensão e o escopo desse obstáculo na devida perspectiva. Ao escrever este livro, conversei algumas vezes com pessoas que presumiam que o sigilo comercial era o maior obstáculo à ciência aberta. Isso é incorreto. Em grande parte da ciência básica, as preocupações dos cientistas com a comercialização são decididamente secundárias em comparação com seu foco incansável na publicação convencional. A comercialização e os direitos de patente são bem-vindos, se vierem, mas o sucesso profissional vem com a conquista do

estima dos pares por meio da publicação. Isso é mais evidente em candidaturas a empregos: cientistas frequentemente listam algumas patentes ou spin-offs resultantes de seu trabalho, mas a ênfase está em artigos, artigos, artigos e bolsas, bolsas, bolsas. Isso se aplica a grande parte da física e astronomia, à matemática, a partes substanciais da química, biologia e a muitos outros campos da ciência. Nesses campos, o obstáculo imediato à ciência aberta não é a comercialização, mas sim uma cultura que valoriza e recompensa apenas o compartilhamento do conhecimento científico na forma de artigos.

Em algumas áreas da ciência básica, o sigilo comercial é primordial. Isso se aplica a alguns dos trabalhos em estágio inicial que podem levar ao desenvolvimento posterior de medicamentos, por exemplo. Nesses campos, a ciência provavelmente permanecerá um assunto fechado e secreto. E há uma área cinzenta muito maior na ciência básica, onde as preocupações com o sigilo comercial são um fator, mas nem sempre um fator dominante. O verdadeiro problema é o trabalho científico que, em princípio, poderia ser aberto, mas onde esperanças infundadas de patentes posteriores impedem a ciência aberta. A longo prazo, há uma discussão a ser travada sobre o papel da propriedade intelectual na ciência básica. Mas a base para a ciência aberta, o ponto de partida, é uma mudança na cultura da ciência para que ela não valorize e recompense apenas a escrita de artigos, mas também novas formas de compartilhamento. Esse é o problema mais crucial, e é para ele que nos voltamos agora.

CAPÍTULO 9

O Imperativo da Ciência Aberta

Imagine que você é um cientista ativo que acredita firmemente que a ciência aberta trará enormes benefícios à ciência e à nossa sociedade.
Você entende que mudar a cultura científica profundamente arraigada será difícil, mas decide, mesmo assim, compartilhar suas ideias e dados online, contribuindo para novas ferramentas, como wikis científicos e sites de comentários de usuários, e disponibilizando gratuitamente o código dos seus programas de computador. Tudo isso exige muito tempo e esforço, e, ainda assim, você descobre que, sem colegas dispostos a retribuir, os benefícios são pequenos. Isso porque muitos dos benefícios da ciência aberta só vêm se ela for adotada coletivamente por um grande número de cientistas. E, como apenas um cientista, você não pode obrigar todos os outros a fazer ciência aberta.

Uma experiência típica é a do meu colega e ex-aluno Tobias Osborne, agora da Universidade de Hannover, na Alemanha. Ansioso por experimentar a ciência aberta, Osborne realizou grande parte de sua pesquisa sobre computação quântica abertamente, por seis meses, em um blog. Ele escreveu muitos posts reflexivos, repletos de ideias perspicazes, e seu blog atraiu seguidores na comunidade de computação quântica, com mais de 50 leitores regulares. Infelizmente, poucos desses leitores estavam dispostos a dar feedback sobre os posts de Osborne ou a compartilhar suas próprias ideias.
E sem uma comunidade de colegas engajados, trabalhar abertamente daria muito trabalho, com um retorno pequeno. Osborne concluiu, por fim, que a ciência aberta não teria sucesso porque "exigiria que a maioria dos cientistas mudasse seu comportamento simultânea e completamente".
Experiências como essa fazem a ciência aberta parecer uma causa perdida.
Embora seja verdade que será difícil avançar em direção à ciência aberta por meio da ação direta de cientistas individuais, isso não significa que outras abordagens não possam ter sucesso. Nossa sociedade resolveu muitos problemas análogos ao problema da ciência aberta, problemas em que a ação individual direta não funciona, e os benefícios só surgem se muitas pessoas em um grande grupo adotarem simultaneamente uma nova maneira de fazer as coisas. Um exemp

O problema de qual lado da estrada dirigir. Se você mora em um país onde as pessoas dirigem do lado esquerdo, não pode um dia começar um movimento para dirigir do lado direito simplesmente trocando o lado em que você dirige.
Mas isso não significa que não seja possível que todos mudem simultaneamente. Foi exatamente o que aconteceu na Suécia em 3 de setembro de 1967, às 5 da manhã. Havia bons motivos para os suecos mudarem: as pessoas nos países vizinhos já dirigiam à direita e, além disso, a maioria dos veículos na Suécia já tinha volante à esquerda, tornando dirigir à direita realmente mais seguro. Mas, assim como acontece com a ciência aberta, o simples fato de que dirigir à direita seria melhor não foi suficiente para causar uma mudança por meio de ação direta de indivíduos. Em vez disso, foi necessária uma campanha prolongada do governo e uma mudança na lei.

Mudar de lado parece muito distante de mudar a cultura da ciência. Mas, na verdade, a primeira revolução da ciência aberta exigiu um tipo semelhante de ação coletiva. Vimos como os cientistas do século XVII frequentemente guardavam seus resultados para si mesmos — a menos que o envio de anagramas seja considerado compartilhamento! Quando o sistema de periódicos científicos foi introduzido pela primeira vez, muitos cientistas ficaram desconfiados, relutantes em compartilhar seus resultados com outros em um novo meio. Embora os cientistas, individualmente, pudessem ver que a ciência como um todo progrediria mais rapidamente se ***todos*** compartilhassem livremente as notícias de suas descobertas, isso não significava que fosse do seu interesse individual publicar nos periódicos. Isso representou um problema para os editores dos primeiros periódicos, pessoas como Henry Oldenburg, que fundou o primeiro periódico científico do mundo, o ***Philosophical Transactions of the Royal Society,*** em 1665. A biógrafa de Oldenburg, Mary Boas Hall, conta como Oldenburg escrevia aos cientistas da época e "implorava por informações", às vezes escrevendo simultaneamente para dois cientistas concorrentes, alegando que seria "melhor dizer a A o que B estava fazendo e vice-versa, na esperança de estimular ambos a trabalhar mais e a ser mais abertos". Dessa forma, Oldenburg incitou alguns dos cientistas mais eminentes de sua época, incluindo Newton, Huygens e Hooke, a publicar no ***Philosophical Transactions.*** A necessidade de tal subterfúgio só cessou após décadas de trabalho de Oldenburg e outros para mudar a cultura da ciência.

O padrão comum subjacente ao problema da troca de lado da estrada e ao problema da ciência aberta — tanto hoje quanto no século XVII — é que os interesses dos indivíduos não estão naturalmente alinhados com o interesse coletivo do grupo. Alguém que acredita que "todos deveriam fazer isso", por exemplo, ciência aberta ou a troca de lado da estrada, não necessariamente acredita também que "eu deveria fazer isso, mesmo que ninguém".

outra pessoa faz.” Cientistas sociais chamam problemas como esse de problemas de ação coletiva. O truque para resolver problemas de ação coletiva é descobrir maneiras de alinhar o interesse individual com o interesse coletivo. No caso da mudança da Suécia para a direita, a solução foi usar a legitimidade do governo — expressa, em parte, pela força da lei — para obrigar as pessoas a mudar. Um dia, era do interesse individual das pessoas dirigir à esquerda, enquanto no dia seguinte era do interesse delas dirigir à direita. Da mesma forma, a genialidade da primeira revolução da ciência aberta foi alinhar os interesses individuais e coletivos, recompensando os cientistas por compartilharem suas descobertas em periódicos científicos. O problema hoje é que, embora agora seja do interesse coletivo que os cientistas adotem novas tecnologias, seus interesses individuais permanecem alinhados com a publicação em periódicos. Precisamos alinhar novamente o interesse individual com o interesse coletivo.

A boa notícia é que muito se sabe sobre como resolver problemas de ação coletiva. Escrevendo na década de 1960, o economista político Mancur Olson analisou o que fundamentava a "lógica da ação coletiva", tentando compreender as condições sob as quais os indivíduos de um grupo trabalharão juntos em prol de seu interesse coletivo, e aquelas sob as quais não o farão.
Na década de 1990, a economista política Elinor Ostrom aprofundou substancialmente a análise de Olson para um tipo específico de ação coletiva, a saber, como grupos podem trabalhar juntos para gerir recursos que possuem em comum, como água e florestas. Os livros que Olson e Ostrom escreveram descrevendo seu trabalho estão entre os mais citados de todos os tempos nas ciências sociais, e a obra foi tão influente que Ostrom recebeu o Prêmio Nobel de Economia de 2009.

Menciono este trabalho como um antídoto ao pessimismo sobre a ciência aberta. Algumas pessoas muito inteligentes dedicaram muito tempo à investigação de exemplos reais em que problemas de ação coletiva foram resolvidos e refletiram profundamente sobre como as estratégias utilizadas nesses exemplos podem ser generalizadas para resolver outros problemas de ação coletiva. O que Olson, Ostrom e seus colegas demonstraram é que, embora resolver problemas de ação coletiva seja difícil, não é impossível. Antes de desistirmos da ciência aberta, devemos nos basear nessas ideias. Veremos agora duas estratégias que podem ser usadas para mudar a cultura científica. Nenhuma delas é uma solução rápida, mas com imaginação e determinação suficientes, essas estratégias podem tornar a ciência muito mais aberta. Embora meu relato seja baseado no trabalho de Olson e Ostrom e seus sucessores, não farei as conexões explicitamente, visto que este não é um livro didático sobre economia política. Se você estiver interessado em explorar as conexões mais a fundo, por favor.

veja “Fontes selecionadas e sugestões para leitura adicional”, começando na página 217.

Ciência Aberta Atraente

Anteriormente neste livro, discutimos as políticas de acesso aberto que algumas agências de fomento científico estão implementando para tornar os resultados de pesquisas científicas amplamente disponíveis. Lembre-se, por exemplo, que os Institutos Nacionais de Saúde (NIH) dos EUA agora exigem que os cientistas tornem seus artigos publicamente acessíveis em até 12 meses após a publicação. Cientistas que não concordarem com essa condição devem buscar financiamento em outras fontes. É uma política de compulsão, semelhante à estratégia usada pelo governo sueco de mudar de lado da estrada. Dessa forma, organizações poderosas, como governos e agências de fomento, podem fazer com que todos em uma comunidade mudem seu comportamento simultaneamente.

Dando continuidade às suas políticas de acesso aberto, diversas agências de fomento agora exigem que os cientistas compartilhem abertamente seus ***dados.*** Essas políticas de dados abertos seguem o espírito do Acordo das Bermudas para compartilhamento de dados genéticos humanos (ver página 7), mas têm um escopo mais amplo. Há muitas maneiras pelas quais isso está acontecendo; então, deixe-me descrever apenas alguns exemplos. Em áreas com foco específico, como a genômica, as políticas podem ser bastante exigentes. Anteriormente neste livro, na página 3, vimos como a genômica pode ser usada para descobrir ligações entre genes e doenças; os estudos resultantes são chamados de estudos de associação genômica ampla (GWAS). Em 2007, o NIH instituiu uma política que exige que os dados dos GWAS sejam disponibilizados abertamente, sujeitos a algumas restrições para garantir a privacidade dos participantes.
Outro grande financiador da pesquisa genômica, o Wellcome Trust, agora exige que todos os dados genéticos sejam disponibilizados abertamente, novamente sujeitos a questões de privacidade e preocupações semelhantes. Além disso, essas agências especificam em quais bancos de dados online os dados devem ser enviados, em quais formatos e assim por diante.
sobre.

Políticas mais amplas sobre compartilhamento de dados geralmente são menos específicas. Por exemplo, desde 2006, o Conselho de Pesquisa Médica do Reino Unido exige que todos os cientistas que financia disponibilizem seus dados abertamente, desde que isso não viole nenhuma norma ética ou legal. Mas essa política não especifica exatamente como ou onde os dados devem ser disponibilizados. Muitas políticas de dados abertos

ainda estão em estágios iniciais de desenvolvimento. Por exemplo, desde janeiro de 2011, a Fundação Nacional de Ciências dos EUA exige que os pedidos de financiamento incluam um plano de gerenciamento de dados de duas páginas. Não se trata de uma política de dados abertos completa, mas um porta-voz afirmou que este anúncio foi apenas a "primeira fase" de um esforço para garantir que todos os dados sejam de acesso aberto. Acima de tudo isso, nos mais altos níveis políticos, há uma compreensão crescente do valor dos dados abertos. Por exemplo, em 2007, a Organização para a Cooperação e Desenvolvimento Econômico (OCDE) recomendou que os países-membros tornassem os dados de pesquisas financiadas publicamente acessíveis abertamente. Essas recomendações levam tempo para serem difundidas, mas, com o tempo, podem ter impacto.

Políticas de acesso aberto e dados abertos são passos poderosos em direção à ciência aberta, passos difíceis para cientistas individuais realizarem sozinhos. As agências de fomento são o mecanismo de governança de fato na república da ciência e têm grande poder para impulsionar mudanças, mais poder até do que cientistas superestrelas como os ganhadores do Prêmio Nobel. O comportamento de muitos cientistas é ditado pela regra de ouro: quem tem o ouro faz as regras. E as grandes agências de fomento têm o ouro. Se as pessoas que administram as agências de fomento decidissem que, como parte do processo de concessão de subsídios, os candidatos a subsídios teriam que dançar uma giga no centro da cidade, as ruas do mundo logo estariam cheias de professores dançarinos. Agora, muitas pessoas — incluindo muitos agentes de fomento — criticam esse sistema, acreditando que ele é muito centralizado e controlador.

Mas, na prática, o sistema de bolsas atualmente rege grande parte da ciência, e se as agências de fomento decidirem levar a ciência aberta a sério, os cientistas também o farão. Imagine, por exemplo, que uma das grandes agências de fomento começasse a solicitar aos candidatos que apresentassem evidências de divulgação pública por meio de blogs e vídeos online. Ou suponha que começassem a solicitar aos candidatos que descrevessem suas contribuições para wikis científicos como evidência de atividade de pesquisa. Tais políticas contribuiriam muito para legitimar novas ferramentas.

Embora as agências de fomento possam ajudar novas ferramentas a serem aceitas, elas não têm poder ilimitado para impor a ciência aberta aos cientistas. Lembre-se novamente da história do Acordo das Bermudas para o compartilhamento de dados genéticos humanos. Esses princípios não foram meramente impostos por decreto aos biólogos moleculares por alguma agência central de fomento. Em vez disso, líderes da comunidade de biologia molecular se reuniram nas Bermudas, onde concordaram que seria do interesse de toda a comunidade compartilhar dados. Essencialmente, cientistas individuais estavam dizendo: "Gostaríamos de abrir — mas somente se todos os outros também o fizerem". As agências de fomento então ajudaram a atingir esse objetivo, aplicando a política de abertura. Mas parte da

A razão pela qual a política foi tão eficaz foi porque já contava com o apoio de importantes biólogos moleculares. Uma situação semelhante ocorreu na Suécia, onde a mudança para o lado direito da estrada só foi implementada após décadas de discussão pública sobre a ideia.

Para ter sucesso, as agências de fomento não podem apenas impor abertura, elas também devem gerar consentimento e acordo dentro da comunidade científica. Se não fizerem isso, será muito fácil para os cientistas responderem seguindo à risca os requisitos das agências de fomento, mas não o espírito. Imagine futuros cientistas divulgando conjuntos de dados "abertos" tão mal documentados que se tornam inúteis para qualquer outra pessoa. Uma coisa é um cientista despejar dados brutos online em algum local obscuro. Outra coisa é documentar e calibrar cuidadosamente esses dados, integrá-los aos dados de outros cientistas e incentivá-los ativamente a encontrar novos usos para eles.

É isso que será necessário para que a rede de dados científicos tenha sucesso. De forma mais geral, para que a ciência em rede atinja seu pleno potencial, os cientistas devem se comprometer, com entusiasmo e sinceridade, com novas formas de compartilhar conhecimento. Para que isso aconteça, as agências de fomento devem trabalhar individualmente com as comunidades científicas, conversando longamente com os cientistas de cada comunidade sobre maneiras pelas quais essa comunidade poderia se tornar mais aberta. Existem dados que poderiam ser compartilhados sistematicamente? E quanto ao código de computador? E quanto às perguntas, ideias e sabedoria popular das pessoas? O que mais poderia ser compartilhado? Com que rapidez poderia ser compartilhado? Quais novas ferramentas precisam ser desenvolvidas para tornar isso eficaz? Se as agências de fomento fizerem isso, poderão atuar como catalisadoras de acordos no estilo das Bermudas para compartilhar conhecimento científico. E, tendo forjado tais acordos, poderão então expressá-los em políticas. Este será um trabalho longo e lento, mas a recompensa será uma tremenda mudança cultural em direção a uma maior abertura.

Incentivar a Ciência Aberta

A perspectiva de as agências de fomento dizerem "Trabalharás mais abertamente" me deixa, como cientista, com sentimentos mistos. Embora promova o uso de novas ferramentas, não provocará uma adoção verdadeiramente entusiástica dessas ferramentas pelos cientistas, a menos que também criemos novos incentivos para seu uso. Os cientistas de hoje demonstram um impulso incansável para escrever artigos científicos porque é isso que a comunidade científica valoriza. Precisamos de nova

Incentivos que criem um impulso semelhante para compartilhar dados, códigos e outros conhecimentos. Como podemos tornar o compartilhamento de conhecimento de novas maneiras tão imperativo para os cientistas quanto a publicação de artigos científicos é hoje?

Ajuda analisar essa questão em termos econômicos. Em uma economia convencional, se eu lhe der um sofá em troca de algum dinheiro, você ganha um sofá e eu perco um sofá. Mas descobertas científicas são diferentes. Se eu compartilhar uma descoberta com você, não perco meu conhecimento sobre essa descoberta.

Esse tipo de compartilhamento é ótimo para a sociedade como um todo, mas tem um problema do ponto de vista do descobridor original: se ele não for recompensado, terá muito menos motivos para investir tempo e esforço para chegar à descoberta em primeiro lugar.

A solução para esse problema adotada pela comunidade científica no século XVII (e ainda usada hoje) é brilhante. Em vez de dar às pessoas direitos exclusivos sobre suas ideias, como em uma economia convencional, criamos uma economia baseada na reputação. Cientistas compartilham abertamente suas descobertas publicando-as em artigos científicos — essencialmente, doando-as —, mas em troca recebem o direito de serem creditados como descobridores. Ao serem creditados, eles podem construir uma reputação, que pode se transformar em um emprego remunerado. É um tipo de direito de propriedade sobre ideias, levando a uma economia baseada na reputação e estabelecendo uma mão invisível para a ciência que motiva fortemente os cientistas a compartilhar seus resultados. A base para essa economia da reputação é um conjunto de normas sociais muito fortes: os cientistas devem dar crédito ao trabalho de outras pessoas; não podem plagiar; e os cientistas julgam o trabalho de outros cientistas por seu histórico de artigos publicados. Mas essas normas se concentram em apenas uma maneira de compartilhar conhecimento científico: o artigo científico. Se pudéssemos estabelecer normas semelhantes e uma economia de reputação que incentivasse um compartilhamento mais amplo do conhecimento científico, a mão invisível da ciência se tornaria mais forte e o processo científico seria grandemente acelerado.

Como podemos expandir a economia de reputação da ciência dessa maneira? Vejamos um exemplo em que tal expansão está começando a acontecer hoje. É uma história que envolve tanto o arXiv — o serviço que vimos anteriormente, que disponibiliza os resultados mais recentes da física para download gratuito — quanto outro serviço para físicos chamado SPIRES. O arXiv e o SPIRES estão juntos criando incentivos para que os físicos compartilhem conhecimento de novas maneiras. Para explicar o que está acontecendo, primeiro preciso explicar o que o SPIRES faz. Suponha que, por algum motivo, você esteja muito interessado em descobrir qual o impacto da última pré-impressão do arXiv de Stephen Hawking no trabalho de outros cientistas. O SPIRES pode ajudar dizendo:

Você sabe quais preprints do arXiv e artigos publicados em periódicos citam o preprint de Hawking. O SPIRES pode informar, por exemplo, que nenhum preprint ou artigo citou o mais recente de Hawking. Ou talvez você descubra que isso estimulou muitos outros físicos a trabalhar em ideias relacionadas.
O SPIRES também pode fornecer uma visão geral da frequência com que os preprints e artigos de Hawking (ou de qualquer outro físico) foram citados em conjunto, e quem os está citando. Isso torna o SPIRES uma ferramenta extremamente útil para avaliar candidatos a empregos científicos. Quando comitês de contratação de físicos se reúnem para avaliar candidatos nas áreas que o SPIRES abrange (física de partículas e algumas áreas relacionadas), não é incomum que todos na reunião estejam com seus laptops à mão, comparando os registros de citações do SPIRES.

O que tudo isso tem a ver com abertura e novos incentivos para compartilhar conhecimento? Bem, algumas décadas atrás, os preprints eram vistos pela maioria dos físicos como meros degraus no caminho para a publicação convencional em periódicos. Eles não eram valorizados como fins em si mesmos. Para construir sua carreira, você precisava de um histórico de artigos de alta qualidade em periódicos. Hoje, por causa do arXiv e do SPIRES, os preprints têm algum status como fins em si mesmos. Não é incomum que físicos, por exemplo, listem preprints que ainda não foram publicados em um periódico em seu currículo. E se um físico descobre alguém trabalhando em um projeto que compete com um dos seus, ele pode se apressar para publicar seu preprint primeiro. Os preprints ainda não têm um status tão alto quanto os artigos de periódicos, mas um preprint com centenas de citações no SPIRES ainda pode ter um grande impacto em termos de carreira. Ao fornecer uma maneira de demonstrar o valor científico e o impacto de uma pré-impressão, o SPIRES e o arXiv criaram um incentivo real para que os físicos produzam pré-impressões, um incentivo diferente do incentivo usual para escrever artigos.

