

EM BUSCA DE MÉTRICAS MAIS REFINADAS

Novos indicadores procuram desvendar tendências e estabelecer comparações mais precisas ao avaliar dados quantitativos de produção científica

Um artigo publicado em fevereiro na revista *Scientometrics* apresentou uma ferramenta computacional que pode ter utilidade na avaliação da produção científica de pesquisadores. Trata-se de um algoritmo capaz de coletar informações sobre o conjunto de *papers* de determinado autor e analisar, por exemplo, até que ponto essa produção está engajada nos temas quentes de sua disciplina ou como ela repercutiu, gerando citações em outros trabalhos, em relação a de colegas com interesses semelhantes. Os autores do manuscrito, o engenheiro de materiais Edgar Dutra Zanotto, da Universidade Federal de São Carlos (UFSCar), e o estudante de computação da universidade Vinícius Carvalho, desenvolveram o algoritmo para compilar e processar dados de dois indicadores fornecidos pela SciVal, plataforma analítica vinculada à base de dados Scopus, da editora Elsevier.

Um deles é o Field-Weighted Citation Index (FWCI), que avalia as citações de um artigo, comparando-as com as de outros *papers* com palavras-chave semelhantes e da mesma idade. O índice pondera o quanto esse trabalho foi mais ou menos citado em comparação com a média dos congêneres. Se o resultado for igual a 1, significa que está exatamente na média – se for superior, sua repercussão é maior. A vantagem da metodologia é que ela permite comparar estudos de qualquer área do conhecimento, ponderando sua posição em relação aos do mesmo assunto. Outros indicadores de impacto científico não permitem esse tipo de analogia, porque cada disciplina tem práticas de publicação peculiares e comunidades de tamanhos distintos, que influenciam a intensidade com que suas pesquisas são citadas sem que isso represente uma diferença de impacto e visibilidade.

O segundo indicador é o Topic Prominence Percentile (TPP), que mostra o quanto os temas de um artigo estão sintonizados com os assuntos mais discutidos no momento em seu campo do conhecimento. A taxa de proeminência é calculada a partir da ponderação do número de citações que o *paper* recebeu nos últimos dois anos, o fator de impacto da revista em que foi publicado e o número de visualizações que teve na internet. Seu conteúdo é cotejado com uma espécie de

ranking dos tópicos de pesquisa, feito também a partir de padrões de citação, sugerindo o quanto ele converge para os assuntos que mais vêm interessando aos editores de periódicos ou atraindo financiamento de agências de fomento. “O problema é que, para calcular o desempenho global de um pesquisador, é preciso pegar cada publicação individualmente, o que toma muito tempo. O software que desenvolvemos coleta, dentro da base Scopus, o FWCI e o TPP do conjunto de *papers* de um autor e busca fornecer um retrato abrangente de sua produção”, explica Zanotto.

Um teste preliminar do algoritmo feito com a produção de um pesquisador que publicou 226 artigos desde 2000 foi concluído em apenas 35 minutos. Alguns resultados foram curiosos. O FWCI do conjunto de trabalhos mais recentes era muito semelhante ao do conjunto dos mais antigos, um sinal de que a ponderação feita pelo índice corrigiu distorções e registrou desempenho estável daquele cientista. Também se observou que a produção feita em colaboração internacional está correlacionada com um FWCI mais elevado e que a sua taxa de proeminência era alta. Em seguida, o software avaliou a produção de 15 pesquisadores seniores, nível 1A no Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) – entre 30 e 50 anos de carreira –, e de 12 jovens pesquisadores prolíficos, selecionados de diferentes áreas, como química, física, astronomia, matemática, biologia, materiais. Constatou-se que praticamente todos da amostra tinham FWCI médio acima de 1. Foi realizada uma segunda análise, normalizando os artigos segundo a quantidade de autores de cada um deles. Dessa forma, dois pesquisadores seniores da área de matemática, que assinaram seus trabalhos com apenas dois ou três colegas, continuaram com o FWCI acima de 1 e ultrapassaram os demais, de outras áreas, que geralmente compartilham a autoria com mais coautores. “O algoritmo proposto e as métricas resultantes dele fornecem uma nova ferramenta em cientometria”, sustenta o pesquisador, referindo-se à disciplina que estuda aspectos quantitativos da ciência.

