



# OPERAÇÃO DE MULTIPLICAR

Estudo avalia produção de cientistas da computação que publicam muitos artigos e identifica práticas anômalas

**U**m grupo de cientistas da computação da Universidade Federal de Minas Gerais (UFMG) criou uma metodologia capaz de identificar comportamentos atípicos e questionáveis em pesquisadores prolíficos, aqueles que publicam uma quantidade de artigos científicos bem maior do que a média dos colegas. Em um estudo divulgado em fevereiro na revista *Scientometrics*, a equipe analisou artigos publicados entre 2010 e 2020 armazenados em um repositório do campo da computação, o DBLP (Digital Bibliography & Library Project), e encontrou algumas centenas de autores que escreveram mais de 19 artigos por ano. Esse desempenho é muito superior à média: 99% dos pesquisadores com trabalhos no repositório DBLP produzem menos de uma dezena de artigos por ano.

Foram mapeados, então, os padrões de publicação dos prolíficos. Viu-se que, em geral, a produção científica deles teve crescimento gradual, distribuiu-se harmonicamente por vários periódicos e envolvia um conjunto circunscrito de colaboradores. Mas havia alguns casos discrepantes. Eram autores tremendamente produtivos – um deles chegou a publicar 127 artigos em 2020 – e com

certas características: repentinamente, duplicaram ou triplicaram sua produção em um período curto, de dois a cinco anos, concentravam muitos trabalhos em poucas revistas e/ou montaram redes de colaboradores muito extensas, alguns deles com quase mil coautores.

Os responsáveis pelo estudo reconhecem que não é possível, com base no que observaram até agora, afirmar que os comportamentos anômalos configuram má conduta – eles pretendem avaliar, em trabalhos futuros, se a integridade científica foi comprometida por esses autores. Mas destacam que as discrepâncias registradas são consideráveis e querem saber se as anomalias detectadas poderiam delimitar a fronteira entre cientistas que publicam mais do que a média de forma autêntica e outros que amplificam de maneira artificial a sua produção. “Obtivemos um conjunto de métricas que aferem um fenômeno e acreditamos que elas podem vir a ter aplicação em várias disciplinas”, afirma Edré Quintão Moreira, aluno de doutorado do Programa de Pós-graduação em Ciência da Computação da UFMG e autor principal do estudo.

Uma das hipóteses levantadas pelo estudo é que alguns autores utilizem truques para inflar seu desempenho, como simular redes de colaboração que não existem de verdade. “Uma possibilidade é que possa haver algum nível de conluio entre os pesquisadores, com o acréscimo de colaboradores que não contribuíram de fato para seus manuscritos, a fim de ampliar a performance de todo o grupo”, diz Alberto Laender, docente aposentado do Departamento de Ciência da Computação (DCC) da UFMG e um dos autores do estudo. A concentração da publicação em determinados periódicos também acende um sinal amarelo. “Um dos altamente prolíficos publicou mais de 140 artigos em uma única revista. Uma pergunta que naturalmente surge é se esses autores estariam se beneficiando de políticas editoriais pouco rigorosas de alguns títulos e se contam com o encorajamento ou com a negligência dos editores para aumentar sua produtividade”, observa Laender. O interesse do grupo da UFMG pelo tema se explica. “Eu estava no comitê assessor de Ciência de Computação da Capes [Coordenação de Aperfeiçoamento de Pessoal de Nível Superior] e surgiram casos de autores

hiperprolíficos que pareciam suspeitos e nós não conseguíamos avaliar se os dados que apresentavam eram robustos”, diz Wagner Meira Júnior, também do grupo do DCC que assina o estudo.