Tenho que admitir que, em termos de mudanças culturais, esta é bem pequena.
A mudança para uma cultura de pré-impressões em física acelera o compartilhamento de conhecimento científico e torna esse conhecimento mais amplamente acessível. Mas não é uma mudança tão grande quanto a substituição de anagramas por periódicos científicos!
Ainda assim, devemos prestar atenção à história do arXiv e do SPIRES, porque ela mostra que é realmente possível criar novos incentivos para que cientistas compartilhem conhecimento. Além disso, isso aconteceu sem qualquer compulsão de uma agência central. Assim que o SPIRES permitiu que o impacto das pré-impressões fosse medido, o novo incentivo surgiu naturalmente, à medida que físicos individuais começaram a usar os relatórios de citação do SPIRES.

Na ciência, como em tantas outras áreas da vida, o que é medido é recompensado, e o que é recompensado é feito.

Uma estratégia semelhante poderia ser usada para incentivar cientistas a compartilhar outros tipos de conhecimento científico? Pensemos, por exemplo, em incentivos para compartilhar dados. Suponhamos que, como aconteceu com as pré-publicações em física, os cientistas comecem a citar regularmente dados de outras pessoas em seus próprios artigos científicos. Isso já está começando a acontecer e acontecerá ainda mais à medida que as políticas de dados abertos se tornarem mais comuns. E suponhamos que alguém crie um serviço de rastreamento de citações que não apenas rastreie citações de artigos e pré-publicações, mas também citações de dados. Se o serviço for bom, as pessoas o usarão para avaliar outros cientistas. E começarão a ver mais vividamente o impacto do compartilhamento de dados. Nesse ponto, o compartilhamento de dados começará a ajudar, em vez de prejudicar, a carreira dos cientistas. De fato, os cientistas não só terão um incentivo para compartilhar seus dados, como também será vantajoso para eles torná-los o mais úteis possível para outros cientistas. Os cientistas começarão a ver a construção da rede de dados como uma parte importante de seu trabalho, não como uma distração da tarefa séria de escrever artigos.

Esse mesmo tipo de incentivo pode ser aplicado a qualquer tipo de conhecimento científico: preprints, dados, código de computador, wikis científicos, mercados de colaboração, o que você quiser. Em cada caso, o padrão geral é o mesmo: citação leva à mensuração, que leva à recompensa, que leva a pessoas motivadas a contribuir. Essa é uma forma de expandir a economia de reputação da ciência. Haverá, na prática, muitas complicações e muitas variações possíveis sobre esse tema. De fato, até mesmo a história arXiv-SPIRES que contei foi simplificada demais: o SPIRES foi apenas um fator entre vários que deram aos preprints um status real na física. Mas o quadro básico é claro.

Um caso de particular importância é o do código de computador. Hoje em dia, cientistas que escrevem e publicam códigos geralmente recebem pouco reconhecimento por seu trabalho. Alguém que criou um excelente programa de software de código aberto usado por milhares de outros cientistas provavelmente receberá pouco crédito dos colegas. "É apenas software" é a resposta de muitos cientistas a esse tipo de trabalho. Do ponto de vista da carreira, o autor do código teria se saído melhor se tivesse dedicado seu tempo a escrever alguns artigos menores que ninguém lê. Isso é loucura: muito conhecimento científico é muito melhor expresso como código do que na forma de um artigo científico. Mas hoje, esse conhecimento muitas vezes permanece oculto ou é forçado a entrar em artigos, porque não há incentivo para fazer o contrário. Mas se tivéssemos um ciclo de citação-mensuração-recompensa para o código, então escrever e compartilhar código começaria a ajudar, em vez de prejudicar, a carreira dos cientistas. Isso

Teria muitas consequências positivas, mas uma delas seria particularmente crucial: daria aos cientistas uma forte motivação para criar novas ferramentas para fazer ciência. Os cientistas seriam recompensados por desenvolver ferramentas como o Galaxy Zoo, o Foldit, o arXiv e assim por diante. E se isso acontecesse, veríamos cientistas se tornando líderes, e não retardatários, no desenvolvimento de novas ferramentas para a construção do conhecimento.

Há limites para a ideia de citação-mensuração-recompensa. Obviamente, não é possível nem desejável julgar uma descoberta com base apenas nas citações que um artigo (ou preprint, dados ou código) recebeu. Quando se trata de avaliar a importância de uma descoberta, nada substitui a compreensão profunda da descoberta. Dito isso, a base para a economia da reputação na ciência é o sistema de citações. É a maneira como os cientistas rastreiam a procedência do conhecimento científico. Se os cientistas quiserem levar a sério as contribuições que vão além dos antigos formatos em papel, devemos ampliar o sistema de citações, criando novas ferramentas e normas para citações, tendo em mente as limitações que as citações têm (e sempre tiveram) como forma de avaliar o trabalho científico.

Hoje, muitos cientistas consideram a ideia de trabalhar de forma mais aberta quase inimaginável. Depois de dar palestras sobre ciência aberta, às vezes sou abordado por céticos que dizem: "Por que eu ajudaria meus concorrentes compartilhando ideias e dados nesses novos sites? Isso não é apenas convidar outras pessoas a roubar meus dados ou a me ultrapassar? Só alguém ingênuo poderia imaginar que isso se tornaria generalizado." Do jeito que as coisas estão atualmente, há muita verdade nesse ponto de vista. Mas também é importante entender seus limites. O que esses céticos esquecem é que ***já*** compartilham livremente suas ideias e descobertas sempre que publicam artigos descrevendo seu próprio trabalho científico. Eles estão tão presos ao sistema de citação-mensuração-recompensa para artigos que o veem como uma lei natural e esquecem que é socialmente construído. É um acordo. E por ser um acordo social, esse acordo pode ser alterado. Tudo o que é necessário para que a ciência aberta tenha sucesso é que o compartilhamento de conhecimento científico em novas mídias tenha o mesmo tipo de prestígio que os artigos têm hoje. Nesse ponto, a recompensa reputacional de compartilhar conhecimento de novas maneiras excederá os benefícios de mantê-lo oculto. A essa altura, os céticos às vezes dirão: "Mas ninguém jamais levará a sério ideias compartilhadas em um ***blog*** [ou wiki, etc.]!". Isso pode ser verdade agora — embora até isso esteja mudando —, mas, a longo prazo, a visão é míope e ignora as lições da primeira revolução da ciência aberta. Temos uma chance real de impulsionar o mesmo tipo de transição que Henry Oldenburg e seus colegas causaram nos séculos XVII e XVIII.

Incentivar cientistas a compartilhar seu conhecimento científico usando as ferramentas mais poderosas disponíveis hoje. Podemos alinhar os interesses individuais dos cientistas com o interesse coletivo da comunidade científica e do público em geral: impulsionar a ciência o mais rápido possível.

Limites à Abertura

Que limites devem ser impostos à abertura na ciência? Embora seja amplamente verdade que, como eu disse anteriormente, informações que não estão na rede não trazem nenhum benefício, alguns limites são necessários. Alguns desses limites são óbvios: médicos não podem compartilhar registros de pacientes à toa, especialistas em segurança não podem compartilhar informações que comprometam a segurança, e assim por diante. É claro que já existem muitas medidas em vigor para impedir a divulgação de informações quando isso violar as expectativas de privacidade, ética, segurança e legalid Mas há preocupações mais sutis sobre abertura que também precisam ser consideradas.

Será que a abertura pode sobrecarregar os cientistas? Um dos maiores matemáticos de todos os tempos, Alexander Grothendieck, acredita que foi sua capacidade de estar sozinho que foi a fonte de sua criatividade. Em notas autobiográficas, ele afirma ter encontrado a verdadeira criatividade como consequência de estar disposto a "alcançar, à minha maneira, as coisas que desejava aprender, em vez de confiar nas noções de consenso, explícito ou tácito, provenientes de um clã mais ou menos extenso do qual eu me tornava membro". Grothendieck não está sozinho nessa crença. Ideias que exigem um desenvolvimento cuidadoso podem murchar e morrer se forem modificadas prematuramente em resposta às opiniões alheias. Talvez, se migrarmos para uma cultura mais aberta e colaborativa, corramos o risco de abrir mão da independência de espírito necessária para as formas mais elevadas de criatividade. Será que menos pessoas tentarão trabalhos ousados que não se enquadram na práxis compartilhada de uma comunidade científica existente, mas que, em vez disso, visam definir uma nova práxis?

Há um problema geral aqui que vai além do desejo de solidão de Grothendieck ou das noções românticas de gênios solitários redefinindo campos.
O problema, que discutimos no final do capítulo 3, é que os cientistas têm tempo limitado, e isso impõe restrições à forma como trabalham com os outros. Devem colaborar pouco, muito ou nada?

Se optarem por colaborar, com quem devem trabalhar? Não importa o quanto gostem de colaborar, sua atenção não é infinita e, portanto, deve ser gerenciada com cuidado. Às vezes, a resolução do problema é, como para Grothendieck, buscar a solidão. Mas para os cientistas que optam por colaborar, o problema se manifesta de outras maneiras. No Projeto Polymath, por exemplo, um pequeno número de contribuições veio de pessoas sem a formação matemática necessária para fazer um progresso significativo no problema. Essas pessoas estavam fora da prática compartilhada pela maioria dos participantes do Polymath. Embora suas contribuições fossem bem-intencionadas, foram de pouca ajuda. Felizmente, houve poucas contribuições de baixa qualidade e elas foram facilmente ignoradas. Mas se houvesse mais, elas teriam sobrecarregado significativamente a atenção de outros participantes do Polymath. Problemas semelhantes podem ser causados por excêntricos, trolls e spammers, ou mesmo pessoas simplesmente desagradáveis.

Esses problemas são sérios, mas não intransponíveis. Um sistema pode ser aberto sem exigir que todos os participantes recebam igual atenção. E você pode compartilhar seu conhecimento abertamente, sem precisar prestar atenção a todos (ou, na verdade, a qualquer outra pessoa). Em geral, para que sistemas colaborativos abertos funcionem com mais eficácia, os participantes devem ter meios eficazes de filtrar informações, para que possam se concentrar nas informações de maior interesse e ignorar o restante. Na competição MathWorks, por exemplo, lembre-se de como a pontuação ajuda os participantes a filtrar ideias inúteis e a se concentrar nas melhores ideias de outros usuários. E se contribuições de baixa qualidade se tornarem um problema maior no Projeto Polymath, elas também poderão ser filtradas. Idealmente, a ciência é aberta, mas fortemente filtrada. Isso é uma consequência natural do fato de que, embora nossa atenção não seja escalável, o compartilhamento de conhecimento é. Em um mundo aberto, mas filtrado, não há problema em pessoas como Grothendieck seguirem seu próprio programa solitário.

A ciência aberta não será, às vezes, usada para fins que muitos cientistas consideram desagradáveis? Em novembro de 2009, hackers invadiram um sistema de computador em um dos principais centros de pesquisa climática do mundo, a Unidade de Pesquisa Climática da Universidade de East Anglia, no Reino Unido. Os hackers baixaram mais de 1.000 mensagens de e-mail trocadas entre cientistas do clima. Em seguida, vazaram os e-mails (e muitos outros documentos) para blogueiros e jornalistas. O incidente recebeu atenção da mídia mundial, com muitos céticos da mudança climática aproveitando os e-mails, alegando que eles continham evidências de que a noção de mudança climática causada pelo homem era uma conspiração entre cientistas do clima. Um dos exemplos usados para apoiar essa afirmação foi um e-mail de Kevin

Trenberth, um renomado cientista climático do Centro Nacional de Pesquisa Atmosférica em Boulder, Colorado. Em seu e-mail, Trenberth afirma: "O fato é que não podemos explicar a falta de aquecimento do sistema computacional, e é uma farsa que não possamos". Na verdade, a frase estava sendo citada de forma inadequada, fora de contexto. No e-mail, Trenberth discutia um artigo publicado recentemente, que analisava as causas da variação anual da temperatura da superfície da Terra — por que temos anos mais quentes e mais frios — e como essa variação se relaciona com o aumento geral da temperatura a longo prazo. As variações anuais são presumivelmente devidas a mudanças na forma como o calor da superfície é redistribuído para o oceano, para o gelo derretido e assim por diante. O e-mail e o artigo de Trenberth apontavam que não compreendemos completamente todos os processos que causam essas variações e, portanto, não podemos necessariamente explicar por que um determinado ano é mais quente ou mais frio. Embora o e-mail expressasse alguma frustração com essa situação, não contradizia de forma alguma sua crença no aumento da temperatura a longo prazo, que ofusca as variações de curto prazo. Observe que a questão aqui não é se você concorda com Trenberth sobre as mudanças climáticas. A questão é que um cético cuidadoso e honesto em relação às mudanças climáticas não poderia interpretar o e-mail de Trenberth como expressando qualquer dúvida de sua parte de que os humanos estão causando as mudanças climáticas. No entanto, muitos céticos optaram por citar a frase fora de contexto, seja maliciosamente, para atingir seus próprios objetivos, ou descuidadamente, por genuína ignorância da intenção original.

Esse tipo de incidente ilustra um grande risco enfrentado por cientistas do clima que estão considerando trabalhar de forma mais aberta. Por um lado, o compartilhamento aberto de ideias e dados tem o potencial de acelerar a descoberta. Por outro lado, toda informação compartilhada por cientistas do clima, não importa quão inocente, corre o risco de ser atacada por grupos que querem desacreditar a ciência do clima exagerando problemas menores ou relatando comentários como os de Trenberth fora de contexto. Diante disso, quão abertamente o trabalho em ciência do clima deve ser feito? Esta não é uma pergunta fácil de responder. Se as questões fossem exclusivamente científicas, então os cientistas do clima deveriam agir rapidamente para trabalhar de forma mais aberta. Mas as questões não são apenas científicas, elas também são políticas. Acredito que a abordagem correta não é fazer uma mudança drástica, mas sim caminhar gradualmente em direção a um sistema mais aberto, diagnosticando e corrigindo os problemas à medida que eles surgem.

A ciência aberta pode levar à disseminação de desinformação? Nas últimas duas décadas, cientistas descobriram mais de 500 planetas orbitando outras estrelas além do nosso Sol. Essas descobertas são empolgantes.

mas até recentemente, todos os planetas extrassolares confirmados eram gigantes gasosos, mais parecidos com Júpiter ou Netuno do que com a Terra. Esperando mudar essa situação, no início de 2009 a NASA lançou a Missão Kepler, um observatório espacial que os astrônomos acreditavam que poderia descobrir os primeiros planetas do tamanho da Terra orbitando outras estrelas. A política da NASA normalmente exige a divulgação aberta de dados dessas missões dentro de um ano, e era amplamente esperado pelos cientistas que os dados do Kepler fossem divulgados em junho de 2010. Mas em abril de 2010, um painel consultivo da NASA concedeu uma extensão incomum, permitindo que a equipe do Kepler retivesse dados sobre os 400 melhores candidatos a planetas até fevereiro de 2011. Isso lhes deu mais tempo para analisar os dados e uma chance melhor de serem os primeiros a descobrir planetas do tamanho da Terra. Em um artigo no ***New York Times,*** o líder da equipe do Kepler, William Borucki, é citado justificando a extensão como uma forma de proteção contra falsas alegações de descoberta por outros astrônomos, dizendo que "Se dissermos: 'Sim, eles são planetas pequenos', você pode ter certeza de que são". Em fevereiro de 2011, a equipe do Kepler anunciou que havia, de fato, descoberto cinco planetas do tamanho da Terra.

Embora praticar ciência abertamente seja, no geral, preferível, Borucki não está totalmente errado em se preocupar com falsas alegações. Em 8 de julho de 2010, o físico de partículas e blogueiro Tommaso Dorigo usou seu blog para relatar rumores de que a tão procurada partícula de Higgs havia finalmente sido descoberta. Sua postagem enfatizou que ele estava apenas repetindo rumores não confirmados, mas, apesar dessa ressalva, os rumores em seu blog foram repercutidos pela grande mídia e levaram a artigos em veículos como o ***Daily Telegraph*** (Reino Unido) e a revista ***New Scientist .*** Apenas nove dias depois, em 17 de julho, Dorigo usou seu blog para retratar o boato: era um alarme falso. Alguns cientistas criticaram Dorigo, alegando que ele agiu de forma irresponsável ou que estava apenas em busca de notoriedade. Mas rumores científicos são um elemento básico da vida científica, o tipo de coisa sobre a qual os cientistas falam durante o almoço ou no corredor. De fato, é por meio desse tipo de discussão especulativa que novas ideias frequentemente nascem. E, portanto, era natural abordar o assunto no ambiente informal de um blog, onde Dorigo pudesse conversar com seus amigos e colegas físicos de partículas. Diante disso, é tentador criticar a grande mídia por suas reportagens irresponsáveis. Mas isso também não é justo. Dorigo é um físico profissional, bem conhecido e bem relacionado na comunidade da física de partículas, alguém que se presume estar por dentro do assunto. É claro que a grande mídia noticiou esses boatos.

Há uma tensão genuína aqui. Os blogs são uma forma poderosa de ampliar a conversa científica informal e explorar ideias especulativas. Mas

Quando essa exploração é realizada abertamente, há o perigo de que a grande mídia, ávida por um furo de reportagem, espalhe notícias dessa especulação, criando a impressão de que se trata de um fato. Felizmente, esse é um problema de escopo limitado. A grande mídia não está interessada na maioria das descobertas científicas e, para aquelas poucas descobertas de amplo interesse, eventos como o incidente Dorigo-Higgs ajudarão a tornar a mídia mais cautelosa ao relatar rumores não confirmados. Embora as pessoas frequentemente sejam céticas em relação ao jornalismo, a maioria das grandes organizações de mídia está profundamente ciente de sua reputação de credibilidade (ou não) e se envergonha se tiver que fazer retratações públicas frequentes. A notícia da retratação de Dorigo-Higgs foi veiculada por mais de meia dúzia de grandes organizações de mídia, muitas das quais apontaram que o boato foi originalmente espalhado pelo ***Telegraph*** e ***pelo New Scientist.***

Esse não é o tipo de publicidade que o ***Telegraph*** e ***o New Scientist*** querem.
Dito isso, veremos esse problema cada vez mais no futuro.
Parece um preço relativamente pequeno pelos benefícios da ciência aberta.

Aumentar a escala da ciência não tornará mais difícil a verificação das descobertas científicas? À medida que a ciência aberta nos permite ampliar o processo de descoberta, a natureza das evidências científicas mudará e se tornará mais complexa. No caso de algumas descobertas, compreender as evidências em detalhes pode estar além da capacidade de qualquer pessoa. Um exemplo inicial disso ocorreu em 1983, quando matemáticos anunciaram a solução de um importante problema matemático, conhecido como a classificação dos grupos finitos simples. A prova levou quase 30 anos para ser concluída, de 1955 a 1983, e envolveu 100 matemáticos escrevendo aproximadamente 500 artigos em periódicos. Muitas pequenas lacunas foram posteriormente encontradas na prova, e pelo menos uma lacuna séria, que agora foi resolvida (esperamos!) por um suplemento de dois volumes e 1.200 páginas à prova. Na década de 1980, era incomum que uma descoberta científica tivesse evidências de tamanha complexidade. Hoje, isso está se tornando comum. Para escolher apenas um exemplo de complexidade, considere que experimentos modernos em muitos campos científicos estão cada vez mais propensos a usar centenas de milhares ou até milhões de linhas de código de computador. É quase impossível eliminar todos os bugs desse código. Como podemos ter certeza de que os resultados gerados por esse código são válidos? Como outros cientistas podem verificar e reproduzir os resultados desses experimentos? Além disso, a situação está se tornando mais desafiadora, à medida que nossos sistemas computacionais se tornam mais complexos. Programas de software individuais estão sendo cada vez mais substituídos por ecologias de software, redes complexas de programas em interação, às vezes mantidas por muitas pessoas em vários locais. Como...

Garantimos que tais ecologias de software produzirão resultados confiáveis e reprodutíveis? Essas e outras preocupações semelhantes afetam descobertas que vão da física de partículas à ciência climática, da biologia à astronomia.

É um tipo de ciência que vai além da compreensão individual. À medida que essa nova escala de evidências se torna a norma, nossos padrões de evidência precisarão evoluir. Estou otimista, porém, de que enfrentaremos o desafio, usando nossa inteligência coletiva ampliada não apenas para fazer novas descobertas, mas também para desenvolver métodos aprimorados para testá-las e validá-las.

Passos práticos em direção à ciência aberta

Que medidas práticas podemos tomar em direção à ciência aberta? Em todo o mundo, nossos governos gastam mais de 100 bilhões de dólares por ano em pesquisa básica. Esse dinheiro é ***nosso ,*** e devemos exigir uma mudança para uma cultura científica mais aberta. Acredito que a ciência financiada com recursos públicos deve ser ciência aberta. Vejamos algumas medidas práticas que todos, desde cientistas em atividade até o público em geral, podem tomar para atingir esse objetivo.

O que você pode fazer se for um cientista? Experimente a ciência aberta! Publique alguns dos seus dados e códigos antigos online. Documente-os, incentive outras pessoas a usá-los e certifique-se de informar como gostaria de ser citado. Experimente blogar. Expanda sua zona de conforto — tente usar seu blog para desenvolver algumas daquelas ideias que você tem guardado na cabeça há anos, mas nunca conseguiu concretizar. Você tem pouco a perder, e trabalhar abertamente pode dar um novo fôlego às suas ideias. Se isso for um compromisso de tempo demais, tente fazer algumas pequenas contribuições para projetos de ciência aberta de outras pessoas — digamos, fazendo um comentário em um blog científico ou uma contribuição para um wiki. Essas contribuições podem ser pequenas, mas seus colegas cientistas perceberão, e isso ajudará a legitimar as novas ferramentas na comunidade científica. E você pode achar isso mais gratificante do que imagina. Se você for aventureiro, tente ultrapassar os limites. Pergunte a si mesmo se você pode ser pioneiro em uma nova maneira de fazer ciência, como o Projeto Polymath, o Foldit e o Galaxy Zoo fizeram. O que você pode conjurar com imaginação e determinação? Mesmo que seus empreendimentos em ciência aberta não sejam bem-sucedidos, pense em seus esforços como um serviço à sua comunidade.

E, claro, você não precisa fazer toda a sua ciência abertamente, nem mais do que uma pequena fração.