Líder de um dos principais grupos de pesquisa em nucleação e cristalização de vidros no mundo, Edgar Zanotto coordena desde 2013 o Centro de

Pesquisa, Educação e Inovação em Vidros (CeR-TEV), um dos Centros de Pesquisa, Educação e Difusão (Cepid) financiados pela FAPESP. Seu engajamento em estudos em cientometria é bissexto – entre os mais de 350 artigos que escreveu, apenas cinco são sobre o tema. O interesse pelo assunto remonta ao início dos anos 2000, quando ele foi membro da coordenação-adjunta de ciências exatas e engenharias da FAPESP. “Naquela época, eu analisava currículos de cientistas e bolsistas o tempo todo e começaram a aparecer uma série de métricas para analisar a produção de pesquisadores, como o índice-h”, diz, referindo-se ao indicador proposto em 2005 pelo físico argentino Jorge Hirsh, que combina o número de artigos com a consistência de suas citações e se tornou largamente usado pela facilidade com que é calculado. Em busca de parâmetros mais abrangentes, Zanotto publicou na mesma *Scientometrics*, em 2006, um estudo sugerindo uma forma de classificar pesquisadores baseada em 11 critérios, como prêmios internacionais conquistados, número de artigos publicados em revistas de alto impacto e volume de financiamento em pesquisa obtido de agências de fomento e empresas (ver Pesquisa FAPESP nº 124). A ideia serviu para fomentar o debate acadêmico e não chegou a ser adotada em larga escala. Mas Zanotto continua utilizando essa classificação até hoje quando avalia currículos de pesquisadores em projetos e prêmios.

De 2017 a 2019, Zanotto presidiu o Conselho Científico do Instituto Serrapilheira, uma instituição privada de fomento à pesquisa sediada no Rio de Janeiro, e renovou seu interesse por métricas de avaliação. Descobriu quase uma centena de indicadores disponíveis. Há, por exemplo, o índice-g, que aponta quando um pesquisador teve, entre seu conjunto de artigos, alguns altamente citados – particularidade que o índice-h não capta. A proliferação de trabalhos com centenas de autores também levou à criação de índices que ponderam o peso da contribuição de cada um, evitando distorções na comparação com *papers* com poucas assinaturas. No Serrapilheira, Zanotto estudou várias métricas para complementar a análise de

projetos de pesquisa do instituto. “A avaliação da produtividade, qualidade e visibilidade de cientistas é relevante, mas muito complexa, e ainda sem uma solução clara, adotada universalmente. Indicadores como o índice-h têm sido utilizados por muitos, mas não todos. Caso não sejam bem entendidos podem e têm causado inúmeras distorções. O FWCI corrige razoavelmente, mas não perfeitamente, tais distorções”, explica Zanotto.

A criação de novos indicadores bibliométricos foi impulsionada pelo desenvolvimento da ciência de dados, que permitiu extrair tendências de grandes volumes de informações. Empresas como a Elsevier e a Clarivate, que mantêm respectivamente as bases de dados Scopus e Web of Science, passaram a comercializar serviços de análise da produção científica para gestores, universidades e agências de fomento com métricas cada vez mais refinadas, ainda que a base da maioria delas continue a ser o número de artigos e citações. Sergio Salles-Filho, coordenador do Grupo de Estudos sobre Organização da Pesquisa e da Inovação da Universidade Estadual de Campinas (Geopi-Unicamp), que trabalha com a avaliação de produção acadêmica e tecnológica, conta que as análises puderam ser aprimoradas nos últimos anos graças a essas novas métricas. “Os resultados da avaliação ficaram mais calibrados, estabelecendo comparações apropriadas e evitando distorções”, afirma. Em um recente relatório de avaliação de três programas da FAPESP (de apoio a pequenas empresas, colaborações internacionais e formação de pesquisadores), Salles-Filho complementou os resultados com dados sobre a taxa de proeminência das pesquisas realizadas, fazendo a ressalva de que um resultado baixo não significa falta de qualidade, mas pode indicar apenas que um tema de pesquisa de grande interesse para o Brasil não está no topo da agenda da comunidade internacional (ver Pesquisa FAPESP nº 297).