O comportamento de autores hiperprolíficos há tempos desafia a compreensão de pesquisadores. Em um estudo publicado em 2018, John Ioannidis, epidemiologista da Universidade Stanford, levantou no repositório Scopus nomes de autores que tinham publicado ao menos 72 artigos em algum ano do período entre 2010 e 2016. Chegou a 9 mil indivíduos que haviam assinado ao menos um artigo a cada cinco dias. Ioannidis encontrou evidências apenas anedóticas de má conduta, como o caso do cientista de materiais japonês Akihisa Inoue, ex-reitor da Universidade de Tohoku, que publicou 2.566 artigos – sete deles foram retratados porque tinham conteúdo duplicado. Em 86% dos casos, os prolíficos eram físicos que participavam de consórcios internacionais, cujos artigos são assinados por centenas de colaboradores, às vezes, milhares. Os exemplos vinculados a essas grandes redes foram descartados por Ioannidis, que enviou questionários para 269 nomes remanescentes, recebendo 85 respostas. Uma de suas constatações é que existem, sim, pesquisadores capazes de escrever uma profusão de artigos, sem incorrer em desvios éticos, embora não dê para garantir que essa produção tenha relevância. Ele observou, contudo, que em certos casos a produtividade estava associada à falta de rigor no padrão de atribuição de autoria em algumas disciplinas.

Para Sigmar de Mello Rode, pesquisador da Universidade Estadual Paulista (Unesp) e presidente da Associação Brasileira de Editores Científicos (Abec-Brasil), ainda é comum em algumas áreas a pressão para incluir entre os autores de um artigo nomes que não se qualificam para assiná-lo, o que configura má conduta. “Acontece em todas as áreas do conhecimento de dois pesquisadores assinarem os artigos um do outro para ampliar a produção científica. O editor de um periódico não dispõe de muitas ferramentas para detectar esse tipo de conluio, menos ainda se ele ocorre em larga escala e envolve muitos pesquisadores”, afirma. No caso do estudo dos cientistas da computação, Rode suspeita também da chamada “produção salame”, em que

os resultados de um projeto de pesquisa são fatiados em diversos achados menores para dar origem a vários *papers*. “Esse tipo de truque é mais fácil de identificar, utilizando, por exemplo, softwares que buscam similaridades de textos”, afirma.

**N**a ciência da computação, a propagação dos autores altamente produtivos é fenômeno recente. De acordo com o estudo da *Scientometrics*, em 2010 somente 38 pesquisadores, o equivalente a 1% dos listados no DBLP, publicaram mais de 19 artigos. Já em 2020, esse desempenho foi observado em 540 autores, ou 6% dos cadastrados no repositório. Enquanto o cientista da computação mais prolífico de 2010 produziu 37 artigos no ano, o de 2020 conseguiu a proeza de publicar 127. “Foi por volta de 2016 e 2017 que houve um aumento repentino”, afirma Marcos André Gonçalves, docente do DCC, também responsável pelo estudo. “É muito estranho publicar dois artigos por semana. É um desempenho que em geral não condiz com uma produção de qualidade. Eu mal consigo ler dois artigos por semana.”

Embora tenha identificado parâmetros para definir comportamentos anômalos, o grupo da UFMG reconhece que outras dimensões do problema possam surgir. “Uma característica desse trabalho é que ele é vivo e dinâmico. Precisamos estar atentos a novos fatores que impulsionem esse fenômeno. Eu apostaria, por exemplo, que o advento do ChatGPT vai turbinar uma nova onda de autores hiperprolíficos”, diz Meira Junior, referindo-se ao famoso programa de inteligência artificial que também está sendo usado para auxiliar na escrita científica. Para o pesquisador, o risco é que essas práticas se tornem disseminadas. “Se isso acontecer, os jovens pesquisadores poderão entender que, em vez de seguir o caminho tradicional, baseado no trabalho e no esforço, é mais compensador adotar truques que turbinam o desempenho individual, mas corrompem o sistema.” ■ **Fabrcio Marques**

#### Artigo científico

MOREIRA, E. et al. The rise of hyperprolific authors in computer science: Characterization and implications. *Scientometrics*. v. 128(5), p. 2945-74. 15 mar. 2023.