Acima de tudo, seja generoso ao dar crédito a outros cientistas quando eles compartilham seu conhecimento científico de novas maneiras. Encontre maneiras de citar as ideias, os dados e os códigos que eles compartilham online. Incentive-os a promover seu trabalho aberto, a destacá-lo em seus currículos e em suas solicitações de financiamento, e a encontrar maneiras de demonstrar seu impacto. Essa é a maneira de iniciar novos ciclos de citação-mensuração-recompensa. É claro que, às vezes, você encontrará colegas com valores científicos antiquados, pessoas que desprezam novas maneiras de compartilhar conhecimento e que pensam que a única medida de sucesso para cientistas é quantos artigos publicaram em periódicos de alto nível, como a ***Nature.*** Converse com essas pessoas sobre o valor de novas maneiras de compartilhar conhecimento e sobre a coragem necessária para cientistas, especialmente jovens cientistas, trabalharem abertamente. Compartilhar ideias, códigos e dados abertamente online é tão importante quanto publicar artigos, e somente valores antiquados dizem o contrário.

Se você é um cientista que também é programador, tem um papel especial a desempenhar, uma oportunidade de construir novas ferramentas que redefinem a forma como a ciência é feita. Seja ousado ao experimentar novas ideias: esta é a era de ouro do software científico. Mas também seja ousado ao afirmar o valor do seu trabalho. Hoje em dia, é provável que seu trabalho seja desvalorizado por colegas antiquados, não por malícia, mas por falta de compreensão. Explique a outros cientistas como eles devem citar seu trabalho. Trabalhem em conjunto com seus amigos cientistas programadores para estabelecer normas compartilhadas para citação e compartilhamento de código. E então trabalhem juntos para aumentar gradualmente a pressão sobre outros cientistas para que sigam essas normas. Não apenas promova seu próprio trabalho, mas também insista de forma mais ampla no valor do código como uma contribuição científica por si só, tão valioso quanto as formas mais tradicionais.

E se você trabalhar em uma agência de fomento? Converse com pessoas nas comunidades científicas que você atende e pergunte qual conhecimento está atualmente retido nas cabeças e laboratórios dos cientistas. Quais ferramentas seriam mais eficazes para compartilhar esse conhecimento? Existe uma oportunidade para desenvolver políticas de acesso aberto, dados abertos e código aberto? Como podemos ir além das políticas atuais de acesso aberto e dados abertos? Podemos usar exemplos como o arXiv e o SPIRES como modelos para ajudar a criar novas normas de citação e novas ferramentas de mensuração, e assim expandir a economia de reputação da ciência? De forma mais geral, se você está envolvido no governo ou no processo de formulação de políticas, pode ajudar:

envolvendo-se, fazendo lobby pelo acesso aberto e dados abertos e, de forma mais geral, conscientizando sobre a questão da ciência aberta.

E o que você pode fazer se não for cientista, não trabalhar para uma agência de fomento ou com políticas públicas, mas for um cidadão interessado em ciência e bem-estar humano? Converse com seus amigos e conhecidos que são cientistas. Pergunte a eles o que estão fazendo para tornar seus dados abertos.

Pergunte a eles o que fazem para compartilhar suas ideias pública e rapidamente. Pergunte a eles como compartilham seu código. Para que a ciência aberta tenha sucesso, o que é necessário é uma mudança nos valores da comunidade científica. Se todos os cientistas acreditarem de todo o coração no valor de trabalhar de forma aberta, online, então a mudança virá. Este é fundamentalmente um problema de mudança de corações e mentes. Não há força maior para alcançar tal mudança do que aumentar a conscientização pública, para que ***todos*** em nossa sociedade entendam o tremendo valor da ciência aberta e entendam que alcançar a ciência aberta é um dos grandes desafios da nossa era. Se todos os cientistas do mundo forem questionados por seus amigos e familiares sobre o que estão fazendo para tornar a ciência mais aberta, então a mudança virá. Se todos os agentes de fomento e todos os líderes em nossas universidades forem questionados por ***seus*** amigos e familiares sobre o que estão fazendo para tornar a ciência mais aberta, então a mudança virá. E se nossos políticos forem pressionados por um público que exige uma cultura científica mais aberta, então a mudança virá.

Isso precisa se tornar uma questão de interesse geral para a nossa sociedade, uma questão política e social que seja entendida por todos como de fundamental importância. Você pode ajudar a alcançar isso usando seu poder pessoal, suas conexões e sua imaginação para pressionar políticos e agências de fomento a elaborarem políticas que incentivem a abertura.

(Listei algumas organizações que já fazem isso na seção "Fontes Selecionadas e Sugestões para Leitura Adicional", no final deste livro.) Que tipos de conhecimento nós, como sociedade, esperamos e incentivaremos os cientistas a compartilhar com o mundo? Continuaremos com nossa abordagem atual? Ou optaremos por criar uma cultura científica que abrace o compartilhamento aberto do conhecimento, o desenvolvimento de novas ferramentas que ampliem nossa capacidade de resolução de problemas e acelerem a descoberta científica?

Os passos que acabei de descrever são todos pequenos. Mas, juntos, criarão um movimento irreversível em direção a formas mais abertas de fazer ciência. O inventor e cientista Daniel Hillis observou que "há problemas que são impossíveis se você pensar neles em termos de dois anos — o que todo mundo faz —, mas são fáceis se você pensar em termos de cinquenta anos". O problema da ciência aberta é um problema desse tipo. Hoje,

Criar uma cultura científica aberta parece exigir uma mudança impossível na forma como os cientistas trabalham. Mas, com pequenos passos, podemos gradualmente causar uma grande mudança cultural.

A Era da Ciência em Rede

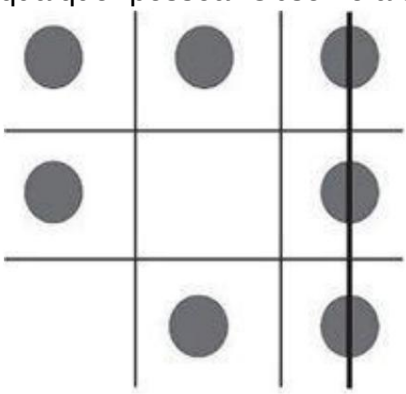
Escrevi este livro com o objetivo de reacender a chama da comunidade científica. Vivemos um momento único na história: pela primeira vez, temos a capacidade ilimitada de construir novas e poderosas ferramentas para o pensamento. ***Temos a oportunidade de mudar a forma como o conhecimento é construído.*** Mas a comunidade científica, que deveria estar na vanguarda, está, em vez disso, na retaguarda, com a maioria dos cientistas apegada à sua forma de trabalhar existente e falhando em apoiar aqueles que buscam uma forma melhor. Assim como na primeira revolução da ciência aberta, como sociedade, precisamos evitar ativamente essa tragédia da oportunidade perdida, incentivando e, quando apropriado, compelindo os cientistas a contribuírem de novas maneiras. Acredito que, com trabalho árduo e dedicação, temos uma boa chance de revolucionar completamente a ciência.

Quando olhamos para a segunda metade do século XVII, podemos ver que uma das grandes mudanças daquela época foi a invenção da ciência moderna. Quando a história do final do século XX e início do século XXI for escrita, veremos este como ***o*** momento em que a informação mundial foi transformada de um estado inerte e passivo para um sistema unificado que a torna viva. A informação mundial está despertando. E essa mudança nos dá a oportunidade de reestruturar a maneira como os cientistas pensam e trabalham, e assim ampliar a capacidade da humanidade de resolver problemas. Estamos reinventando a descoberta, e o resultado será uma nova era de ciência em rede que acelerará a descoberta, não em um pequeno canto da ciência, mas em toda a ciência.

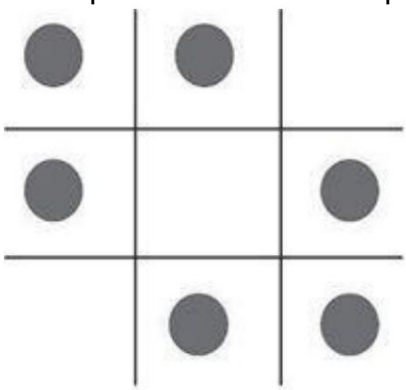
Essa reinvenção aprofundará nossa compreensão de como o universo funciona e nos ajudará a abordar nossos problemas humanos mais críticos.

Apêndice: O Problema Resolvido pelo Polímata Projeto

O Projeto Polymath teve como objetivo provar um resultado matemático conhecido como teorema de Hales-Jewett (DHJ) da densidade. Embora a demonstração do DHJ seja complexa, a afirmação básica pode ser compreendida por qualquer pessoa. Observe a seguinte grade três por três:



Marquei sete dos quadrados na grade com um ponto; como você pode ver, é possível traçar uma linha passando por três desses pontos. Em contraste, a configuração na imagem a seguir não tem ***linhas — você*** não pode traçar uma linha passando por nenhum dos três pontos:



Se você brincar um pouco, descobrirá que esta configuração é a maior configuração sem linhas possível. Em particular, se você marcar sete pontos na grade, não importa como você os posicione, sempre será possível traçar uma linha passando por três deles, em algum lugar da grade.

Imagine agora que estendemos a grade para três dimensões, ou seja, uma grade de três por três por três. Acontece que, em três dimensões, a maior configuração possível sem linhas tem 16 locais preenchidos. Se preenchermos 17 locais na grade, independentemente de quais locais preenchermos, será possível traçar uma linha através de três pontos em algum lugar da grade. Você pode acreditar em mim ou, se preferir, com um pouco de trabalho e imaginação tridimensional, pode se convencer de que é esse o caso.

Vamos dar um salto agora e imaginar a extensão da grade de três dimensões para um número arbitrário de dimensões espaciais. Daremos um rótulo ao número de dimensões — chamaremos de n. Essa extensão é difícil de visualizar, difícil o suficiente para que a maioria dos matemáticos não consiga fazê-la, e, em vez disso, traduzam o problema para uma forma algébrica. Não farei a tradução algébrica aqui, mas apenas explicarei a questão que nos preocupa: qual é o tamanho da maior configuração sem linhas em uma grade em ***n*** dimensões? Daremos um nome a esse tamanho, chamando-o de ***sn .*** Nossa discussão acima indica que s2 = 6 e s3 = 16, os tamanhos das maiores configurações sem linhas em duas e três dimensões. Em dimensões maiores, rapidamente se torna extremamente difícil descobrir o valor de ***sn .*** Matemáticos calcularam o valor de s^2 e s^3 , como vimos, e também, com mais esforço, s^2 , s^2 e s^2 , mas ninguém no mundo sabe qual é o valor exato de s^2 . E a situação se complica ainda mais em dimensões ainda maiores. Mas, embora seja difícil descobrir um valor exato para ***s² ,*** o teorema de DHJ nos fornece algumas informações parciais sobre o quão grande é ***s² .***

Em particular, uma consequência do teorema DHJ é que, à medida que o número de dimensões ***n*** aumenta, o tamanho ***sn*** da maior configuração sem linhas é apenas uma pequena fração do número total de localizações na grade. Em outras palavras, à medida que ***n*** aumenta, preencher até mesmo uma pequena fração da grade força uma linha em algum lugar. Não importa o quão inteligente você seja em preencher as localizações, haverá uma linha em algum lugar. Para colocar a afirmação em termos um pouco mais formais, o teorema DHJ nos diz que a fração da grade ***sn /*** 3n ocupada pela maior configuração sem linhas

a configuração fica extremamente pequena à medida que ***n*** se torna grande — ela vai a zero no limite de n grande, para usar o jargão matemático.

Esta é uma afirmação surpreendente. Como vimos, em duas e três dimensões podemos preencher a maior parte da grade antes de sermos forçados a colocar três peças em uma linha. No entanto, em dimensões elevadas, a DHJ nos diz que uma linha é forçada em algum lugar da grade, mesmo que apenas uma pequena fração da grade seja preenchida. Não é nada óbvio que esse seja o caso, e ainda assim o teorema da DHJ nos diz que isso é verdade.

Tenho descrito as consequências do teorema DHJ para dar a vocês uma ideia do que ele diz. Na verdade, a afirmação completa do teorema DHJ é mais forte do que as consequências que descrevi até agora. Ele não funciona apenas para grades de três por três por...; uma afirmação análoga é verdadeira para grades m por m por ..., onde ***m*** é qualquer número.

Além disso, o teorema DHJ nos diz que a reta será um tipo especial de reta, chamado de reta ***combinatória .*** Não definirei retas combinatórias aqui — veja as referências nas notas de rodapé se quiser uma explicação sobre o que é uma reta combinatória. Por enquanto, basta que sejam um tipo especial de reta. O que a declaração completa do teorema DHJ diz é que, à medida que o número de dimensões ***n*** aumenta, a fração da grade ***m*** × ***m ...*** ocupada pelo maior subconjunto sem uma reta combinatória tende a zero. Em outras palavras, à medida que ***n*** aumenta, preencher até mesmo uma pequena fração da grade forçará uma reta combinatória em algum lugar.

Por que você deveria se importar com o DHJ? Se você está chegando ao DHJ sem muito conhecimento matemático, talvez pareça um problema obscuro. O DHJ parece o tipo de quebra-cabeça que pode ser uma diversão potencialmente divertida (embora difícil), se você tiver uma mente para resolver quebra-cabeças.

Mas por que o teorema DHJ é mais importante do que resolver um quebra-cabeça de Sudoku?

As aparências enganam. DHJ é um teorema profundo. Ele acaba tendo como consequência muitos outros resultados matemáticos importantes e difíceis de provar, alguns em áreas que parecem não ter relação entre si. Pense nele como um dominó: quando cai, faz com que muitos outros dominós matemáticos importantes e, de outra forma, difíceis de mover também caiam. Deixe-me dar um exemplo de como DHJ se conecta a outra parte da matemática que parece não ter relação: o problema de compreender a estrutura dos números primos. Acontece que DHJ implica um resultado profundo da teoria dos números chamado teorema de Szemerédi. Esse teorema foi provado pela primeira vez em 1975 pelo matemático Endre Szemerédi; desde então, matemáticos encontraram várias provas adicionais. Usando ideias extraídas de várias dessas provas, em 2004 os matemáticos Ben Green e Terence Tao provaram um

Novo resultado importante sobre a estrutura dos números primos. Para entender o que diz o teorema de Green-Tao, considere a sequência de números 199, 409, 619, 829, 1039, 1249, 1459, 1669, 1879, 2089. Todos esses são números primos e são uniformemente espaçados; cada membro da sequência é 210 vezes maior que o que o precede. O que o teorema de Green-Tao diz é que você pode encontrar sequências uniformemente espaçadas de números primos de qualquer comprimento. Quer uma sequência uniformemente espaçada de um milhão de números primos? Green-Tao garante que tal sequência existe. O teorema não fornece, na verdade, uma receita facilmente utilizável para encontrar tal sequência, mas garante que, se você procurar por uma sequência por tempo suficiente, você a encontrará eventualmente. Agora, os resultados sobre os números primos provavelmente parecem bastante alheios à preocupação com configurações sem linhas em alta dimensão. E, no entanto, as conexões DHJ-Szemerédi e Szemerédi-Green-Tao sugerem que realmente há uma conexão entre DHJ e a estrutura dos números primos.

O teorema de DHJ foi provado pela primeira vez em 1991 pelos matemáticos Hillel Furstenberg e Yitzhak Katznelson. Portanto, quando Tim Gowers propôs o Projeto Polímata, ele não estava propondo que os polímatas encontrassem a primeira prova de DHJ. Em vez disso, ele estava propondo que encontrassem uma nova prova. Você pode se surpreender ao saber que um matemático renomado como Gowers estaria interessado em encontrar uma nova prova para um resultado já conhecido. Mas a prova existente de DHJ utilizava técnicas indiretas e bastante avançadas de um ramo da matemática chamado teoria ergódica. Embora fosse uma prova perfeitamente boa, Gowers acreditava que uma compreensão adicional da DHJ poderia ser obtida encontrando uma nova prova que se baseasse em técnicas diferentes. Em particular, Gowers estava interessado em encontrar uma prova que se baseasse apenas em técnicas elementares, isto é, técnicas que não exigissem matemática sofisticada, como as ferramentas da teoria ergódica.

Às vezes, encontrar novas provas pode nos dar insights significativos que nos ajudam a entender por que um resultado é verdadeiro em primeiro lugar. De fato, foi exatamente isso que aconteceu com as múltiplas provas do teorema de Szemerédi. Quando Green e Tao provaram seu teorema sobre números primos, eles se basearam em ideias de diversas provas diferentes do teorema de Szemerédi. Isso fez com que encontrar uma nova prova do teorema DHJ usando apenas técnicas elementares fosse uma meta desafiadora e valiosa para o Projeto Polymath.

Agradecimentos

Ao escrever este livro, beneficiei-me enormemente do entusiasmo, da perspicácia e do apoio de muitas pessoas. Agradeço especialmente a Peter Tallack, meu agente, cujo entusiasmo pelo projeto, feedback perspicaz e talento para fazer as perguntas certas aprimoraram drasticamente o livro. Agradeço também à equipe da Princeton University Press, tanto pelo entusiasmo quanto pela paciência em me ajudar a transformar este livro em realidade. Sou particularmente grato à minha editora, Ingrid Gnerlich, bem como a Jodi Beder, Bob Bettendorf, Christopher Chung, Kathleen Cioffi, Peter Dougherty, Jessica Pellien e Julie Shawvan. Agradeço a Simon Capelin, Kelly McNees e Lee Smolin pelos comentários úteis sobre as primeiras versões da minha proposta de livro. As palavras de incentivo de Lee Smolin foram especialmente bem-vindas em um momento em que eu estava prestes a abandonar o livro por completo. Um grande agradecimento a Eva Amsen, Rob Dodd, Eric Drexler, John Dupuis, Hassan Masum, Christina Pikas, Dorothea Salo, Lee Smolin e Rob Spekkens por fornecerem comentários ponderados sobre um rascunho do livro inteiro. Rob Spekkens não apenas forneceu comentários detalhados sobre o livro inteiro, mas também fez várias sugestões gerais que melhoraram drasticamente todo o trabalho. Agradecimentos a Harvey Brown, Amy Dodd, Danielle Fong, Chris Ing, Chris Lintott, Garrett Lisi, Cameron Neylon, Tobias Osborne, Peter Rohde, Mickey Schafer, Carlos Scheidegger, Arfon Smith, John Stockton e Mark Tovey pelos comentários detalhados sobre os primeiros rascunhos de um ou mais capítulos. Os capítulos 8 e 9 deste livro são parcialmente adaptados de um ensaio que escrevi para o meu blog [152] e que foi posteriormente republicado na ***Physics World*** [150]. Meus agradecimentos a João Medeiros e Matin Durrani, da ***Physics World,*** pela ajuda com o artigo e por torná-lo possível. Meu trabalho foi grandemente enriquecido pelas muitas pessoas na comunidade online de

a todos dessa comunidade. Agradeço especialmente a Cameron Neylon, Peter Suber e a muitos outros que tanto contribuíram para criar uma comunidade online próspera para a ciência aberta. Obrigado também a muitas outras pessoas cujos insights impulsionaram meu pensamento, incluindo: Scott Aaronson, Hal Abelson, Richard Akerman, Dave Bacon, Gavin Baker, Travis Beals, Pedro Beltrao, Mic Berman, Michael Bernstein, Peter Binfield, Robin Blume-Kohout, Jean-Claude Bradley, Björn Brembs, Titus Brown, Zacary Brown, Howard Burton, Carl Caves, Ike Chuang, Ken Coates, Alessandro Cosentino, John Cumbers, Wim van Dam, Amy Dodd, Rob Dodd, Michael Duschenes, Drew Endy, Steven van Enk, Steve Flammia, Connie French, Chris Fuchs, Joshua Gans, Alexei Gilchrist, Benjamin Good, Daniel Gottesman, Tim Gowers, Christopher Granade, Ilya Grigorik, Nicholas Gruen, Melissa Hagemann, Timo Hannay, Aram Harrow, Andrew Hessel, Daniel Holz, Tad Homer-Dixon, Bill Hooker, Sabine Hossenfelder, Jonathan Hunt, Heather Joseph, Jason Kelly, Marius Kempe, Manny Knill, Steve Koch, Matt Leifer, Hope Leman, Daniel Lemire, Debbie Leung, Mike Loukides, Sean McGee, Bob McNees, Hassan Masum, Gerard Milburn, Len Mlodinow, Peter Murray-Rust, Brian Myers, Béla Nagy, Anders Norgaard, Jill O'Neill, Tobias Osborne, Seb Paquet, Heather Piwowar, Jorge Pullin, Srinivasan Ramasubramanian, Neil Saunders, Kevin Schawinski, Cosma Shalizi, Alice Sheppard, John Sidles, Deepak Singh, Rolando Somma, Hilary Spencer, Graham Steel, Victoria Stodden, Dan Stowell, Brian Sullivan, Pawel Szcz sny, Terry Tao, Kaitlin Thaney, Matthew Todd, Ben Toner, Umesh Vazirani, Ricardo Vidal, Christian Weedbrook, Andrew White, John Wilbanks, Greg Wilson, Shirley Wu, Carl Zimmer e Bora Zivkovic. Às outras pessoas cujos nomes deveriam estar nesta lista, mas não estão, minhas desculpas e meus agradecimentos.

Mais do que posso dizer, agradeço à minha família, por seu amor e apoio: Howard e Wendy Nielsen; Stuart e Shelly Nielsen; Kate Nielsen e Scott Andrews; meus sobrinhos Zie, Cooper, Blake e Bowen; minha avó Ricardo; Rob e Diane Dodd; Amy Dodd; e Candace e Jonny Marzano. Meus maiores agradecimentos vão para minha esposa, Jen Dodd. Os comentários e críticas de Jen aprimoraram muito todos os aspectos deste livro, desde os mínimos detalhes até a estrutura em grande escala e o argumento geral. Sem seu incentivo e apoio, eu nunca teria começado o livro e certamente nunca o teria terminado.

Fontes selecionadas e sugestões para futuras pesquisas

Leitura

Este livro é, em grande parte, uma obra de síntese e tem uma enorme dívida com o trabalho de outros autores. Notas detalhadas sobre minhas fontes podem ser encontradas a partir da página 221. Aqui, descrevo algumas das fontes que influenciaram de forma mais decisiva meu pensamento e sugiro leituras adicionais.