Além de indicadores que normalizam a contribuição de pesquisadores de acordo com as características de sua disciplina, também ganharam espaço em processos de avaliação as chamadas métricas alternativas, ou altmetrias, que expõem o alcance de artigos científicos na imprensa e em mídias sociais (ver Pesquisa FAPESP nº 250). A

base de dados Dimensions, criada em 2018, fornece um índice sintético de altmetria composto por mais de uma dezena de parâmetros, entre citações em blogs, referências em tweets, likes no Facebook ou compartilhamentos na rede social acadêmica Mendeley. “A Dimensions só consegue detectar a repercussão de um trabalho na imprensa ou em redes sociais se ele contiver seu registro DOI [Identificador de Objeto Digital]. Por conta disso, os pesquisadores passaram a incluir o DOI quando divulgam seus estudos no Twitter ou no Facebook”, explica Salles-Filho.

A pesar da utilidade dos novos indicadores, persiste o debate sobre até que ponto essas métricas conseguem reconhecer a qualidade de um estudo científico e se têm potencial para substituir a revisão por pares, em que o conteúdo e o grau de inovação e de originalidade da produção de um autor são analisados por pesquisadores de sua área para mensurar o valor de sua contribuição. Para o bioquímico Jorge Guimarães, que presidiu a Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (Capes) entre 2004 e 2015, indicadores bibliométricos oferecem informações importantes e tornam possível a avaliação de um volume muito extenso de dados. “Mesmo criticadas, essas métricas são cada vez mais usadas pelo segmento acadêmico porque há poucas alternativas. Imagine um pesquisador que escreveu 200 *papers*. Um avaliador vai ler todos eles? Não é viável. Primeiro eu preciso saber quais artigos repercutiram na comunidade científica e receberam mais citações. E quantos não foram citados – se a proporção for alta, esse é um dado relevante”, diz Guimarães, que atualmente preside a Empresa Brasileira de Pesquisa e Inovação Industrial (Embrapii).

Nos anos 2000, a Capes criou o Qualis, sistema de classificação de revistas científicas usado para avaliar a produção dos programas de pós-graduação do Brasil. O sistema está sendo revisto e costuma ser criticado por medir o impacto de um artigo não pelo número de citações que ele de fato recebeu, mas por um parâmetro indireto: o índice de citação do periódico que o publicou. “É uma crítica injusta, porque a avaliação da Capes se debruça sobre o que

foi publicado nos quatro anos anteriores e, nesse breve espaço de tempo, o número de citações que cada artigo recebe é pequeno e não serviria como medida de avaliação”, afirma.

O matemático e engenheiro Renato Pedrosa, coordenador do programa especial de indicadores da FAPESP, reconhece a utilidade de índices bibliométricos em processos de avaliação, mas considera prudente utilizá-los com parcimônia. “No fundo, todos se baseiam nos mesmos parâmetros: artigos e suas citações. É razoável utilizá-los, por exemplo, quando se vai avaliar a produção de uma universidade ou de um departamento, mas a análise da produção individual de um pesquisador requer detalhamento e cuidado maiores”, diz Pedrosa, que é pesquisador do Departamento de Política Científica e Tecnológica da Unicamp. Ele menciona como exemplo o processo de avaliação das universidades e laboratórios do Reino Unido, realizado a cada cinco anos, que, em vez de se debruçar sobre toda a produção acadêmica do período, pede que os pesquisadores selecionem os dois ou três trabalhos mais significativos para que possam ser analisados em profundidade por revisores.

A preocupação está alinhada com o Manifesto de Leiden sobre métricas de pesquisa, concebido nos Países Baixos em 2015, que alerta para o uso indiscriminado dos indicadores na tomada de decisões de universidades e agências de fomento. “Corremos o risco de prejudicar o sistema da ciência com as próprias ferramentas projetadas para melhorá-lo, uma vez que a avaliação é cada vez mais realizada por instituições sem o devido conhecimento sobre as boas práticas e a interpretação adequada de indicadores”, afirma o manifesto. Pedrosa também chama a atenção para uma grande limitação dos indicadores bibliométricos, que é a sua incapacidade de avaliar pesquisadores em início de carreira. “Ainda que tenham talento e potencial, jovens pesquisadores não tiveram tempo de produzir artigos e receber citações”, afirma. Uma das vantagens do FWCI é que normaliza o número de citações recebidas segundo a idade do *paper*. ■

Artigo científico

ZANOTTO, E. D. *et al.* Article age- and field-normalized tools to evaluate scientific impact and momentum. *Scientometrics*. 25 fev. 2021.