Inteligência coletiva: A ideia de usar computadores para ampliar a inteligência humana individual e coletiva tem uma longa história. Trabalhos iniciais influentes incluem o célebre artigo de Vannevar Bush, “As We May Think” [31], que descreveu seu sistema memex imaginado e inspirou o trabalho seminal de Douglas Engelbart [63] e Ted Nelson [145]. Embora esses trabalhos tenham muitas décadas, eles expõem muito do que vemos na internet atual e revelam perspectivas além. Além desses trabalhos fundamentais, minhas ideias sobre inteligência coletiva foram fortemente influenciadas por ideias econômicas. Herbert Simon [197] parece ter sido a primeira pessoa a apontar o papel crucial da atenção como um recurso escasso em um mundo rico em informações. Também apreciei muito o artigo provocativo de Michael Goldhaber [75] sobre “A Economia da Atenção e a Rede”. Complementando isso, está o trabalho do teórico da complexidade Scott Page, que demonstra o valor da diversidade cognitiva na resolução de problemas em grupo [168], e a noção de “conhecimento oculto” de Hayek e o uso de preços como sinais para agregar esse conhecimento [93]. Outros trabalhos influentes sobre temas relacionados incluem a análise antropológica detalhada de Hutchins sobre a inteligência coletiva na navegação de um navio [95], o livro de Lévy sobre inteligência coletiva [124] e a estimulante coleção de ensaios sobre inteligência coletiva recentemente reunida por Mark Tovey [224]. Escrevendo de um ponto de vista muito diferente

Nesta perspectiva, David Easley e Jon Kleinberg escreveram um excelente livro didático, " ***Networks, Crowds, and Markets*** " [59], que resume grande parte da pesquisa matemática e quantitativa sobre redes. Por fim, recomendo o livro de Nicholas Carr, ***"The Shallows*** " [35]. Ele levanta a questão fundamental: como as ferramentas online estão mudando a maneira como pensamos (individualmente)? Acredito que a resposta de Carr esteja incompleta, mas é uma exploração estimulante dessa importante questão.

Código aberto: A melhor maneira de se informar sobre código aberto é participando de alguns projetos de código aberto. Você também pode aprender muito lendo o código e os arquivos de discussão de projetos de código aberto, como Linux e Wikipédia. Enquanto escrevia este livro, passei muitas horas felizes fazendo exatamente isso, e posso dizer que não é apenas informativo, mas também surpreendentemente divertido, uma espécie de entretenimento barato para geeks. Recomendo também dar uma boa olhada no GitHub (http://github.com). que é o locus atual mais importante para o trabalho de código aberto. Uma boa visão geral do código aberto é ***The Success of Open Source*** [235], de Steven Weber. Sua única desvantagem é que está se tornando um pouco datado (2004), mas há muito no livro que é relativamente atemporal. Indo ainda mais longe, há o famoso ensaio de Eric Raymond, “The Cathedral and the Bazaar” [178]. O ensaio de Raymond foi o que primeiro despertou meu interesse (e de muitos outros) em código aberto, e continua valendo a pena lê-lo. Os perspicazes “Coase's Penguin, or, Linux and ***The Nature of the Firm”*** [12] e ***The Wealth of Networks*** [13], de Yochai Benkler, influenciaram fortemente muitas ideias sobre código aberto, especialmente na comunidade acadêmica. Finalmente, recomendo o fascinante relato de Ned Gulley e Karim Lakhani [87] sobre a competição de programação Mathworks.

Limites da inteligência coletiva: Resumos informativos são ***Infotopia*** [212] , de Cass Sunstein, e ***The Wisdom of Crowds*** [214], de James Surowiecki. Textos clássicos incluem ***Extraordinary Popular Delusions and the Madness of Crowds [Delírios Populares Extraordinários e a Loucura das Multidões],*** de Charles Mackay , publicado pela primeira vez em 1841 e reimpresso inúmeras vezes desde então [130], e ***Groupthink*** [99], de Irving Lester Janis. É claro que uma fração considerável de nossa cultura escrita lida, direta ou indiretamente, com os desafios da resolução de problemas em grupo. Entre os relatos mais formativos para mim estão ***Skunk Works*** [184], de Ben Rich, ***The Making of the Atomic Bomb*** [183], de Richard Rhodes, e ***The Pentium Chronicles*** [45], de Robert Colwell. Um pouco mais adiante, o livro de Peter Block , ***Community: The structure of belonging*** [18], contém muitos insights sobre os problemas da construção de uma comunidade. E, finalmente, a obra-prima de Jane Jacobs, ***The Death and Life of Great American***

Cidades [98] é um relato soberbo de como grupos muito grandes abordam um problema humano fundamental: como criar um lugar para viver.

Ciência em rede, em geral: O potencial dos computadores e das redes para mudar a forma como a ciência é feita tem sido discutido por muitas pessoas, e por um longo período de tempo. Tal discussão pode ser encontrada em muitos dos trabalhos descritos acima, em particular o trabalho de Vannevar Bush [31] e Douglas Engelbart [63]. Outros trabalhos notáveis incluem os de Eric Drexler [57], Jon Udell [227], Christine Borgman [23] e Jim Gray [83]. Veja também a proposta original de Tim Berners-Lee para a World Wide Web, reimpressa em [14]. Uma representação ficcional estimulante e agradável da ciência em rede é ***Rainbows End*** [231], de Vernor Vinge.

Ciência orientada por dados: Uma das primeiras pessoas a compreender e articular claramente o valor da ciência orientada por dados foi Jim Gray, da Microsoft Research. Muitas de suas ideias estão resumidas no ensaio [83], que também mencionei acima. Esse ensaio faz parte de um estimulante livro de ensaios intitulado ***O Quarto Paradigma*** [94]. O livro pode ser baixado gratuitamente na internet e oferece uma boa visão geral de muitos aspectos da ciência orientada por dados. Outro artigo instigante é “A Eficácia Irracional dos Dados” [88], de Alon Halevy, Peter Norvig e Fernando Pereira. Os três autores trabalham para o Google, que talvez tenha a cultura mais orientada por dados de qualquer organização do mundo, e o artigo transmite bem a mudança radical de perspectiva que advém do pensamento orientado por dados. Se você tem experiência em programação, também recomendo o excelente ensaio curto de Norvig [157] sobre como escrever um corretor ortográfico (orientado por dados, naturalmente!) em apenas 21 linhas de código. Existem muitos, muitos textos e artigos sobre tópicos relacionados à inteligência orientada por dados. (Observe, porém, que a maioria não usa o termo.) Uma boa introdução prática é ***Programming Collective Intelligence*** [191], de Toby Segaran.

A democratização da ciência e da ciência cidadã: A democratização da ciência tem análogos no mundo dos negócios, em fenômenos como a inovação gerada pelo usuário e modelos de inovação aberta para negócios. Veja, por exemplo, o livro de Eric von Hippel, ***Democratizing Innovation*** [233], cujo título inspirou o título do capítulo 7, e ***Open Innovation*** [36], de Henry Chesbrough. O ponto de vista desenvolvido no capítulo 7 também deve muito à noção de Clay Shirky de que nossa sociedade possui um excedente cognitivo [195, 194; veja também 196] que pode ser usado a serviço de novas formas de ação coletiva.

Ciência aberta: A minha análise da ciência aberta é fortemente influenciada pelo trabalho de Mancur Olson [161] sobre a acção colectiva e pelo trabalho de

Elinor Ostrom [165] sobre a gestão de recursos de uso comum, como pesca e florestas. Ambos os trabalhos têm muito mais implicações para a ciência aberta do que descrevi. Em particular, abordei apenas brevemente muitos dos princípios detalhados que Ostrom identifica para a gestão de recursos de uso comum. Muitos desses princípios podem ser aplicados ou adaptados de forma proveitosa à ciência aberta. Também fui estimulado pelo trabalho de Robert Axelrod [9] sobre as condições sob as quais as partes cooperarão; o problema da cooperação em larga escala é um exemplo de um problema de ação coletiva. Sobre a história inicial da ciência aberta, fui estimulado por muitas fontes, mas especialmente por Paul David [49], Elizabeth Eisenstein [61] e Mary Boas Hall [89].

Uma coisa que me incomodou ao escrever este livro foi que as restrições narrativas me obrigaram a omitir quase todos os ***milhares*** de projetos de ciência aberta em andamento. Felizmente, existem muitas fontes excelentes para acompanhar o que está acontecendo na ciência aberta hoje. Deixe-me mencionar apenas algumas. Uma das mais valiosas é o site de Peter Suber (http://www.earlham.edu/~peters/hometoc.htm). que é um recurso tremendo sobre todos os aspectos da ciência aberta, mas especialmente sobre publicação em acesso aberto. Blog excelente (http://www.earlham.edu/~peters/fos/fosblog.html) não é mais atualizado, mas continua sendo um recurso histórico valioso. E o Boletim de Acesso Aberto de Suber (http://www.earlham.edu/~peters/fos/newsletter/archive.htm) é essencial.

Suber's

Outra excelente fonte sobre ciência aberta é o blog de Cameron Neylon (http://cameronneylon.net/). Neylon é um dos pioneiros da ciência de cadernos abertos e tem muitas coisas estimulantes a dizer sobre ciência aberta em geral. Você também pode encontrar muitos cientistas e projetos de ciência aberta usando serviços como Twitter e FriendFeed. Uma boa maneira de entrar nesse mundo é usar o Google para pesquisar "ciência aberta no Twitter".

Além dessas pessoas, existem muitas organizações que trabalham pela ciência aberta. A Aliança para o Acesso do Contribuinte (http://www.taxpayeraccess.org/) pressionou o governo dos EUA por políticas de acesso aberto a artigos e dados científicos. Por exemplo, foi em parte por meio desse lobby que surgiu a política de acesso aberto do NIH, descrita no capítulo 7. Outras organizações que trabalham nesse sentido incluem o Science Commons Open (http://sciencecommon.sorg), que faz parte da organização Creative Commons e da Open Knowledge Foundation (http://ciência.org).

O desafio de criar uma cultura mais aberta não se limita à ciência. Ele também está sendo enfrentado na cultura em geral. Pessoas como Richard Stallman [202], Lawrence Lessig [122] e muitos outros descreveram os benefícios que a abertura traz em um mundo em rede. Eles desenvolveram ferramentas como o licenciamento Creative Commons (http://creativecommons.org). e licenças “copyleft” para ajudar a promover uma cultura mais aberta. Meu pensamento foi especialmente influenciado por Lessig [122]. No entanto, embora a ciência aberta tenha muitos paralelos com o movimento da cultura aberta, a ciência enfrenta um conjunto único de forças que inibem o compartilhamento aberto. Isso significa que ferramentas como as licenças Creative Commons, que têm sido tremendamente eficazes na transição para uma cultura mais aberta, não abordam diretamente o principal desafio subjacente na ciência: o fato de que os cientistas são recompensados pela publicação de artigos, e não por outras formas de compartilhar conhecimento. Portanto, embora a ciência aberta possa aprender muito com o movimento da cultura aberta, ela também requer uma nova forma de pensar.

Notas

Algumas das referências a seguir incluem páginas da web cujos URLs podem expirar após a publicação deste livro. Essas páginas da web podem ser recuperadas usando o Wayback Machine do Internet Archive (http://www.archive.org/web/web.php). Fontes online geralmente são escritas informalmente, e reproduzi erros de ortografia e outros erros literalmente ao citar tais fontes.

Capítulo 1. Reinventando a Descoberta

p 1: Gowers propôs o Projeto Polymath em uma postagem em seu blog [79]. Para mais informações sobre o Projeto Polymath, veja [82].

p 2: Anúncio de Gowers sobre o provável sucesso do primeiro Projeto Polymath: [81]. ***p 2 O processo***

Polymath era "para a pesquisa normal o que dirigir é para empurrar um carro": [78]. ***p 3:*** O termo

inteligência coletiva foi introduzido pelo filósofo Pierre Lévy [124]. Uma tentativa recente e estimulante de medir a inteligência coletiva e relacioná-la às qualidades dos participantes do grupo é [243]. ***p 3 o processo da ciência irá... mudar mais nos próximos vinte anos do que nos***

últimos 300 anos: o autor Kevin Kelly fez uma afirmação semelhante em [108] (veja também [109]): "Haverá mais mudanças nos próximos 50 anos da ciência do que nos últimos 400 anos." Há uma ampla sobreposição entre meu raciocínio e o de Kelly, por exemplo, ambos enfatizamos a importância da colaboração e da coleta de dados em larga escala. Há também algumas diferenças consideráveis em nosso raciocínio, por exemplo, Kelly enfatiza mudanças como experimentos triplo-cegos e mais prêmios em ciências, enquanto eu acredito que estes desempenharão um papel relativamente menor na mudança, e que as três áreas a seguir são as mais críticas:

(1) inteligência coletiva e ciência orientada por dados, e a maneira como elas mudam a forma como a ciência é feita; (2) a relação mutável entre ciência e sociedade; e (3) o desafio de alcançar uma cultura científica muito mais aberta. ***p 4:*** GenBank está em http://www.ncbi.nlm.nih.gov/genbank/. O ser humano no genoma http://www.ncbi.nlm.nih.gov/projects/genome/assembly/grc/human/index.shtml, e o mapa de haplótipos está disponível em http://hapmap.ncbi.nlm.nih.gov/.

p . 7: Um relato em primeira mão da reunião das Bermudas, incluindo uma declaração do Acordo das Bermudas, pode ser encontrado em [211]. A declaração Clinton-Blair sobre o compartilhamento de dados genéticos não menciona explicitamente o Acordo das Bermudas, mas os princípios defendidos são essencialmente os princípios acordados nas Bermudas. A declaração pode ser encontrada em [102].

p . 7: Usei o Acordo das Bermudas como exemplo de um acordo coletivo que impulsiona o compartilhamento de dados. De fato, a quantidade de dados genéticos depositados no GenBank dobrou aproximadamente a cada 18 meses desde a sua fundação, e essa tendência não foi significativamente acelerada pelo Acordo das Bermudas. Você pode se perguntar se o Acordo das Bermudas foi realmente tão importante para o aumento do compartilhamento de dados. É claro que parte do aumento no compartilhamento de dados se deve a uma melhor tecnologia de sequenciamento. Mas o aumento também se deve, em parte, a um amplo impulso da comunidade biológica para compartilhar dados com mais liberdade. O Acordo das Bermudas é apenas parte desse amplo movimento, embora talvez seja a manifestação mais visível.

p 7: Sobre as extensões do Acordo das Bermudas, ver especialmente o Acordo de Fort Lauderdale [237]. ***p 7:*** Sobre a partilha de dados sobre a gripe, ver por exemplo [20] e [60] sobre o Acordo das Bermudas [237]. surto de gripe aviária de 2006 e [32] sobre a pandemia de gripe suína de 2009-2010.

p 10 Vivemos na época da transição para a segunda era da ciência: uma afirmação relacionada foi feita pelo pesquisador de bancos de dados Jim Gray [83] (ver também o volume em que o ensaio de Gray aparece [94]). Gray afirmou que estamos entrando hoje no que ele chama de um “quarto paradigma” da descoberta científica, um paradigma baseado em uma ciência altamente intensiva em dados, na qual os computadores nos ajudam a encontrar significado nos dados. Na descrição de Gray, esse quarto paradigma é uma extensão do que ele chama de primeiro paradigma (observação empírica), segundo paradigma (a formação de modelos para explicar a observação) e terceiro paradigma (o uso de simulação para compreender fenômenos complexos) da ciência. É verdade que a ciência intensiva em dados é importante, e a discutiremos no capítulo 6. Mas a concepção de Gray sobre a mudança atual na ciência é muito limitada. A ciência envolve muito mais do que apenas encontrar significado em dados. Trata-se também das maneiras pelas quais os cientistas trabalham juntos para construir conhecimento e como a comunidade científica se relaciona com a sociedade como um todo. Esses aspectos da ciência também estão sendo transformados por ferramentas online. Além disso, cada uma dessas mudanças impacta e reforça as demais. Assim, por exemplo, para realmente compreender o impacto da ciência intensiva em dados, precisamos compreender as mudanças na forma como os cientistas trabalham.

juntos. O quarto paradigma de Gray é apenas parte das mudanças que estão sendo promovidas pela ciência em rede.

Capítulo 2. Ferramentas online nos tornam mais inteligentes

p 15: Meu relato de Kasparov versus o Mundo é baseado principalmente no livro de Kasparov (com Daniel King) [107] e no relato do jogo de Irina Krush (com Kenneth Regan) [115].

p 15 "a maior partida da história do xadrez": de uma entrevista da Reuters com Kasparov conduzida durante a partida [186], no lance número 37. Faz parte de um comentário mais longo e interessante de Kasparov: "É a maior partida da história do xadrez. O grande número de ideias, a complexidade e a contribuição que fez ao xadrez fazem dela a partida mais importante já jogada." ***p 19:*** James Surowiecki, ***The Wisdom of Crowds,*** [214]. ***p 20:*** O livro de Nicholas Carr ***The Shallows*** [35] é uma versão expandida de um artigo anterior, "Is Google Making Us Stupid?" [34]. Argumentos relacionados também foram feitos por Jaron Lanier [117].

Capítulo 3. Reestruturação da Atenção Especializada

p 22: Sobre a ASSET Índia, InnoCentive e Zacay Brown: [29, 222]. O texto sobre a InnoCentive é uma versão muito expandida e adaptada do texto do meu artigo [153]. ***p 23 Muitos dos solucionadores bem-sucedidos relatam, como Zacary Brown, que os desafios que resolvem correspondem intimamente às suas habilidades e interesses:*** veja [116] para mais informações sobre as características dos solucionadores bem-sucedidos. Observe que este estudo também descobriu que as pessoas frequentemente resolvem desafios que estão nominalmente fora de seu domínio de especialização. Um químico pode, por exemplo, resolver um problema de biologia.

Isso parece uma contradição à afirmação sobre uma correspondência próxima à expertise, mas não é: a principal dificuldade em resolver o problema biológico pode residir em uma expertise muito específica da química. Portanto, quando se analisa as soluções do Desafio em um nível mais detalhado, a correspondência com a expertise costuma ser excepcionalmente próxima.

p 24 É porque Zacary Brown tem uma vantagem comparativa tão enorme que ele e a ASSET podem trabalhar juntos para benefício mútuo:

“vantagem comparativa” é um termo técnico da economia, e estou usando o termo nesse sentido. Em outros lugares, quando falo de pessoas aplicando seus conhecimentos da “melhor” maneira possível (ou linguagem similar), quero dizer da melhor maneira possível no sentido de maximizar a vantagem comparativa, não a vantagem absoluta. ***p. 24:*** O caráter crítico da

atenção humana como um recurso escasso em um mundo rico em informações foi apontado em um artigo premonitório de Herbert Simon [197]. Um trabalho especulativo impressionante sobre a economia da atenção é o artigo de Michael Goldhaber [75]. Veja também [151]. ***p. 27:*** Em relação ao termo

“serendipidade projetada”, Jon Udell usou o termo “serendipidade fabricada” para descrever um conceito semelhante em [228]. Usei “serendipidade projetada” em vez disso porque enfatiza a maneira como a serendipidade pode ser alcançada como resultado de escolhas deliberadas de design. A ideia de serendipidade projetada parece ter se originado no movimento do software de código aberto e foi sucintamente capturada na observação de Eric Raymond [178] de que, ao depurar software de código aberto, “com olhos suficientes, todos os bugs são superficiais”. Raymond apelidou essa observação de Lei de Linus, em homenagem ao criador do Linux, Linus Torvalds. Podemos generalizar a Lei de Linus para outras formas de resolução de problemas: “Com olhos suficientes, todos os problemas são fáceis”. Não é literalmente verdade, mas captura algo da essência da serendipidade projetada. ***p. 27 “Grossmann, você precisa me ajudar, senão eu enlouqueço!”:*** a história de Einstein-Grossmann

é contada na íntegra em [169]. ***p. 30:*** A discussão sobre massa crítica conversacional é inspirada em parte por

capítulo 3 de [189].

p 30 Os participantes do Polymath frequentemente “se encontravam tendo pensamentos que não teriam tido sem algum comentário casual de outro colaborador”: [80].

p 31: Sobre o valor da diversidade cognitiva, veja, por exemplo, o trabalho de Scott Page [168] e Friedrich von Hayek [93]. ***p 32:*** A frase

“arquitetura da atenção” é inspirada na elegante frase de Tim O'Reilly “arquitetura da participação” [162]. O'Reilly usa seu termo “para descrever a natureza dos sistemas que são projetados para a contribuição do usuário”. Estamos interessados em sistemas projetados para a resolução criativa de problemas e, em tais sistemas, é a alocação de atenção especializada que é mais crucial. ***p 34:*** O número de funcionários no ***Avatar***

é de [65]. ***p 36:*** A descoberta do bóson Z em 1983 é descrita em

[4]. ***p 37 “quem é responsável pelo fornecimento de pão para a população de Londres?”:*** ver ***The Company of Strangers*** [190] de Paul Seabright .

p 37 O que torna os preços úteis é que... eles agregam uma enorme quantidade de conhecimento oculto: [93].

p 38: A “pergunta idiota” foi feita pelo participante do Polymath Ryan O'Donnell: [159].

p 39: Sobre o ponto de que as ferramentas online estão subsumindo e expandindo tanto os mercados convencionais quanto as organizações convencionais: um ponto relacionado foi levantado pelo teórico Yochai Benkler em seu artigo “O Pinguim de Coase, ou Linux e ***a Natureza da Firma*** [12]”. Benkler tem um foco diferente, preocupando-se não tanto com a solução de problemas criativos, mas com a produção de bens. Ele propõe que a colaboração online possibilitou uma terceira forma de produção, além dos mercados e das organizações convencionais, que ele chama de “produção entre pares”. Acredito que este seja um ponto de vista muito restrito, tanto para a resolução criativa de problemas quanto para a produção de bens. Ferramentas online podem ser usadas para ***subsumir*** tanto mercados quanto organizações convencionais como casos especiais, e também possibilitar muitas novas formas de produção e resolução criativa de problemas.

Portanto, não é que agora tenhamos uma terceira forma de produção. É que agora temos um meio de produção que inclui todas as nossas formas anteriores como casos especiais e possibilita novas formas.

Capítulo 4. Padrões de colaboração online

p 44: Relatos esclarecedores sobre o desenvolvimento de software de código aberto incluem [12, 13, 178, 235]. Ainda mais úteis são os inúmeros projetos de código aberto mantidos online em sites como o GitHub (http://github.com). e SourceForge (http://sourceforge.net).

p 44: Minha história sobre o Linux baseia-se, em grande parte, em postagens nos grupos de notícias comp.os.minix, alt.os.linux e comp.os.linux em 1991 e 1992. Achei a leitura desses fóruns surpreendentemente agradável, e até mesmo envolvente: à medida que se lê, começa-se a ter uma noção visceral do que foi necessário para produzir uma maravilha do software moderno. Meu relato sobre o Linux também foi amplamente influenciado por [235], bem como por muitas outras fontes para detalhes (veja abaixo). ***p 45 Logo após a postagem de Torvalds...*** postagem no grupo de notícias comp.os.minix, 13 de janeiro de 1992.

p 45 80 pessoas foram nomeadas como contribuidoras no arquivo Linux Credits: Veja [226] para o histórico do arquivo Credits. Março de 1994 é a primeira vez que tal arquivo foi incluído no Linux. ***p 45 No início de 2008, o kernel Linux . .*** ***p 45:*** Sobre o papel do · : [114].

Linux nas empresas de animação e efeitos visuais de Hollywood, veja [90] para um relato de 2002, época em que o Linux estava entrando na indústria e começando a dominar. [187] afirma que em 2008, o Linux era usado em “mais de 95% dos servidores e desktops em grandes empresas de animação e efeitos visuais”. ***p 45 Projetos de software de código aberto têm dois atributos principais:*** Alguns defensores do código aberto preferem uma descrição mais matizada do código aberto do que

descrição que dei. Muitas discussões complexas e, às vezes, acaloradas têm ocorrido sobre quais projetos devem ser considerados verdadeiramente de código aberto. De fato, uma organização sem fins lucrativos chamada Open Source Initiative existe, em parte, para decidir se um projeto deve ser rotulado como código aberto e, em caso afirmativo, para fornecer certificação. Visto de fora, isso pode parecer uma crítica pedante, mas há boas razões para isso. O código aberto às vezes é visto como uma ameaça a algumas grandes empresas de software: por exemplo, Linus Torvalds disse certa vez no ***New York Times:*** "Não estou aqui para destruir a Microsoft. Isso será apenas um efeito colateral completamente não intencional" [52]. Algumas das empresas ameaçadas pelo código aberto reagiram tentando quebrar a marca do código aberto, lançando produtos que chamam de "código aberto", mas sem recursos cruciais encontrados em projetos verdadeiramente de código aberto. Em maio de 2001, o vice-presidente sênior da Microsoft, Craig Mundie [142], anunciou que a Microsoft lançaria alguns produtos como "Código Compartilhado", afirmando que "Código Compartilhado é Código Aberto". Uma análise mais atenta das licenças do Microsoft Shared Source mostra que elas são fortemente direcionadas aos usuários de produtos Microsoft e, em alguns casos, impedem que os programadores modifiquem o código. Isso certamente não é código aberto! Esse tipo de incidente mostra por que os defensores do código aberto às vezes se irritam quando as pessoas usam o termo "código aberto" de forma desleixada. Adotaremos uma abordagem mais relaxada que, acredito, aborda a essência do código aberto, mas sem nos prendermos às complexidades de saber se os projetos que descrevemos passariam em todos os rigorosos testes exigidos por alguns defensores do código aberto.

p 46: O número de 4.300 linhas de código adicionadas ao kernel Linux por dia é de uma palestra informativa sobre o processo de desenvolvimento do kernel Linux, por Greg Kroah-Hartman [113]. ***p 46 um desenvolvedor experiente normalmente escreverá af: para insousan linhas de código por ano:*** esta estimativa é baseada no modelo de software COCOMO II [19]. ***p 46*** SourceForge abriga mais de 230.000 projetos ***de código aberto :*** [239]. ***p 46 código aberto é uma metodologia geral de design que pode ser aplicada a qualquer projeto que envolva informações digitais:*** A metodologia de código aberto também pode ser aplicada a informações não digitais. Você poderia imaginar, por exemplo, usar plantas impressas de arquitetos como base para o design de edifícios de código aberto. O problema com informações analógicas é que elas tendem a se degradar à medida que são copiadas repetidamente, o que limita sua utilidade para a metodologia de código aberto.

p 46 Rede de Arquitetura Aberta: http://www.openarchitecturenetwork.org. A Open Architecture Network foi apresentada em uma palestra de Cameron Sinclair: [198].

p 48:

Sobre biologia de código aberto, veja, por exemplo, o capítulo 13 de [33]. ***p 49:*** Meu relato sobre a bifurcação próxima do Linux é baseado principalmente no online Lista de discussão do kernel Linux, com algumas informações adicionais de [235].

p 51: Sobre a dificuldade de tornar o desenvolvimento de código aberto modular, um comentário que às vezes ouço de não programadores interessados em código aberto é que a programação é “naturalmente modular”.

equívoco e parece basear-se numa confusão terminológica. É verdade que muitas linguagens de programação incentivam uma estrutura modular no desenvolvimento e, para programas ***pequenos,*** isso facilita o design modular. Mas, para sistemas de grande escala, como o Linux, modularidade significa algo bem diferente e é muito mais difícil de alcançar. Sistemas de grande escala não são mais naturalmente modulares do que uma pintura, porque a tinta é construída a partir de unidades modulares (moléculas). Em vez disso, a modularidade na engenharia de software em larga escala requer um design inteligente por meio de vários níveis de abstração, o que, por sua vez, requer um forte compromisso com o princípio da modularidade por parte dos desenvolvedores. ***p 52:*** Para Linus Torvalds sobre modularidade, veja [223]. ***p 53:*** Pinguins

O blog http://www.amillionpenguins.com/blog/, e Milhão é no possui links para outros recursos associados ao projeto Million Penguins, incluindo o wiki usado para escrever o romance. Soube do projeto por meio de [139], que publicou o mesmo trecho do romance que usei. ***p. 55:*** O rastreador de problemas online do Firefox pode ser encontrado em http://bugzilla.mozilla.org.

p 55: O bug do favicon no Firefox https://bugzilla.mozilla.org/show_bug.cgi?id=411966. é descrito em ***p 56 O rastreador de problemas não serve apenas para corrigir bugs, ele também é usado para propor e implementar novos recursos:*** na verdade, o rastreador de problemas é apenas uma das várias maneiras pelas quais os desenvolvedores do Firefox podem propor novos recursos. Outros fóruns usados para propor novos recursos incluem uma lista de discussão online, um wiki e até mesmo uma teleconferência semanal com desenvolvedores do Firefox. ***p 58 Mais de um bilhão de linhas:*** Isso e a estimativa da taxa de crescimento do código são estimativas conservadoras, baseadas no trabalho de Deshpande e Riehle [51], atualizado até o final de 2006.

p 58: A história de Alan Kay sobre Donald Knuth está na página 101 de [192]. ***p 59 “Bons programadores codificam; grandes programadores reutilizam o código de outras pessoas”:*** Variantes desse ditado circulam pelo mundo do código aberto há anos, mas não consegui rastrear a fonte original. Isso é apropriado.
Há mais: a citação é uma paráfrase de uma citação frequentemente atribuída a Picasso: “Bons artistas copiam; grandes artistas roubam”. Não consegui encontrar uma fonte verificável para a citação de Picasso, mas compare com “Poetas imaturos imitam; poetas maduros roubam” de T. S. Eliot [62].

p 59 Para mais informações sobre a competição MathWorks, veja [87] e, especialmente, [88]. ***p 61 “Comecei a ficar 'obcecado' ”:*** [86]. ***p 63: Um estudo realizado por dois cientistas da empresa de software SAP, Oliver e Dirk Riehle*** . . ***Arafat . :***[3]. ***p 63:*** Você pode concluir da discussão sobre microcontribuição que o software de código aberto é construído principalmente a partir de pequenas contribuições. Mas só porque pequenas contribuições são mais frequentes não significa que elas constituem

a maior parte do produto final. Pode ser que algumas contribuições grandes sobrecarreguem as muitas contribuições menores. E, de fato, em muitos projetos de código aberto é isso que acontece: pequenas mudanças são as mais frequentes, mas o produto final ainda é dominado por pedaços relativamente grandes de código. É tentador, então, reverter a direção e concluir que pequenas contribuições não são tão importantes, que na verdade são uma distração. Mas isso também está errado, um pouco como argumentar que ***Hamlet*** seria uma peça melhor sem tudo, exceto os grandes solilóquios. Tanto as contribuições grandes quanto as pequenas são cruciais. As grandes contribuições importam pelo motivo óbvio, e as pequenas contribuições importam porque elas fazem a conversa avançar e ajudam a colaboração a explorar uma gama mais ampla de ideias. É das melhores dessas ideias que surgem as grandes contribuições.

p 66 uma colaboração precisa saber o que , A aration sabe: Esta observação, muitas vezes sob diferentes formas, parece ter sido feita muitas vezes. Eu a apreciei plenamente pela primeira vez depois de ler [28]. ***p***

67 "Se alguma coisa na minha vida da qual participei · . .": Esta citação é de um comentário feito pelo comentarista AdmiralBumblebee [30] no site reddit.
Vale mencionar que o comentário foi motivado por uma versão inicial do material que abre o capítulo 2 deste livro, que, segundo a AdmiralBumblebee, refletia "hype comercial" e a visão de um patrocinador sobre o jogo. Meu relato, no entanto, não se baseia em informações do patrocinador, a Microsoft, mas principalmente nos relatos em primeira mão de Kasparov e Krush, corroborados por diversas outras fontes.

Capítulo 5. Os Limites e o Potencial da Inteligência Coletiva

p 69: Os experimentos de Stasser-Titus são descritos em [204], que contém muito mais detalhes do que meu relato resumido. Uma revisão do trabalho que acompanha esses experimentos encontra-se em [203]. Um resumo mais amplo e informativo sobre como a inteligência coletiva pode falhar encontra-se no livro de Sunstein, ***Infotopia*** [212].

p 75 os jogadores mais fortes da Equipe Mundial geralmente conseguiam concordar sobre quais análises eram melhores: Houve uma exceção significativa a isso, que é que no início do jogo a Microsoft pediu aos conselheiros da Equipe Mundial que não consultassem uns aos outros, e assim eles não tiveram a oportunidade de chegar a um acordo.
Mas muitos dos jogadores mais fortes da Equipe Mundial estavam em contato próximo e frequentemente conseguiam chegar a um acordo. ***p***

78: Sobre os limites da inteligência coletiva e problemas como pensamento de grupo, cascatas de informação, etc., veja [99, 212, 213, 214] e referências aí citadas.

p. 79: Em relação à rápida aceitação das ideias de Einstein, o fato de cientistas renomados como Lorentz e Poincaré terem chegado a conclusões semelhantes quase ao mesmo tempo ajudou. Mas, embora a formulação da relatividade de Einstein fosse ainda mais radical do que as formulações de Lorentz e Poincaré, ela rapidamente foi aceita como a maneira correta de pensar a relatividade.

p 79: Sobre a descoberta do DNA e o erro de Pauling, veja as memórias de Watson, ***A Dupla Hélice*** [234].

p 80 "Se Feynman diz isso três vezes, está certo": [72]. ***p 84:***

Meus agradecimentos a Mark Tovey pela ajuda na construção deste exemplo em ilusões de ótica e ciência cognitiva.

p 85: Sobre mercados de colaboração, veja também [246] e [146]. ***p 85:*** A discussão sobre computações quânticas topológicas é inspirada em [22]. Os computadores quânticos topológicos foram originalmente propostos em um artigo notável de Kitaev [111].

Capítulo 6. Todo o Conhecimento do Mundo

p 91: A descoberta de Swanson da conexão magnésio-enxaqueca é descrita em [215] e revisada em [216]. ***p 92:*** Uma questão interessante sobre a conexão enxaqueca-magnésio é por que ela não foi descoberta por, digamos, cientistas trabalhando em epilepsia, alguns dos quais estavam presumivelmente cientes da conexão da epilepsia com enxaquecas e deficiência de magnésio. Especulando, parece provável que a razão pela qual essa conexão passou despercebida é que (1) esses cientistas estavam focados principalmente em entender a epilepsia, não outras condições; e (2) uma única conexão ligando enxaquecas e deficiência de magnésio não é um padrão suficiente para inferir qualquer coisa. A epilepsia está conectada a muitas condições diferentes, a maioria das quais não tem relação direta entre si.

p 92: Sobre o procedimento de Swanson, é claro que não há nada de novo em inferir conhecimento desconhecido a partir do conhecimento científico existente. É uma prática padrão em áreas como a minha, a física teórica. Mas a aplicação sistemática dessa ideia mediada por computador por Swanson na medicina era nova e prenunciava uma explosão no uso de técnicas semelhantes de mineração de dados em muitas áreas da ciência. ***p 93:*** A noção de mente estendida foi discutida em [43]. ***p 93:*** O artigo que descreve o uso de consultas de pesquisa do Google para rastrear a gripe é [71]. ***p 93:*** As taxas anuais de mortalidade por gripe são da Organização Mundial da Saúde.

Organização [244]. ***p 93:*** A taxa de mortalidade da gripe espanhola é de [219].

p 94: O site Google Flu Trends é http://www.google.org/flutrends. ***p 94:*** O sistema CDC/General Electric para rastrear a gripe é descrito em [136]. ***p 94:*** O estudo de acompanhamento mostrando que o Google Flu Trends é melhor em rastrear doenças semelhantes à gripe do que em rastrear casos de gripe confirmados em laboratório é [163]. ***p 94:*** Sobre o uso de consultas de pesquisa para prever o desemprego, veja [6]. Sobre o uso de consultas de pesquisa para prever preços de imóveis, veja [245]. Sobre o uso de consultas de pesquisa para ajudar a melhorar as previsões de quão bem as músicas se sairão nas paradas, veja [73]. Para uma ampla gama de aplicações, veja [42]. Um estudo usando o Twitter para prever a receita de bilheteria de filmes é [7]. Por fim, veja [11] para uma discussão instigante do Google como um "banco de dados de intenção [humana]".

p 95: Para Eric Schmidt sobre privacidade, veja [64]. ***p 96:*** A frase "desconhecidos conhecidos" foi sugerida neste contexto por Jen Dodd e Hassan Masum, inspirados pelo famoso uso [188] do antigo Secretário de Defesa dos EUA, Donald Rumsfeld, de "desconhecidos desconhecidos". ***p 97:*** A descoberta da Grande Muralha de galáxias de Sloan é descrita em [77].
As galáxias da Grande Muralha de Sloan não parecem estar gravitacionalmente unidas e, por isso, alguns astrofísicos não as consideram uma estrutura única.
No entanto, grande parte da história contada nesta seção se estende a várias outras características de grande escala do universo — minha escolha da Grande Muralha de Sloan foi um tanto arbitrária.

p 100: A descoberta de muitas galáxias anãs próximas à Via Láctea foi descrita em vários artigos. Para um visão geral, ver http://www.sdss.org/signature.html.

p 100: A descoberta dos buracos negros em órbita foi descrita em [25]. No texto, afirmo que Boroson e Lauer pesquisaram imagens de galáxias do SDSS. Para ser um pouco mais preciso, eles pesquisaram uma seleção de 17.500 quasares, um tipo especial de galáxia conhecido por conter buracos negros supermassivos. Para mais informações sobre o que são quasares e por que são interessantes, veja a descrição na página 130. Observe que houve considerável discussão subsequente na comunidade de astronomia e astrofísica sobre se a descoberta em [25] é, de fato, de um par de buracos negros em órbita, ou talvez algo mais. Esta conversa está em andamento. ***p 101:*** O Sloan Digital Sky Survey foi descrito em [247]. Os números de citação para este artigo são do serviço Google Acadêmico. Os números são conservadores, uma vez que não incluem citações de lançamentos de dados subsequentes e muitos outros artigos importantes do SDSS. ***p 102:*** O SDSS codificou muitas de suas políticas sobre colaboração e compartilhamento de dados em http://www.sdss.org/collaboration/. É uma leitura surpreendentemente estimulante.

p 102: O SDSS SkyServer está em http://skyserver.sdss.org.

p 104: Sobre Watson, Crick e Franklin, veja as memórias de Watson, ***The Double Helix*** [234]. ***p 105:***

A página da web para o estágio III do SDSS está em http://www.sdss3.org. ***p 105:*** Meu relato sobre a Iniciativa ***de Observatórios Oceânicos baseia-se no site http://www.oceanleadership.org/programs-and-project partnerships/ocean-observing/ooi/ e [50]. p 106:*** Mapear o cérebro é um

assunto muito amplo para que eu possa fornecer uma lista abrangente de referências. Uma visão geral do trabalho sobre o Atlas Cerebral de Allen pode ser encontrada no excelente artigo de Jonah Lehrer [120]. A maioria dos fatos que relato são desse artigo. O artigo que anuncia o atlas da expressão gênica no cérebro de camundongos é [121]. Visões gerais de alguns dos progressos e desafios no mapeamento do conectoma humano podem ser encontradas em [119] e [125].

p 108: Bioinformática e quimioinformática são campos já bem estabelecidos, com uma literatura significativa, e não tentarei destacar nenhuma referência específica para menção especial. A astroinformática surgiu mais recentemente. Veja especialmente [24] para um manifesto sobre a necessidade da astroinformática.

p 113: Um relatório sobre o torneio de xadrez estilo livre Playchess.com de 2005 pode ser encontrado em [37], com comentários subsequentes sobre os vencedores em [39]. Os comentários de Garry Kasparov sobre o resultado estão no fascinante artigo [106], que contém muito interesse sobre o assunto de computadores e xadrez. Comentários adicionais sobre o envolvimento da Hydra podem ser encontrados em [38]. Curiosamente, a Hydra jogou e perdeu duas vezes em partidas de xadrez por correspondência, contra o grande mestre de xadrez por correspondência Arno Nickel. Nickel foi, no entanto, autorizado a usar programas de xadrez de computador nessas partidas. Um registro completo das partidas da Hydra pode ser encontrado em [40]. ***p 119:*** O livro de Chuck

Hansen é [92]. A história que conto sobre Hansen

a metodologia é contada no livro de Richard Rhodes ***How to Write,*** [182], página 61. ***p 120:***

Sobre a web semântica, veja [16, 15] e http://www.w3.org/standards/semanticweb/. Um ponto de vista alternativo estimulante é [88]. ***p 120:*** Para o memorando de Obama sobre transparência e governo

aberto, veja [158]. ***p 123:*** O belo resumo da teoria geral da relatividade de Einstein, “O espaço-

tempo diz à matéria como se mover; a matéria diz ao espaço-tempo como se curvar”, é devido a John Wheeler [240]. ***p 125 esses modelos não têm compreensão do significado de “hola” ou “olá”:*** eu uso o termo

“compreensão” aqui em seu sentido cotidiano. Suspeito, porém, que um dia descobriremos que o que queremos dizer com “compreensão” é capturado em parte (mas apenas em parte) pelo tipo de associação estatística nesses modelos. ***p 125 ninguém na equipe do Google Translate falava chinês ou árabe:*** [69]. ***p 128:*** O comentário de Planck “Eu realmente não pensei

muito nisso [a teoria quântica]” é do artigo de Helge Kragh [112].

Capítulo 7. Democratizando a Ciência

p 129: Meu relato do Galaxy Zoo é baseado no blog do Galaxy Zoo, http://blogs.zooniverse.org/galaxyzoo/, Fórum do Galaxy Zoo, http://www.galaxyzooforum.org, e um artigo de Chris Lintott e Kate Land [127]. O material sobre Hanny's Voorwerp também se baseia no blog de Hanny van Arkel http://www.hannysvoorwerp.com/, e o tópico de discussão original iniciado por Hanny van Arkel [67]. O primeiro artigo do Galaxy Zoo sobre o voorwerp é [128]. ***p 131:*** A explicação alternativa do voorwerp é dada em
[105, 177].
Alguns comentários sobre a explicação alternativa do cofundador e tratador do Galaxy Zoo, Chris Lintott, podem ser encontrados em [126].

p 135: O relato de Alice Sheppard sobre a descoberta das galáxias ervilhas verdes está em [193]. Observe que as imagens de galáxias vistas pelos zoóitos estão em cores falsas, e as “ervilhas verdes” estão na verdade mais
próximas do vermelho. ***p 138:*** Um artigo curto e agradável sobre a descoberta
do hélio está em [118]. ***p 141:*** A citação de Bob Nichol, “Posso fazer a pergunta 'quantas galáxias têm uma barra no meio delas' e normalmente eu embarcaria em uma busca ao longo da carreira para responder a essa pergunta fundamental...”, é de [149].

p 143: Foldit está em http://fold.it. Boas visões gerais do Foldit são [46, 21]. ***p 147:*** Para Aotearoa no Foldit, veja [1] e [2]. ***p 148:***
Os resultados do Foldit para o CASP de 2008 estão em [174].

p 149: Sobre a descoberta do grande cometa por John Caister Bennett em 1968, veja [104].

p 149: Página inicial do caçador de cometas Rainer Kracht, em http://www.rkracht.de/, tem uma lista de cometas que descobriu. Informações sobre o sucesso do SOHO na caça a cometas podem ser encontradas em http://sungrazer.nrl.navy.mil/.

p 150: O site do eBird está em http://ebird.org, e o projeto é descrito em [210]. As informações sobre o número de contribuições e contribuidores são de http://www.avianknowledge.net/content/datasets e [209]. ***p 150:*** O
projeto dinossauro aberto está em http://opendino.wordpress.com/. Uma visão geral do projeto pode ser encontrada em [220]. ***p 151:***
O uso de dados do Galaxy Zoo para treinar um algoritmo de computador é descrito em [10].

p 153: A análise da Wikipédia feita por Clay Shirky apareceu em [195]. Esse artigo também é a origem da expressão “excedente cognitivo”. Shirky desenvolveu essas ideias em um livro em [194]. ***p***
153 Em média, os americanos assistem cinco horas de televisão por dia:
[156]. ***p 154:*** A ideia de Clay Shirky de fazer “grandes coisas por amor” é desenvolvida detalhadamente em seu perspicaz livro ***Here Comes Everybody*** [196]. A citação “Nós

acostumado a um mundo onde pequenas coisas acontecem por amor e grandes coisas acontecem por dinheiro...” é da página 104 daquele livro.

p 155 “a minha vida mudou para sempre ...”:[132]. ***p***

155: Os dados de 1988 sobre a incidência da poliomielite são de [141]. ***p***

155: Os dados sobre a incidência da poliomielite em 2003 são do Programa Mundial de Erradicação da Poliomielite Relatório anual de 2003 da iniciativa, disponível em http://polioeradication.org.

p 155: O boicote nigeriano ao programa de vacinação contra a poliomielite é descrito em [101].

p 156: Uma revisão da literatura sobre a conexão entre vacinas e autismo é [68]. As evidências nesta revisão sugerem fortemente que não há ligação causal. ***p 156:*** Os números sobre as taxas de vacinação para sarampo-caxumba-rubéola e a taxa de infecção por sarampo são de [135], com base em dados da Agência de Proteção à Saúde.

p 160: O melhor recurso sobre acesso aberto é o notável blog de Peter Suber, disponível em http://www.earlham.edu/~peters/fos/fosblog.html. O blog foi descontinuado em abril de 2010, mas vale a pena navegar pelos arquivos. Suber preparou uma visão geral do Acesso Aberto [207] e uma linha do tempo [208], ambos muito úteis para obter uma visão geral do acesso aberto. Suber e outros continuam com o Projeto de Rastreamento do Acesso Aberto, cujos arquivos podem ser encontrados em http://oatp.tumblr.com/. Para uma visão geral do acesso aberto em formato de livro, consulte [241]. ***p 161:*** O arXiv está online em http://www.arxiv.org. Observe que o arXiv começou no campo da física, mas desde então se espalhou para outras disciplinas, como matemática e ciência da computação. Neste livro, concentrei-me nos aspectos da física e, às vezes, me refiro a ele como arXiv da física, visto que a física é o campo em que o arXiv é mais dominante.

p 162: O site da Biblioteca Pública de Ciências (PLoS) está em http://plos.org. A PLoS não foi a primeira revista de acesso aberto, mas foi uma das primeiras, e me concentrei nela porque abriu caminhos em muitos aspectos. ***p 162:*** Para uma visão geral da Política de Acesso Público do NIH, consulte [206]. É curta, mas contém muitos links informativos. ***p 162:*** O orçamento do NIH http://www.nih.gov/about/budget.htm.

p 164: Os números de receita e lucro da Elsevier são baseados no Relatório Anual da Reed Elsevier de 2009 [181]. ***p 164:*** Os números de receita e lucro da American Chemical Society são de [131]. ***p 164:*** Meu relato sobre Eric Dezenhall e a associação comercial dos editores (a Association of American Publishers) é baseado em [70], com informações adicionais de [100]. As citações do PRISM são de [176].

p 165: O artigo original de Simon Singh, no qual ele critica a Associação Britânica de Quiropraxia (BCA), está em [199]. O artigo de Dougans e Green sobre o caso Singh está em [56]. Minha discussão também se beneficiou dos artigos de Ben Goldacre [74] e Martin Robbins [185]. A descrição da BCA sobre as evidências da eficácia dos tratamentos quiropráticos está em [221]. Um caso semelhante de litígio wiki no mundo do software de código aberto envolveu afirmações de uma empresa chamada SCO de que um código de sua propriedade havia sido incorporado ao Linux, resultando na ação da SCO contra empresas como a Novell e a IBM. Os casos foram abordados com detalhes notáveis em um site comunitário chamado Groklaw (http://groklaw.net). iniciado por uma paralegal chamada Pamela Jones.

p 167: Pharyngula está em http://scienceblogs.com/pharyngula/. Os números da circulação do ***Des Moines Register*** e do ***Salt Lake Tribune*** são do Audit Bureau of Circulations [8].

p 170: Meu relato sobre a Ilha de Páscoa é baseado no livro ***Collapse*** [53], de Jared Diamond. A reconstrução da história da Ilha de Páscoa é difícil e complexa, e objeto de muita controvérsia entre os estudiosos; sem surpresa, alguns discordam do relato de Diamond. ***p 171:*** Sobre a redução da expectativa de vida devido ao HIV/AIDS nos países africanos mais afetados, ver [103].

p 171: Sobre como colmatar a lacuna da engenhosidade, ver [133].

Capítulo 8. O Desafio de Fazer Ciência Aberta

p 173: Meu relato do trabalho de Galileu é baseado em [238]. ***p 174:*** Para mais informações sobre o caso de Galileu e Baldassare Capra, veja [17]. ***p 175:*** Meu relato das origens do compartilhamento aberto de descobertas na ciência é baseado em parte no artigo de Paul David [49]. David aponta que não há nada logicamente inevitável sobre o surgimento da abertura na ciência, e que ela foi em grande parte resultado de forças externas agindo sobre a comunidade científica, não meramente forças dentro da ciência. A análise de David foca nos primórdios da ciência moderna, e enfatiza como a busca de prestígio por monarcas e outros patronos foi uma motivação para a divulgação aberta de resultados. Em meu relato, também enfatizei a motivação vinda do benefício público derivado da ciência aberta. Essa motivação parece ter adquirido mais força em tempos posteriores, à medida que o poder dos monarcas diminuía.

p 176: O qwiki está online em http://qwiki.stanford.edu/wiki/Main_Page. Na minha descrição do qwiki, afirmo que apenas algumas páginas são atualizadas regularmente. Na verdade, há uma parte do site que recebe atenção bastante regular: o “Zoológico da Complexidade”, um recurso para cientistas da computação que descreve diferentes tipos de problemas computacionais. O Zoológico da Complexidade precisa de recursos separados.

consideração, no entanto, pois se baseia em um projeto que originalmente era totalmente desconectado do qwiki e que posteriormente se fundiu com o qwiki. Como resultado, para os propósitos desta discussão, estou tratando o "Zoológico da Complexidade" como uma entidade separada. É, claro, interessante perguntar por que o Zoológico da Complexidade teve sucesso enquanto o restante do qwiki fracassou. Uma resposta completa para essa pergunta é complexa, mas, em resumo, o Zoológico da Complexidade tem um escopo muito mais restrito do que o qwiki e, devido a esse escopo mais restrito, uma única pessoa dedicada (Scott Aaronson, agora no MIT) foi capaz de desenvolvê-lo a ponto de se tornar um recurso extremamente útil e conhecido na comunidade da ciência da computação.
A combinação de seu perfil já elevado e seu escopo limitado ajudou a atrair algumas pessoas para fazer contribuições ocasionais à sua manutenção.

p 176: O termo “wiki-ciência” parece ter sido introduzido em um ensaio de Kevin Kelly [108]. Ideias semelhantes foram propostas independentemente (e, em alguns casos, anteriormente) por muitas pessoas. Uma discussão mais aprofundada envolvendo alguns dos primeiros colaboradores de wikis pode ser encontrada no wiki Meatball: [137] e [138]. ***p 178:*** Os dados de empregos e formaturas em física são baseados nos “Dados Mais Recentes de Emprego para Físicos e Cientistas Relacionados” do Instituto Americano de Física, disponíveis em http://www.aip.org/statistics/. Escolhi física porque há dados confiáveis disponíveis. Impressões anedóticas de outras áreas confirmam que a situação é semelhante.

p 178 Os wikis científicos que têm sucesso geralmente desempenham um papel de apoio a algum projeto mais convencional: uma exceção notável a essa regra é o Gene Wiki, um projeto wiki bem-sucedido para anotar genes. Parte do que ajudou o Gene Wiki a ter sucesso é que ele não é um wiki independente, mas sim um subprojeto da Wikipédia: se você já pesquisou um gene na Wikipédia, é provável que tenha visto trabalhos feitos como parte do projeto Gene Wiki. O Gene Wiki se beneficia das muitas pessoas que já dedicam tempo à edição e ao aprimoramento da Wikipédia, e da alta visibilidade que as páginas da Wikipédia costumam ter nos mecanismos de busca. ***p 179:*** Para outra perspectiva sobre sites de comentários contribuídos por usuários para ciência, ver [148]. ***p 179:*** O relatório final sobre o teste ***da Nature*** de revisão aberta por pares: [167].

p 180: Embora os sites de comentários científicos com contribuições de usuários estejam falhando, os cientistas nem sempre se mostram relutantes em comentar online sobre o trabalho de outros cientistas. Vimos um exemplo nesse sentido a partir da página 259, com blogueiros científicos investigando as evidências em favor da quiropraxia oferecidas pela Associação Britânica de Quiropraxia em sua disputa com Simon Singh. Outros exemplos incluem (1) uma colaboração no estilo Polymath [173] em 2010, na qual um grupo de matemáticos, cientistas da computação e físicos trabalharam juntos online para analisar uma suposta solução para um dos maiores problemas em aberto na ciência da computação; (2) uma discussão online baseada em blog [180] analisando o anúncio da NASA em 2010 [242] de que eles haviam descoberto formas de vida que incorporam arsênio; (3) Faculty of 1000 (http://f1000.com/), um site que recruta ativamente um número limitado

número de pesquisadores de alto nível para escrever revisões de artigos biomédicos; e (4) MathSciNet (http://www.ams.org/mathscinet/), um site semelhante para matemática. Em cada caso, os incentivos para potenciais colaboradores são bastante diferentes daqueles para os sites de comentários de usuários que descrevi. Não analisarei os incentivos aqui — o objetivo desta seção não é analisar exaustivamente os hábitos de comentários online dos cientistas —, mas observe que, em cada caso, uma análise detalhada mostra que os incentivos para os cientistas comentarem são muito mais fortes do que para os sites de comentários de usuários.

p. 182 ***"publique [artigos] ou pereça", e não "publique [dados] ou pereça"*** é de [171].

Capítulo 9. O Imperativo da Ciência Aberta

="0em" width="1em" align="justify">p ***187:*** O blog de pesquisa de Tobias Osborne sobre computação quântica está em http://tjoresearchnotes.wordpress.com/. A ideia da ciência de cadernos abertos foi desenvolvida em detalhes por Jean-Claude Bradley [26] e Cameron Neylon [147]. Veja também o blog de Bradley (http://usefulchem.blogspot.com/) e Blog do Neylon (http://cameronneylon.net/). ***p 187 a ciência aberta "exigiria que a maioria dos cientistas mudasse simultânea e completamente seu comportamento:*** [164]. ***p 188:*** Detalhes sobre a mudança sueca de dirigir pela esquerda para dirigir pela direita podem ser encontrados em [217] e [97]. A linguagem em meu relato é inspirada por uma frase maravilhosa de Stephen Pinker [170], que escreveu: "Uma mudança de dirigir pela esquerda para dirigir pela direita não poderia começar com um movimento ousado não-conformista ou de base, mas teria que ser imposta de cima para baixo (que foi o que aconteceu na Suécia às 5 da manhã de domingo, 3 de setembro de 1967)." ***p 188:*** De fato, o ***Journal des Sçavans*** tem a pretensão de ser o primeiro periódico científico do mundo, tendo começado a ser publicado alguns meses antes das ***Philosophical Transactions of the Royal Society.*** No entanto, a questão é discutível, visto que o ***Journal des Sçavans*** misturava conteúdo científico e não científico. ***p 188:*** Comentários de Mary Boas Hall sobre Oldenburg implorando por informações dos cientistas da época são fornecidas em [89] (página 159).

p 191: A situação política para compartilhamento de dados genéticos está evoluindo rapidamente. Para uma ampla visão geral da política dos Institutos Nacionais de Saúde, incluindo a política sobre estudos de associação genômica ampla, consulte [143]. Para a política específica do Instituto Nacional de Pesquisa do Genoma Humano em apoio ao Acordo das Bermudas, consulte [96]. Para a política do Wellcome Trust, consulte [236].

p 191: A política do Conselho de Pesquisa Médica do Reino Unido sobre dados abertos está em [229].

p 191** um porta-voz disse que este anúncio era apenas a "primeira fase" de um esforço para garantir que todos os dados fossem abertamente acessíveis:* [140]. ***p 191: As recomendações da OCDE sobre o acesso aberto a dados financiados publicamente os dados da pesquisa estão em [160].

p. 191: A expressão ***"a república da ciência"*** é retirada do excelente ensaio de Michael Polanyi [172] com o mesmo título. Entre outras coisas, o ensaio descreve os perigos de um controle centralizado excessivo na ciência, exatamente o tipo de controle centralizado que as agências de fomento têm hoje. (Quando Polanyi escrevia, as agências de fomento tinham orçamentos muito menores e, consequentemente, muito menos poder.) Concordo com as preocupações de Polanyi — de fato, é tentador escrever um ensaio complementar sobre "A Oligarquia da Ciência" —, mas o objetivo da discussão atual é, obviamente, encontrar as melhores ações no mundo em que nos encontramos, não em algum mundo idealizado.

p 193: Sobre direitos de propriedade em ideias e a mão invisível na ciência, veja [172,48]; um artigo geral interessante sobre explicações da mão invisível é [230].
Não sei de onde vem o termo "economia da reputação"; ele é amplamente utilizado desde a década de 1990 (e talvez antes), mas a ideia é muito mais antiga.

p 194: SPIRES está em http://www.slac.stanford.edu/spires/. A pré-impressão de física arXiv está, como observado anteriormente, em http://arxiv.org. ***p 195:*** Sobre novas formas de medir a ciência, ver, por exemplo, [175] e referências nele contidas.

p 196: Em relação ao desenvolvimento de novas ferramentas para a construção do conhecimento, atribuí a maior parte do ônus à construção dessas ferramentas aos cientistas. Você pode objetar que o desenvolvimento dessas ferramentas é tarefa de bibliotecas acadêmicas e editoras científicas. No entanto, há muitas razões para pensar que o lugar certo para a origem dessas ferramentas é com os próprios cientistas. Considere, por exemplo, que quase todos os exemplos que descrevi neste livro — do Projeto Polymath ao GenBank e ao arXiv — foram criados por cientistas.

Bibliotecas e editoras científicas não estão, em sua maioria, preparadas para trabalhar em inovações tão arriscadas e radicais. Em vez disso, estão voltadas para melhorias constantes nas formas existentes de fazer as coisas. Embora as bibliotecas e editoras empreguem muitas pessoas talentosas, quando essas pessoas tentam desenvolver ferramentas radicalmente novas, muitas vezes se veem lutando contra uma tremenda inércia institucional. Como resultado, o melhor lugar para novas ferramentas surgirem é com os próprios cientistas. Acredito que o papel apropriado para bibliotecas e editoras seja posterior, como parceiras que podem ajudar a sustentar e desenvolver ainda mais as ferramentas mais bem-sucedidas.

Foi exatamente isso que aconteceu com projetos como o arXiv e o GenBank, iniciados por cientistas, mas cujo crescimento e desenvolvimento se deram por meio de parcerias com a Biblioteca da Universidade Cornell e a Biblioteca Nacional de Medicina dos EUA, respectivamente. ***p. 196:*** Continuando o tema da última nota, você também pode se perguntar se o desenvolvimento de novas ferramentas de software não seria uma tarefa para uma agência centralizada. Isso foi tentado na biologia, por exemplo, onde muitas ferramentas de software são

desenvolvido no Centro Nacional de Informações sobre Biotecnologia (NCBI) dos EUA, que por sua vez faz parte da Biblioteca Nacional de Medicina dos EUA. O NCBI é responsável pela administração do GenBank e também ajudou a criar ou dar suporte a muitos outros importantes bancos de dados biológicos online. Mas, embora o NCBI ofereça um serviço valioso, ele também centraliza a inovação e afasta potenciais concorrentes, que não podem esperar competir com os recursos do NCBI. A longo prazo, acredito que a ciência precisa de uma abordagem mais descentralizada para a inovação. ***p. 196:*** Sobre os limites da fusão científica, ver [154]. ***p. 198:*** Sobre as expectativas sobre privacidade, ética, segurança e legalidade, essas expectativas, é claro, evoluirão. Sites como Patients Like Me (http://patientslikeme.com) pedir aos pacientes médicos que compartilhem voluntariamente suas informações médicas, e muitos pacientes fizeram isso, em parte para que as informações pudessem ser usadas para fins de pesquisa.

p 198: A citação de Grothendieck é de [85]. Veja também a discussão em capítulo 18 de [200], onde aprendi sobre esta citação.

p 198: O problema de gerenciar a atenção em colaboração foi estudado experimentalmente em [76]. Seus resultados são consistentes com a análise aqui, e mostram que a resolução de problemas em grupo pode, na verdade, se tornar menos eficaz se todos se comunicarem com todos os outros. ***p 200:*** Um relato do e-mail de Trenberth, juntamente com um link para o e-mail original (aparentemente genuíno), pode ser encontrado em [44]. O artigo original de Trenberth [225] é bastante legível. ***p 201:*** Sobre o gerenciamento dos dados do Kepler, veja [91, 166]. Para o anúncio de fevereiro de 2011 de planetas do tamanho da Terra, veja [129]. Observe que em setembro de 2010 outra equipe anunciou independentemente [232] a descoberta de um planeta semelhante à Terra, ao redor da estrela Gliese 581. Essa descoberta foi contestada desde então [110].

p 201: O anúncio de Dorigo de que estava ouvindo rumores de que a partícula de Higgs havia sido descoberta está em [54], e sua retratação está em [55]. A cobertura na grande mídia inclui [47, 41]. ***p 202:*** Uma discussão sobre a história da classificação dos grupos simples finitos é dada em [201]. O status atual da classificação é discutido em [5]. ***p 203 Como outros cientistas podem verificar e reproduzir os resultados de tais experimentos?:*** Veja, por exemplo, [205] e as referências nele contidas.

p 203: Sobre “ciência além da compreensão individual”, ver [155]. ***p 203 Em todo o mundo, nossos governos gastam mais de 100 bilhões de dólares por ano em pesquisa básica:*** baseei esta afirmação no capítulo 4 de um relatório da Fundação Nacional de Ciências dos EUA [144]. De acordo com os números incluídos naquele relatório, o governo dos EUA gasta 39 bilhões de dólares por ano em pesquisa básica. O relatório não calcula diretamente o gasto governamental total mundial em pesquisa básica e, portanto, o valor de 100 bilhões de dólares é uma estimativa, baseada em vários outros números daquele relatório.

p 206: A citação de Daniel Hillis “há problemas que são impossíveis&0;. . .” é da página 157 do livro de Stewart Brand, ***The Clock of the Long Now*** [27].

Apêndice

p 211: Uma introdução suave ao teorema de densidade de Hales Jewett (DHJ), incluindo uma explicação do conceito de linhas combinatórias, pode ser encontrada em
[66]. ***p 212:*** Para o teorema de Szemerédi, veja [218]. O teorema de Green-Tao é provado em
[84]. ***p 212:*** A prova original de DHJ estava em [66].

Referências

[1] Aotearoa. Comentário sobre uma postagem do blog ***Boing Boing*** (blog), 3 de maio de 2009. http://www.boingboing.net/2009/05/03/wasting-time-for-ag.html#comment-481275.

[2] Aotearoa. Comentário em uma postagem do blog ***Nature newsblog,*** 23 de março de 2009. http://blogs.nature.com/news/2008/05/addictive_protein_folding_game.html#comment-97818 .

[3] Oliver Arafat e Dirk Riehle. A distribuição do tamanho de commits de software de código aberto. ***Anais da 42ª Conferência Internacional Havaiana sobre Ciências de Sistemas,*** 2008.

[4] G. Arnison et al. Observação experimental de pares de léptons de massa invariante no colisor SPS 95 GeV/c^2 do CERN. ***Physics Letters*** B, 126(5):398–410, 1983.

[5] Michael Aschbacher. O status da classificação dos grupos simples finitos. ***Avisos da Sociedade Matemática Americana,*** 51(7):736–740, agosto de 2004. http://www.ams.org/notices/200407/fea-aschbacher.pdf.

[6] Nikolaos Askitas e Klaus F. Zimmermann. Econometria do Google e previsão de desemprego. ***Applied Economics Quarterly,*** 55:107–120, 2009.

[7] Sitaram Asur e Bernardo A. Huberman. Prevendo o futuro com as mídias sociais. ***impressão eletrônica arXiv:1003.5699,*** 2010.

[8] Audit Bureau of Circulations. ACCESS ABC: eCirc para jornais dos EUA, 2010. http://abcas3.accessabc.com/ecirc/newstitlesearchus.asp.

[9] Robert M. Axelrod. ***A evolução da operação.*** Nova York: Basic Books, 1984.

[10] Manda Banerji, Ofer Lahav, Chris J. Lintott, Filipe B. Abdalla, Kevin Schawinski, Steven P. Bamford, Dan Andreescu, Phil Murray, M. Jordan Raddick, Anze Slosar, Alex Szalay, Daniel Thomas e Jan Vandenberg. Galaxy Zoo: Reproduzindo morfologias de galáxias via aprendizado de máquina. ***Monthly Notices of the Royal Astronomical Society,*** 406(1):342–353, julho de 2010. eprint arXiv:0908.2033.

[11] J. Battelle. ***A Busca: Como o Google e seus rivais reescreveram as regras dos negócios e transformaram nossa cultura.*** Boston: Nicholas Brealey, 2005.

[12] Yochai Benkler. O pinguim de Coase, ou Linux e ***a natureza da empresa. The Yale Law Journal,*** 112:369–446, 2002.

[13] Yochai Benkler. ***A riqueza das redes.*** New Haven: Yale University Press, 2006.

[14] Tim Berners-Lee. ***Tecendo a teia.*** Nova Iorque: Harper Business, 2000.

[15] Tim Berners-Lee e James Hendler. Publicação na web semântica. ***Nature,*** 410:1023–1024, 26 de abril de 2001.

[16] Tim Berners-Lee, James Hendler e Ora Lassila. A web semântica. ***Scientific American,*** 17 de maio de 2001.

[17] Mario Biagioli. ***Instrumentos de crédito de Galileu: telescópios, imagens, segredo.*** Chicago: Universidade de Chicago Press, 2006.

[18] Peter Block. ***Comunidade: A Estrutura do Pertencimento.*** São Francisco: Berrett Koehler, 2008.

[19] Barry Boehm, Bradford Clark, Ellis Horowitz, Ray Madachy, Richard Shelby e Chris Westland. Modelos de custo para futuros processos de ciclo de vida de software: COCOMO 2.0. ***Annals of Software Engineering,*** 1(1):57–94, 1995.

[20] Peter Bogner, Ilaria Capua, David J. Lipman e Nancy J. Cox et al. Uma iniciativa global para compartilhar dados sobre a gripe aviária. ***Nature,*** 442:981, 31 de agosto de 2006.

[21] John Bohannon. Jogadores desvendam a vida secreta das proteínas. ***Wired,*** 17(5), 20 de abril, http://www.wired.com/medtech/genetics/magazine/17-05/ff_protein? 2009. currentPage=todos.

[22] Parsa Bonderson, Sankar Das Sarma, Michael Freedman e Chetan Nayak. Um projeto para um computador quântico topologicamente tolerante a falhas. ***eprint arXiv:1003.2856,*** 2010.

[23] Christine L. Borgman. ***Bolsa de estudos na era digital.*** Cambridge, MA: MIT Press, 2007.

[24] Kirk D. Borne et al. Astroinformática: Uma abordagem do século XXI para a astronomia. ***eprint arXiv: 0909.3892,*** 2009. Documento de posicionamento para o Astro2010 Decadal Survey State, disponível em http://arxiv.org/abs/0909.3892.

[25] Todd A. Boroson e Tod R. Lauer. Um sistema candidato a buraco negro binário supermassivo subparsec. ***Nature,*** 458:53–55, 5 de março de 2009.

[26] Jean-Claude Bradley. Ciência do caderno aberto. ***Drexel CoAS E-Learning*** (blog), 26 de setembro de 2006. http://drexel-coas-elearning.blogspot.com/2006/09/open-notebook-science.html.

[27] Stewart Brand. ***O relógio do longo agora.*** Nova Iorque: Basic Books, 2000.

[28] John Seely Brown e Paul Duguid. ***A vida social da informação.*** Boston: Harvard Business School Press, 2000.

[29] Zacary Brown. Sou um solucionador. ***Perspectivas sobre Inovação*** (blog), 4 de fevereiro de 2009. http://blog.innocentive.com/2009/02/04/im-a-solver-zacary-brown/.

[30] Almirante Bumblebee. Comentário sobre a submissão “Kasparov versus o Mundo”, 2007. http://www.reddit.com/r/reddit.com/comments/2hvex/kasparov_versus_the_world/.

[31] Vannevar Bush. Como podemos pensar. ***Atlantic Monthly,*** julho de 1945.

[32] Declan Butler. A discussão sobre a gripe no banco de dados aumenta. ***The Great Beyond*** (blog), 14 de setembro de 2009. http://blogs.nature.com/news/thegreatbeyond/2009/09/flu_database_row_escalates.html.

[33] Robert H. Carlson. ***Biologia é tecnologia.*** Cambridge, MA: Universidade de Harvard Imprensa, 2010.

[34] Nicholas Carr. O Google está a tornar-nos estúpidos? ***Atlantic Monthly,*** Julho/Agosto de 2008.

[35] Nicholas Carr. ***The Shallows: O que a Internet está fazendo com nossos cérebros.*** Nova York: WW Norton & Company, 2010.

[36] Henry William Chesbrough. ***Opennovation: O novo imperativo para criar e lucrar com a tecnologia.*** Boston: Harvard Business Press, 2006.

[37] Base de Xadrez. O azarão ZackS vence o torneio de xadrez estilo livre, 19 de junho de 2005. http://www.chessbase.com/newsdetail.asp?newsid=2461.

[38] Base de Xadrez. Hydra perde as quartas de final do torneio Freestyle, 11 de junho de 2005. http://www.chessbase.com/newsdetail.asp?newsid=2446.

[39] Base de Xadrez. Relatório PAL/CSS da boca do azarão, 22 de junho de 2005. http://www.chessbase.com/newsdetail.asp?newsid=2467. [40] O xadrez de jogos de Hydra (Computador). http://www.chessgames.com/perl/chessplayer?pid=87303.

[41] Tom Chivers. O rival do Grande Colisor de Hádrons, Tevatron, “encontrou o bóson de Higgs”, dizem rumores. ***Daily Telegraph,*** 12 de julho de 2010.

[42] Hyunyoung Choi e Hal Varian. Prevendo o presente com as tendências do Google. ***Google*** 12, 2009. ***Pesquisa*** ***blog,*** abril http://googleresearch.blogspot.com/2009/04/predicting-present-with-google-trends.html.

[43] Andy Clark e David J. Chalmers. A mente estendida. ***Análise,*** 58:10–23, 1998.

[44] “Climategate” exposto: mídia conservadora distorce e-mails roubados no mais recente ataque ao consenso sobre o aquecimento global. ***Media Matters,*** 1º de dezembro de 2009. http://mediamatters.org/research/200912010002.

[45] Robert P. Colwell. ***As Crônicas do Pentium.*** Hoboker, NJ: Sociedade de Computação IEEE, 2006.

[46] Seth Cooper, Firas Khatib, Adrien Treuille, Janos Barbero, Jeehyung Lee, Michael Beenen, Andrew Leaver-Fay, David Baker, Zoran Popoviÿ e jogadores do Foldit. Previsão de estruturas de proteínas com um jogo online multijogador. ***Nature,*** 466:756–760, 5 de agosto de 2010.

[47] Rachel Courtland. Bóson de Higgs: um resultado é iminente? ***New Scientist,*** 9 de julho de 2010.

[48] Partha Dasgupta e Paul A. David. Rumo a uma nova economia da ciência. ***Pesquisa Política,*** 23:487–521, 1994.

[49] Paul A. David. As origens históricas da “ciência aberta”: Um ensaio sobre clientelismo, reputação e contratos de agência comum na revolução científica. ***Capitalismo e Sociedade,*** 3(2), 2008.

[50] John R. Delaney e Roger S. Barga. Uma visão para a ciência oceânica em 2020. Em Tony Hey, Stewart Tansley e Kristin Tolle, editores, ***O Quarto Paradigma:*** Ciência com Uso Intensivo de Dados. Seattle: Microsoft Research, ***Descoberta*** 2009. http://research.microsoft.com/en-us/collaboration/fourthparadigm/

[51] Amit Deshpande e Dierk Riehle. O crescimento total do código aberto. Em ***Anais da Quarta Conferência sobre Sistemas de Código Aberto,*** 2008.

[52] David Diamond. A forma como vivemos agora: perguntas para Linus Torvalds. ***Nova Iorque Times,*** 28 de setembro de 2003.

[53] Jared Diamond. ***Colapso.*** Nova Iorque: Penguin Books, 2005.

[54] Tommaso Dorigo. Rumores sobre um Higgs leve. ***Um sobrevivente de Diários Quânticos*** (blog), 8 de julho de 2010. http://www.science20.com/quantum_diaries_survivor/rumors_about_light_higgs .

[55] Tommaso Dorigo. Então, o boato era mais do que apenas um boato, ou era um boato honesto? ***Um sobrevivente de Diários Quânticos*** (blog), 17 de julho de 2010. http://www.science20.com/quantum_diaries_survivor/so_was_rumor_more_just_rumor_or_was_it_honest_rumor.

[56] Robert Dougans e David Allen Green. Veracidade virtual. ***The Lawyer,*** 5 de julho de 2010.

[57] K. Eric Drexler. Publicação de hipertexto e a evolução do conhecimento. ***Social Inteligência,*** 1:87–120, 1991.

[58] Jason Dyer. Uma introdução suave ao Projeto Polímata. ***The Number Warrior*** (blog), 25 de março de 2009. http://numberwarrior.wordpress.com/2009/03/25/a-gentle-introduction-to-the-polymath-project/.

[59] David Easley e Jon Kleinberg. ***Redes, multidões e mercados.*** Cambridge: Cambridge University Press, 2010.

[60] Editorial da Nature. Sonhos de dados sobre a gripe. ***Nature,*** 440:255–256, 16 de março de 2006.

[61] Elizabeth L. Eisenstein. ***A Revolução da Imprensa na Europa Moderna (2ª ed.).*** Cambridge: Cambridge University Press, 2005.

[62] T. S. Eliot. ***A Floresta Sagrada: Ensaios sobre Poesia e Crítica.*** Londres: Methune, 1920.

[63] Douglas C. Engelbart. Aumentando o intelecto humano: uma estrutura conceitual. ***Relatório do Instituto de Pesquisa de Stanford,*** outubro de 1962.

[64] Jon Fortt. Os 5 melhores momentos da palestra de Eric Schmidt em Abu Dhabi. ***Fortune Tech*** (blog), 2010. Marchar 11, http://brainstormtech.blogs.fortune.cnn.com/2010/03/11/top-five-moments-from-eric-schmidt%27s-talk-in-abu-dhabi/.

[65] Elenco completo e equipe de Avatar. ***Internet Movie Database (IMDb).*** http://www.imdb.com/title/tt0499549/fullcredits.

[66] Hillel Furstenberg e Yitzhak Katznelson. Uma versão de densidade do Hales-Jewett teorema. ***Journal d'Analyse Mathematique,*** 57:64–119, 1991.

[67] Galáxia Jardim zoológico Fórum. O Voorwerp de Hanny, http://www.galaxyzooforum.org/index.php?topic=3802.0, 2007–ÿ.

[68] Jeffrey S. Gerber e Paul A. Offit. Vacinas e autismo: uma história de mudança hipóteses. ***Doenças Infecciosas Clínicas,*** 48:456–461, 2009.

[69] Jim Giles. Google lidera ranking de tradução. ***Nature News,*** 7 de novembro de 2006. http://www.nature.com/news/2006/061106/full/news061106-6.html.

[70] Jim Giles. O “pit bull” das Relações Públicas enfrenta o acesso aberto. ***Nature,*** 445:347, 1 de fevereiro de 2007.

[71] Jeremy Ginsberg, Matthew H. Mohebbi, Rajan S. Patel, Lynnette Brammer, Mark S. Smolinski e Larry Brilliant. Detectando epidemias de gripe usando mecanismos de busca dados de consulta. ***Nature,*** 457:1012–1015, 19 de fevereiro de 2009.

[72] James Gleick. ***Gênio: A vida e a ciência de Richard Feynman.*** Toronto: Random Casa do Canadá, 1993.

[73] Sharad Goel, Jake M. Hofman, Sébastien Lahaie, David M. Pennock e Duncan J. Watts. O que pode pesquisa prever? http://www.cam.cornell.edu/~sharad/papers/searchpreds.pdf, 2009.

[74] Ben Goldacre. Um grupo intrépido e maltrapilho de blogueiros. ***Guardian,*** 29 de julho de 2009. http://www.badscience.net/2009/07/we-are-more-possible-than-you-can-powerfully-imagine/.

[75] Michael H. Goldhaber. A economia da atenção e a rede. ***Primeira segunda-feira,*** 2(4–7), Abril de 1997.

[76] Robert L. Goldstone, Michael E. Roberts e Todd M. Gureckis. Emergente processos no comportamento de grupo. ***Current Directions in Psychological Science,*** 17(1):10–15,2008.

[77] J. Richard Gott III, Mario Juriÿ, David Schlegel, Fiona Hoyle, Michael Vogeley, Max Tegmark, Neta Bahcall e Jon Brinkmann. Um mapa do universo. ***Revista Astrofísica,*** 624(2):463–484, 2005.

[78] W. Timothy Gowers. Comentário no ***weblog de Gowers,*** 2 de fevereiro de 2009. http://gowers.wordpress.com/2009/02/01/questions-of-procedure/#comment-1701.

[79] W. Timothy Gowers. É possível uma matemática massivamente colaborativa? ***A teoria de Gowers weblog,*** 27 de janeiro de 2009. http://gowers.wordpress.com/2009/01/27/is-massively-collaborative-mathematics-possible/.

[80] W. Timothy Gowers. Polymath1 e matemática colaborativa aberta. ***De Gowers weblog,*** 10 de março de 2009. http://gowers.wordpress.com/2009/03/10/polymath1-and-open-collaborative-mathematics/.

[81] W. Timothy Gowers. Problema resolvido (provavelmente). ***Blog de Gowers,*** 10 de março de 2009. http://gowers.wordpress.com/2009/03/10/problem-solved-probably/.

[82] W. Timothy Gowers e Michael Nielsen, Matemática colaborativa massiva, ***Nature*** 461, 15 de outubro de 2009.

[83] Jim Gray. eScience: Um método científico transformado. Em Tony Hey, Stewart Tansley, e Kristin Tolle, editores, ***O Quarto Paradigma: Ciência Intensiva em Dados Descoberta.*** Seattle: Microsoft Research, 2009. http://research.microsoft.com/en-us/collaboration/fourthparadigm/.

[84] Ben Green e Terence Tao. Os primos contêm números aritméticos arbitrariamente longos progressões. ***Annals of Mathematics,*** 167:481–547, 2008.

[85] Alexander ***Recoltes*** http://brewfermentmagazine.org/ ***e*** ***Semailles.*** 1986. rands/recoltes1.html

[86] Ned Gulley. Em louvor ao ajuste: um concurso de programação semelhante a um wiki. ***Interações,*** 11(3):18–23, maio–junho de 2004.

[87] Ned Gulley e Karim R. Lakhani. Os determinantes do desempenho individual e valor coletivo na inovação de software privado-coletivo. Harvard Business School Documento de trabalho 10–65, 2010.

[88] Alon Halevy, Peter Norvig e Fernando Pereira. A eficácia irracional de dados. ***IEEE Sistemas Inteligentes,*** 24:8–12, 2009.

[89] Marie Boas Hall. ***Henry Oldenburg: Moldando a Royal Society.*** Oxford: Oxford Imprensa universitária, 2002.

[90] Michael J. Hammel. Indústria da mudança: Linux invade Hollywood. ***Linux Journal,*** Fevereiro de 2002. http://www.linuxjournal.com/article/5472

[91] Eric Hand. A equipe do telescópio pode ser autorizada a trabalhar com dados de exoplanetas. ***Nature News,*** 14 de abril de 2010. http://www.nature.com/news/2010/100414/full/news.2010.182.html

[92] Chuck Hansen. ***Armas nucleares dos EUA: a história secreta.*** Arlington, Aerofax, 1988.

[93] Friedrich von Hayek. O uso do conhecimento na sociedade. ***American Economic Review,*** 35(4):519–530, 1945.

[94] Tony Hey, Stewart Tansley e Kristin Tolle, editores. ***O Quarto Paradigma: Descoberta Científica Intensiva em Dados.*** Seattle: Microsoft Research, 2009. http://research.microsoft.com/en-us/collaboration/fourthparadigm/

[95] Edwin Hutchins. ***Cognição na Natureza.*** Cambridge, MA: MIT Press, 1995.

[96] Instituto Nacional de Pesquisa do Genoma Humano. Reafirmação e extensão de Políticas de divulgação rápida de dados do NHGRI: sequenciamento em larga escala e outras comunidades projetos de recursos, fevereiro de 2003. http://www.genome.gov/10506537

[97] Num instante, uma nação que seguia a faixa da esquerda desvia para a direita. ***Life,*** 15 de setembro de 1967.

[98] Jane Jacobs. ***A morte e a vida das grandes cidades americanas.*** Toronto: Random House do Canadá, 1961.

[99] Irving Lester Janis. ***Pensamento de grupo: estudos psicológicos de decisões políticas e Fiascos.*** Boston: Houghton Mifflin, 1983.

[100] Eamon Javers. O pitbull das relações públicas. ***Business Week,*** 17 de abril de 2006.

[101] Ayodele Samuel Jegede. O que levou ao boicote nigeriano à vacinação contra a poliomielite campanha? ***PLoS Medicina,*** 4(3):e73, 2007. http://www.plosmedicine.org/article/info:doi/10.1371/journal.pmed.0040073.

[102] Declaração conjunta do Presidente Clinton e do Primeiro-Ministro Blair. 14 de março de 2000. http://clinton4.nara.gov/WH/EOP/OSTP/html/00314.html

[103] Programa Conjunto das Nações Unidas sobre o VIH/SIDA. Impacto socioeconómico da epidemia e o fortalecimento das capacidades nacionais de combate ao VIH/SIDA, Junho 15, 2001. http://data.unaids.org/Publications/External-Documents/GAS26-rt3_en.pdf

[104] Jonathan Spencer Jones. JC Bennett (1914–1990). ***Jornal trimestral do Royal Sociedade Astronômica,*** 35:353, 1994.

[105] GIG Jozsa, MA Garrett, TA Oosterloo, H. Rampadarath, Z. Paragi, H. van Arkel, C. Lintott, WC Keel, K. Schawinski e E. Edmondson. Revelador Hanny's Voorwerp: Observações de rádio de IC 2497. ***Astronomia e Astrofísica,*** 500(2):L33–L36, 2009. eprint arXiv:0905.1851.

[106] Garry Kasparov. O mestre de xadrez e o computador. ***New York Review of Books,*** 57(2), 11 de fevereiro de 2010.

[107] Garry Kasparov com Daniel King. ***Kasparov contra o mundo.*** KasparovChess Online, 2000.

[108] Kevin Kelly. Especulações sobre o futuro da ciência. ***Edge: A Terceira Cultura,*** 2006. http://www.edge.org/3rd_culture/kelly06/kelly06_index.html.

[109] Kevin Kelly. ***O que a tecnologia quer.*** Nova Iorque: Viking, 2010.

[110] Richard A. Kerr. Mundo habitável descoberto recentemente pode não existir. ***Science Now,*** 12 de outubro de 2010. http://news.sciencemag.org/sciencenow/2010/10/recently-discovered-habitable-world.html.

[111] A. Yu Kitaev. Computação quântica tolerante a falhas por ânions. ***Annals of Physics,*** 303(1):2–30, 2003.

[112] Helge Kragh. Max Planck: O revolucionário relutante. ***Physics World,*** dezembro 2000. http://physicsworld.com/cws/article/print/373.

[113] Greg Kroah-Hartman. O kernel Linux. Vídeo online do Google Tech Talks. http://www.youtube.com/watch?v=L2SED6sewRw.

[114] Greg Kroah-Hartman, Jonathan Corbet e Amanda McPherson. Desenvolvimento do kernel Linux. ***The Linux Foundation,*** abril de 2008.

[115] Irina Krush com Kenneth W. Regan. A maior partida da história do xadrez, partes I, II e III. Disponível em http://www.cse.buffalo.edu/~regan/chess/KW/KHR99i.html, 1999.

[116] Karim R. Lakhani, Lars Bo Jeppesen, Peter A. Lohse e Jill A. Panetta. O valor da abertura na resolução de problemas científicos. Harvard Business School Working Paper 07–050, 2007.

[117] Jaron Lanier. ***Você não é um gadget: um manifesto.*** Toronto: Random House of Canada, 2010.

[118] Hadley Leggett. 18 de agosto de 1868: Hélio descoberto durante eclipse solar total. 2009. ***Wired,*** agosto 18, http://www.wired.com/thisdayintech/2009/08/dayintech_0818/.

[119] Jonah Lehrer. Estabelecendo conexões. ***Nature,*** 457:524–527, 28 de janeiro de 2009.

[120] Jonah Lehrer. Cientistas mapeiam o cérebro, gene por gene. ***Wired,*** 17, 28 de março de 2009. http://www.wired.com/medtech/health/magazine/17-04/ff_brainatlas.

[121] Ed S. Lein ***et al.*** Atlas de expressão gênica em todo o genoma no cérebro de camundongos adultos. ***Nature,*** 445:168–176, 11 de janeiro de 2007.

[122] Lawrence Lessig. ***Cultura livre: como a grande mídia usa a tecnologia e a lei para bloquear a cultura e controlar a criatividade.*** Nova York: Penguin, 2004. http://www.free-culture.cc/freecontent/.

[123] Pierre Levy. ***Coletivo de inteligência.*** Paris: La Découvte, 1994.

[124] Pierre Lévy. ***Inteligência Coletiva.*** Cambridge, MA: Perseus Books, 1997. Traduzido do original francês [123] por Robert Bononno.

[125] Jeff W. Lichtman, R. Clay Reid, Hanspeter Pfister e Michael F. Cohen. Descobrindo o diagrama de conexões do cérebro. Em Tony Hey, Stewart Tansley e Kristin Tolle, editores, ***O Quarto Paradigma: Descoberta Científica Intensiva em Dados.*** Seattle: Microsoft Research, 2009. http://research.microsoft.com/en-us/collaboration/fourthparadigm/.

[126] Chris Lintott. Ele disse que eles disseram que ele disse. . · . ***Galaxy Zoo*** (blog), 2009. http://blogs.zooniverse.org/galaxyzoo/2009/07/09/he-said-that-they-said-that-he-said/.

[127] Chris Lintott e Kate Land. Olhando o universo. ***Physics World,*** 21:27–30, 2008.

[128] Chris J. Lintott, Kevin Schawinski, William Keel, Hanny van Arkel, Nicola Bennett, Edward Edmondson, Daniel Thomas, Daniel JB Smith, Peter D. Herbert, Matt J. Jarvis, Shanil Virani, Dan Andreescu, Steven P. Bamford, Kate Land, Phil Murray, Robert C. Nichol, M. Jordan Raddick, Anže Slosar, Alex Szalay e Jan Vandenberg. Galaxy Zoo: “Hanny's Voorwerp”, uma luz quasar eco? ***Avisos mensais da Royal Astronomical Society,*** 399(1):129–140, Outubro de 2009. eprint arXiv:0906.5304.

[129] Jack J. Lissauer, Daniel C. Fabrycky, Eric B. Ford, William J. Borucki, François Fressin, Geoffrey W. Marcy, Jerome A. Orosz, Jason F. Rowe, Guillermo Torres, William F. Welsh, Natalie M. Batalha, Stephen T. Bryson, Lars A. Buchhave, Douglas A. Caldwell, Joshua A. Carter, David Charbonneau, Jessie L. Christiansen, William D. Cochran, Jean-Michel Desert, Edward W. Dunham, Michael N. Fanelli, Jonathan J. Fortney, Thomas N. Gautier III, John C. Geary, Ronald L. Gilliland, Michael R. Haas, Jennifer R. Hall, Matthew J. Holman, David G. Koch, David W. Latham, Eric Lopez, Sean McCauliff, Neil Miller, Robert C. Morehead, Elisa V. Quintana, Darin Ragozzine, Dimitar Sasselov, Donald R. Short e Jason H. Steffen. Um sistema compacto de baixa massa, planetas de baixa densidade transitando por Kepler-11. ***Nature,*** 470:53–58, 3 de fevereiro de 2011.

[130] Charles Mackay. ***Delírios populares extraordinários e a loucura das multidões*** (1841). Toronto: Random House do Canadá, 1995.

[131] Emma Marris. Sociedade Química Americana: Reação química. ***Natureza,*** 437:807–809, 6 de outubro de 2005.

[132] Karen Masters. Ela é uma astrônoma: Aida Berges. ***Galaxy Zoo*** (blog), 1º de outubro, 2009. http://blogs.zooniverse.org/galaxyzoo/2009/10/01/shes-an-astronomer-aida-berges/.

[133] Hassan Masum e Mark Tovey. Dadas mentes suficientes... lacuna. · . : Unindo a engenhosidade ***Primeira segunda-feira*** 11 (7), julho de 2006.

[134] Hassan Masum e Mark Tovey, editores. ***A Sociedade da Reputação.*** Cambridge, MA: MIT Press, a ser publicado.

[135] Peter Mclntyre e Julie Leask. Melhorar a aceitação da vacina MMR. ***Britânico Revista Médica,*** 336:729–730, 2008.

[136] Lucas Mearian. O CDC adota um novo sistema de rastreamento da gripe quase em tempo real. ***Computador Mundo,*** 5 de novembro de 2009.

[137] Wiki de almôndegas. WikiAsScience. http://meatballwiki.org/wiki/WikiAsScience.

[138] Wiki de Almôndegas. Publicação WikiScience. http://meatballwiki.org/wiki/WikiSciencePublication.

[139] David Mehegan. Autor(es)! autor(es)! ***Off the Shelf*** (blog), 10 de abril de 2007. http://www.boston.com/ae/books/blog/2007/04/authors_authors_2.html.

[140] Jeffrey Mervis. A NSF solicitará a cada candidato a subsídio um plano de gerenciamento de dados. ***ScienceInsider,*** 5 de maio de 2010. http://news.sciencemag.org/scienceinsider/2010/05/nsf-to-ask-every-grant-applicant.html.

[141] Relatório semanal de morbidade e mortalidade. Centro de Controle de Doenças dos Estados Unidos, 13 de outubro de 2006.

[142] Craig Mundie. O modelo de software comercial. Discurso no Congresso de Nova York Universidade Stern Business, 2001. Escola de Poderia http://www.microsoft.com/presspass/exec/craig/05-03sharedSource.mspx.

[143] Institutos Nacionais de Saúde. Política de compartilhamento de dados do NIH (em 17 de abril de 2007). http://grants.nih.gov/grants/policy/data_sharing/.

[144] Fundação Nacional de Ciências. Indicadores de ciência e engenharia, 2010. http://www.nsf.gov/statistics/seind10/pdfstart.htm.

[145] Theodor Holm Nelson. Máquinas ***LiteráriasSausalito ,*** CA: Mindful Press, 1987.

[146] Cameron Neylon. A troca científica. ***Ciência em Aberto*** (blog), 16 de abril, 2008. http://cameronneylon.net/blog/the-science-exchange/.

[147] Cameron Neylon. Cientistas lideram o movimento para o compartilhamento aberto de dados. ***Pesquisa Informações,*** Abril/Maio 2009. http://www.researchinformation.info/features/feature.php?feature_id=214.

[148] Cameron Neylon e Shirley Wu. Métricas de nível de artigo e a evolução de impacto científico. ***PLoS Biology*** 7(11): e1000242, 2009.

[149] Bob Nichol. Esta é a minha primeira vez. . . . ***Galaxy Zoo*** (blog), 19 de fevereiro de 2009. http://blogs.zooniverse.org/galaxyzoo/2009/02/19/this-is-my-first-time/.

[150] Michael Nielsen. Fazer ciência ao ar livre. ***Physics World,*** maio de 2009. http://physicsworld.com/cws/article/print/38904.

[151] Michael Nielsen. A economia da colaboração científica. ***Michael Nielsen blog,*** 29 de dezembro de 2008. http://michaelnielsen.org/blog/the-economics-of-scientific-collaboration/.

[152] Michael Nielsen. O futuro da ciência. ***Blog de Michael Nielsen,*** 17 de julho de 2008. http://michaelnielsen.org/blog/o-futuro-da-ciência-2/.

[153] Michael Nielsen, Despertar da informação, Nature Physics 5, abril de 2009.

[154] Michael Nielsen. A medição errada da ciência. ***Blog de Michael Nielsen,*** 29 de novembro de 2010. http://michaelnielsen.org/blog/the-mismeasurement-of-science/, e a ser publicado em [134].

[155] Michael Nielsen. Ciência além da compreensão individual. ***Blog de Michael Nielsen,*** 24 de setembro de 2008. http://michaelnielsen.org/blog/science-beyond-individual-understanding/.

[156] Nielsen Media Research. Relatório de três telas. ***Nielsen Wire*** (blog), 1º trimestre de 2009. http://blog.nielsen.com/nielsenwire/online_mobile/americans-watching-more-tv-than-ever/.

[157] Peter Norvig. Como escrever um corretor ortográfico. http://norvig.com/spell-correto.html.

[158] Barack Obama. Transparência e governo aberto. ***Federal Register,*** 74(15), Janeiro 26, 2009. http://www.whitehouse.gov/the_press_office/TransparencyandOpenGoverno/.

[159] Ryan O'Donnell. Comentário no ***weblog de Gowers,*** 6 de fevereiro de 2009. http://gowers.wordpress.com/2009/02/06/dhj-the-triangle-removal-approach/#comment-1913.

[160] OCDE. Princípios e diretrizes da OCDE para acesso a dados de pesquisa de fontes públicas financiamento. OCDE Relatório, Abril de 2007. http://www.oecd.org/document/55/0,3343,en_2649_201185_38500791_1_1_1_1,00.html.

[161] Mancur Olson. ***A lógica da ação coletiva: bens públicos e a teoria da Grupos.*** Cambridge, MA: Harvard University Press, 1965.

[162] Tim O'Reilly. A arquitetura da participação, junho de 2004. http://www.oreillynet.com/pub/a/oreilly/tim/articles/architecture_of_participação.html.

[163] Justin R. Ortiz, Hong Zhou, David K. Shay, Kathleen M. Neuzil e Christopher H. Goss. O rastreamento da gripe pelo Google está correlacionado com testes de laboratório positivos para gripe? Resumo da conferência. ***American Journal of Respiratory and Critical Medicina do Cuidado,*** 181:A2626, 2010.

[164] Tobias J. Osborne. Mais de 6 meses depois. ***Notas de pesquisa de Tobias J. Osborne*** (blog), 4 de outubro de 2010. http://tjoresearchnotes.wordpress.com/2009/10/04/over-6-

meses depois/.

[165] Elinor Ostrom. ***Governando os bens comuns: a evolução das instituições para a ação coletiva.*** Cambridge: Cambridge University Press, 1990.

[166] Dennis Overbye. Na busca por planetas, quem detém os dados? ***New York Times,*** 14 de junho de 2010. http://www.nytimes.com/2010/06/15/science/space/15kepler.html.

[167] Visão geral: ensaio de revisão por pares da Nature. ***Nature,*** dezembro de 2006. http://www.nature.com/nature/peerreview/debate/nature05535.html.

[168] Scott E. Page. ***A diferença: como o poder da diversidade cria grupos melhores.*** Princeton, NJ: Princeton University Press, 2008.

[169] A. Pais. ***Sutil é o Senhor: a ciência e a vida de Albert Einstein.*** Oxford: Oxford University Press, 1982.

[170] Stephen Pinker. ***A tábula rasa: a negação moderna da natureza humana.*** Nova Iorque: Penguin, 2003.

[171] Elizabeth Pisani e Carla AbouZahr. Compartilhamento de dados de saúde: boas intenções não são suficientes. ***Boletim da Organização Mundial da Saúde,*** 88(6), 2010.

[172] Michael Polanyi. A república da ciência: sua teoria política e econômica. ***Minerva,*** 1:54–74, 1962. http://www.missouriwestern.edu/orgs/polanyi/mp-repsc.htm.

[173] Participantes do Polymath. Artigo de Deolalikar P vs NP. ***Wiki do Polymath,*** 2010-. http://michaelnielsen.org/polymath1/index.php?title=Deolalikar's_P!%3DNP_paper.

[174] Zoran Popoviÿ. Resultados CASP8. ***Blog Foldit,*** 17 de dezembro de 2008. http://fold.it/portal/node/729520.

[175] Jason Priem, Dario Taraborelli, Paul Groth e Cameron Neylon. Altmetrics: Um manifesto. 26 de outubro de 2010. http://altmetrics.org/manifesto/.

[176] PRISM: Questões atuais. http://web.archive.org/web/20080330235026/http://www.prismcoalition.org/topics.htm.

[177] H. Rampadarath, MA Garrett, GIG Józsa, T. Muxlow, TA Oosterloo, Z. Paragi, R. Beswick, H. van Arkel, WC Keel e K. Schawinski. Hanny's Voorwerp: Evidências de atividade de AGN e uma explosão estelar nuclear nas regiões centrais de IC 2497. ***eprint arXiv: 1006.4096,*** 2010.

[178] Eric S. Raymond. A Catedral e o Bazar. Publicado online e reimpresso em [179]. http://www.catb.org/~esr/writings/cathedral-bazaar/cathedral-bazaar/.

[179] Eric S. Raymond. ***A Catedral e o Bazar: Reflexões sobre Linux e Código Aberto por um Revolucionário Acidental.*** Sebastopol, CA: O'Reilly Media, 2001.

[180] Rosie Redfield. Bactérias associadas ao arsênio (afirmações da NASA). ***Rrresearch*** (blog), 4 de dezembro de 2010. http://rrresearch.blogspot.com/2010/12/arsenic-associated-bacteria-nasas.html.

[181] Reed Elsevier. Relatórios anuais e demonstrações financeiras, 2009. http://reports.reedelsevier.com/ar09/ [182]

Richard Rhodes. ***Como escrever.*** Nova Iorque: Harper Collins, 1995.

[183] Richard Rhodes. ***A fabricação da bomba atômica.*** Nova York: Simon & Schuster, 1986.

[184] Ben Rich. ***Skunk Works: Uma memória pessoal dos meus anos na Lockheed.*** Boston: Little, Brown e Companhia, 1996.

[185] Martin Robbins. Uma revisão das evidências da BCA para a quiropraxia. ***O Lay Cientista: Blog de Martin,*** 2009. http://www.layscience.net/node/598.

[186] Bill Rosato. O campeão de xadrez Kasparov encontra-se com a partida na internet. ***Reuters (Londres),*** 3 de setembro de 1999.

[187]Robin Rowe. Linux #1 http://www.linuxmovies.org, 2008. sistema operacional em Hollywood.

[188] Donald Rumsfeld. Briefing de notícias do Departamento de Defesa dos Estados Unidos, 12 de fevereiro de 2002. http://www.defense.gov/transcripts/transcript.aspx?transcriptid=2636.

[189] Thomas C. Schelling. ***Micromotivos e Macrocomportamento.*** Nova Iorque: WW Norton & Company, 1978.

[190] Paul Seabright. ***A Companhia dos Estranhos: Uma História Natural da Vida Económica.*** Princeton, NJ: Princeton University Press, 2004.

[191] Toby Segaran. ***Programação de Inteligência Coletiva.*** Sebastopol, CA: O'Reilly Mídia, 2007.

[192] D. Shasha e C. Lazere. ***Fora de si: as vidas e descobertas de 15 grandes cientistas da computação.*** Nova Iorque: Springer-Verlag, 1998.

[193] Alice Sheppard. Ervilhas no universo, boa vontade e uma história de colaboração da Zooite no projeto das ervilhas. ***Galaxy Zoo*** (blog), 7 de julho de 2009. http://blogs.zooniverse.org/galaxyzoo/2009/07/07/peas-in-the-universe-goodwill-and-a-history-of-zooite-collaboration-on-the-peas-project/.

[194] Clay Shirky. ***Excedente cognitivo: criatividade e generosidade numa era conectada.*** Pinguim, 2010.

[195] Clay Shirky. Gim, televisão e excedente social. ***Here Comes Everybody*** (blog), 26 de abril de 2008. http://www.shirky.com/herecomeseverybody/2008/04/looking-for-the-mouse.html.

[196] Clay Shirky. ***Aí vem todo mundo: O poder de organizar sem organizações.*** Nova Iorque: Penguin, 2008.

[197] Herbert A. Simon. Projetando organizações para um mundo rico em informação. Em Martin Greenberger, editor, ***Computadores, Comunicação e o Interesse Público.*** Baltimore: Johns Hopkins Press, 1971.

[198] Cameron Sinclair. Cameron Sinclair sobre arquitetura de código aberto. ***TED: Ideias que valem a pena espalhar,*** 2006. http://www.ted.com/talks/cameron_sinclair_on_open_source_architecture. html.

[199] Simon Singh. Cuidado com a armadilha espinhal. ***Guardian,*** 19 de abril de 2008.

[200] Lee Smolin. ***O problema da física.*** Londres: Allen Lane, 2006.

[201] Ron Solomon. Sobre grupos simples finitos e sua classificação. ***Avisos da Sociedade Matemática Americana,*** 42(2):231–239, fevereiro de 1995. http://www.ams.org/notices/199502/solomon.pdf.

[202] Richard M. Stallman. ***Software Livre, Sociedade Livre: Ensaios Selecionados de Richard M. Stallman.*** Boston: Livre Software Fundação, 2002. http://www.gnu.org/philosophy/fsfs/rms-essays.pdf.

[203] Garol Stasser e William Titus. Perfis ocultos: uma breve história. ***Psychological Inquiry,*** 14(3&4):304–313, 2003.

[204] Garold Stasser e William Titus. Agrupamento de informações não compartilhadas na tomada de decisão em grupo: Amostragem tendenciosa de informações durante a discussão. ***Journal of Personality and Social Psychology,*** 48(6):1467–1478, 1985. [199] w Roman">[205]

Victoria Stodden, David Donoho, Sergey Fomel, Michael P. Friedlander, Mark Gerstein, Randy LeVeque, Ian Mitchell, Lisa Larrimore Ouellette, Chris Wiggins, Nicholas W. Bramble, Patrick O. Brown, Vincent J. Carey, Laura DeNardis, Robert Gentleman, J. Daniel Gezelter, Alyssa Goodman, Matthew G. Knepley, Joy E. Moore, Frank A. Pasquale, Joshua Rolnick, Michael Seringhaus e Ramesh Subramanian. Pesquisa reproduzível: abordando a necessidade de compartilhamento de dados e códigos na ciência computacional. ***Computação em Ciência e Engenharia,*** p. 12(5):8–12, set./out. 2010.

[206] Peter Suber. Um dia que vale a pena celebrar. ***Open Access News*** (blog), 17 de abril de 2008. http://www.earlham.edu/~peters/fos/2008/04/day-worth-celebrating.html.

[207] Peter Suber. http://www.earlham.edu/~peters/fos/overview.htm. Abrir acesso visão geral.

[208] Peter Suber. Linha do tempo do movimento de acesso aberto. http://www.earlham.edu/~peters/fos/timeline.htm.

[209] Brian Sullivan. Você usa eBird? — tópico aberto. ***Chip Notes: eBird Buzz*** (blog), maio 23, 2009. http://ebirdforum.blogspot.com/2009/05/do-you-ebird-tell-us-about-yourself.html.

[210] Brian L. Sullivan, Christopher L. Wood, Marshall J. Iliff, Rick E. Bonney, Daniel Finka e Steve Kellinga. eBird: Uma rede de observação de pássaros baseada em cidadãos no ciências biológicas. ***Conservação Biológica,*** 142(10):2282–2292, 2009.

[211] John Sulston. Patrimônio da humanidade. ***Le Monde Diplomatique (edição em inglês),*** Novembro de 2002.

[212] Cass R. Sunstein. ***Infotopia: Como Muitas Mentes Produzem Conhecimento.*** Nova Iorque: Imprensa da Universidade de Oxford, 2006.

[213] Cass R. Sunstein. ***Republic.com 2.0.*** Princeton University Press, 2007.

[214] James Surowiecki. ***A sabedoria das multidões.*** Nova Iorque: Doubleday, 2004.

[215] Don R. Swanson. Enxaqueca e magnésio: onze conexões negligenciadas. ***Perspectivas em Biologia e Medicina,*** 31(4):526–557, 1988.

[216] Don R. Swanson. Literatura médica como fonte potencial de novos conhecimentos. ***Boletim da Associação de Bibliotecas Médicas,*** 78(1):29–37, 1990.

[217] Suécia: Mude para o horário da direita Conse>Time, 15 de setembro de 1967.

[218] Endre Szemerédi. Sobre conjuntos de inteiros que não contêm ***k*** elementos em aritmética progressão. ***Acta Arithmetica,*** 27:199–245, 1975.

[219] Jeffery K. Taubenberger e David M. Morens. Gripe de 1918: a mãe de todas Pandemias. ***Doenças Infecciosas Emergentes,*** 12:15–22, 2006.

[220] Michael P. Taylor, Andrew A. Farke e Mathew J. Wede. O dinossauro aberto projeto. ***Boletim da Associação Paleontológica,*** (73), 2010. http://opendino.wordpress.com/2010/04/22/novo-artigo-do-odp-na-newsletter-da-associação-palaeontológica/.

[221] Terceira atualização sobre BCA v Simon Singh, junho de 2009. http://www.chiropractic-uk.co.uk/cfx/uploads/textbox/Singh/BCA%20Statement%20170609.pdf.

[222] Nativo do Texas de 31 anos desenvolve roteador sem fio movido a energia solar para a ASSET Índia, uma organização sem fins lucrativos focada na educação de crianças marginalizadas na Índia em tecnologia. ***Comunicado de Imprensa da InnoCentive,*** 25 de setembro de 2008. http://www.marketwire.com/press-release/Texas Native, de 31 anos, desenvolve roteador sem fio movido a energia solar-ASSET-India-Non-Profit-903974.htm.

[223] Linus Torvalds. A vantagem do Linux. Em ***fontes abertas: vozes do código aberto Revolução.*** editores, Chris DiBona, Sam Ockman e Mark Stone, Sebastopol, CA: O'Reilly Media, 1999.

[224] Mark Tovey, editor. ***Inteligência Coletiva: Criando um Mundo Próspero em Paz.*** Oakton, VA: Earth Intelligence Network, 2008.

[225] Kevin Trenberth. Um imperativo para o planeamento das alterações climáticas: monitorizar a evolução da Terra energia global. ***Opinião Atual em Sustentabilidade Ambiental,*** 1:19–27, 2009. http://www.cgd.ucar.edu/cas/Trenberth/trenberth.papers/EnergyDiagnostics09final2.pdf.

[226] Ilkka Tuomi. Evolução do arquivo de créditos do Linux: desafios metodológicos e dados de referência para pesquisa de código aberto. ***Primeira segunda-feira,*** 9(6), 2004.

[227] Jon Udell. Groupware da Internet para colaboração científica. 2000. http://jonudell.net/GroupwareReport.html.

[228] Jon Udell. O encontro de Sam com a serendipidade fabricada. ***Blog de rádio de Jon Udell,*** 4 de março de 2002. http://radio-weblogs.com/0100887/2002/03/04.html.

[229] Política do Conselho de Pesquisa Médica do Reino Unido sobre compartilhamento e preservação de dados. http://www.mrc.ac.uk/Ourresearch/Ethicsresearchguidance/ Iniciativa de compartilhamento de dados/Política/index.htm.

[230] Edna Ullmann-Margalit. Explicações da mão invisível. ***Síntese,*** 39(2): 263–291, 1978.

[231] Vernor Vinge. ***Rainbows End.*** Nova Iorque: Tor, 2007.

[232] Steven S. Vogt, R. Paul Butler, Eugenio J. Rivera, Nader Haghighipour, Gregory W. Henry e Michael H. Williamson. O Levantamento de Exoplanetas Lick-Carnegie: Uma 3.1 Planeta M_Terra na zona habitável da estrela M3V próxima, Gliese 581. ***impressão eletrônica arXiv:1009.5733,*** 2010.

[233] Eric von Hippel. ***Democratizar a inovação.*** Cambridge, MA: MIT Press, 2005.

[234] James D. Watson. ***A Dupla Hélice: Um Relato Pessoal da Descoberta da Estrutura do DNA.*** Nova Iorque: Simon e Schuster, 1980.

[235] Steven Weber. ***O sucesso do código aberto.*** Cambridge, MA: Universidade de Harvard Imprensa, 2004.

[236] Wellcome Trust. Política de gestão e partilha de dados. 2010. http://www.wellcome.ac.uk/About-us/Policy/Policy-and-position-statements/WTX035043.htm.

[237] Wellcome Trust. Partilha de dados de projectos de investigação biológica em larga escala: Uma sistema de responsabilidade tripartite. 2003. http://www.wellcome.ac.uk/stellent/groups/corporatesite/@policy_comunicacões/documentos/web_document/wtd003207.pdf.

[238] Richard S. Westfall. Ciência e mecenato: Galileu e o telescópio. ***Ísis,*** 76:11–30, 1985.

[239] O que é SourceForge.net? é http://sourceforge.net/apps/trac/sourceforge/wiki/O que é SourceForge.net?

[240] John A. Wheeler. ***Uma viagem pela gravidade e pelo espaço-tempo.*** Nova York: Scientific Biblioteca Americana, 1990.

[241] John Willinsky. ***O Princípio do Acesso.*** The MIT Press, Cambridge, Massachusetts, 2006.

[242] Felisa Wolfe-Simon, Jodi Switzer Blum, Thomas R. Kulp, Gwyneth W. Gordon, Shelley E. Hoeft, Jennifer Pett-Ridge, John F. Stolz, Samuel M. Webb, Peter K. Weber, Paul CW Davies, Ariel D. Anbar e Ronald S. Oremland. Uma bactéria que pode crescer usando arsênio em vez de fósforo. ***Science,*** 2 de dezembro de 2010.

[243] Anita Williams Woolley, Christopher F. Chabris, Alex Pentland, Nada Hashmi e Thomas W. Malone. Evidências de um fator de inteligência coletiva na desempenho de grupos humanos. ***Science,*** 330(6004):686–688, 29 de outubro de 2010.

[244] Organização Mundial da Saúde. Ficha informativa sobre a gripe n.º 211, março de 2003. http://www.who.int/mediacentre/factsheets/2003/fs211/en/.

[245] Lynn Wu e Erik Brynjolfsson. O futuro da previsão: como as pesquisas do Google prenunciam os preços e as vendas de imóveis. Apresentado no Workshop de 2009 sobre Economia de Sistemas de Informação (WISE 2009), 2009. http://pages.stern.nyu.edu/~bakos/wise/papers/wise2009-3b3_paper.pdf.

[246] Shirley Wu. Imaginando a comunidade científica como um grande laboratório. ***Um grande laboratório*** (blog), 14 de abril de 2008. http://onebiglab.blogspot.com/2008/04/envisioning-scientific-community-as-one.html.

[247] Donald G. York ***et al.*** O levantamento digital do céu de Sloan: Resumo técnico. ***Astronomical Journal,*** 120(3):1579–1587, setembro de 2000.

Índice