



Pecha Kucha

DOI: [10.21680/2447-7842.2023v9n2ID33682](https://doi.org/10.21680/2447-7842.2023v9n2ID33682)

Preprints na Ciência da Computação: um estudo preliminar

Preprints in Computer Science: a preliminary study

Rosane Teles Lins Castilho ¹

Eloisa Príncipe ²

Submetido em: 17/04/2023	Aprovado na ConfOA: 14/06/2023	Publicado em: 04/12/2023
--------------------------	--------------------------------	--------------------------

Resumo: O movimento do acesso aberto, na última década, tem assistido à intensificação do processo de mudanças no cenário da comunicação científica. A esse movimento vêm se somando as iniciativas dos cientistas da computação, inclusive no Brasil, de depositarem seus manuscritos em servidores de *preprints*, como o *arXiv*. O objetivo desta pesquisa é apresentar questões que têm sido temas de discussão recorrentes na Ciência da Computação e refletidas na literatura internacional sobre o tema *preprints*. A metodologia adotada foi uma pesquisa bibliográfica realizada no Google Acadêmico e nas bases de dados *Scopus* e *Dimensions* sobre o tema *preprints* na Ciência da Computação, onde foi encontrada expressiva coleção de referências, algumas delas selecionadas e analisadas no decorrer deste texto. Como conclusão, aponta para a necessidade de aprofundamento dos estudos da comunicação científica da Ciência da Computação brasileira para entender melhor o seu comportamento e o perfil da sua produção em *preprints*.

¹ Doutoranda e Mestre em Ciência da Informação, Programa de Pós-Graduação em Ciência da Informação (convênio UFRJ-IBICT).

² Doutora e Mestre em Ciência da Informação (UFRJ), Docente do Programa de Pós-Graduação em Ciência da Informação (convênio UFRJ-IBICT). Tecnologista Sênior do Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT).



Palavras-chave: *preprints*; comunicação científica na Ciência da Computação; repositório de *preprints arXiv*.

Abstract: The open access movement, in the last decade, has witnessed the intensification of a process of changes in the scientific communication scenario. To this movement have been added the initiatives of computer scientists communities, including in Brazil, to deposit their manuscripts in preprint servers, such as arXiv. This research aims at presenting issues that have been recurrent topics of discussion on the theme preprints in Computer Science as reflected in the international literature. The methodology adopted is a bibliographic search carried out in Google Scholar and in the Scopus and Dimensions databases on the topic of preprints in Computer Science, where an expressive collection of references was found, some of them selected and analyzed throughout this text. As a conclusion, it points to the need to deepen the studies of scientific communication in Brazilian Computer Science to better understand its behavior and the profile of its production in preprints.

Keywords: *preprints*; Computer Science scientific communication; *arXiv preprints* repository.

1 INTRODUÇÃO

Resultados de estudos sobre a produção e a comunicação científica da Ciência da Computação, nacional e internacional, revelam uma rápida e intensa evolução da área e do crescimento de sua produção, bem como as características da comunicação entre pares que, apesar de não retirarem a importância do periódico científico avaliado por pares como veículo chave da disseminação da produção científica dos cientistas da computação, seu ritmo de publicação é lento, o que vai contra a resposta rápida necessária. Diferentemente de outras áreas, é nos anais de eventos científicos que a comunidade tem reunido maior interesse em publicar trabalhos de pesquisa por serem mais ágeis em termos do



acompanhamento e divulgação dos avanços científicos que, historicamente, têm se dado em ritmo acelerado (Sutton & Gong, 2017).

Na Ciência da Computação, no entanto, os trabalhos submetidos a eventos científicos também passam por avaliação rigorosa do comitê científico, um processo similar ao dos artigos científicos, resultando normalmente em taxas de aceitação entre 10% e 30% (Sutton & Gong, 2017). Mais recentemente, os cientistas têm expressado frustração com o atraso na disseminação de sua produção quando os artigos são rejeitados e têm que esperar vários meses até a data de submissão da próxima edição do evento, o que pode afetar até mesmo pesquisas inovadoras, pois podem envolver artigos mais difíceis para os revisores entenderem e apreciarem. Entre algumas das alternativas mais ágeis que os cientistas têm encontrado para divulgarem seus trabalhos encontram-se os *preprints* (LeCun, 2012).

A metodologia adotada para este estudo foi a pesquisa bibliográfica realizada no Google Acadêmico e nas bases de dados *Scopus* e *Dimensions* com as palavras-chave *preprints*, *Ciência da Computação/Computer Science*, *comunicação científica/scientific communication*, *arXiv*. Como resultado, foi encontrado um expressivo número de referências a questões sobre os *preprints* que têm sido temas de discussão recorrentes e refletidas na literatura da computação, algumas delas analisadas ao longo do presente texto.

2 DESENVOLVIMENTO

Preprints, pré-publicações, em português, são manuscritos que descrevem pesquisas que ainda não foram submetidas à revisão por pares, foram aceitos para publicação pelo editor de conformidade em periódicos ou em anais de eventos científicos, mas aguardam o resultado final da avaliação pelo comitê de revisores para serem publicados (Sutton & Gong, 2017).

É consenso entre as comunidades que a avaliação por pares é pedra angular do processo de publicação acadêmica, porém, os créditos que conferem garantia de autoria imediata à pesquisa, a agilidade com que, em pouco tempo, o documento se torna disponível online à comunidade científica, e a visibilidade de divulgação da



pesquisa, que amplia a possibilidade de citação do manuscrito, são fatores que têm atraídos novos adeptos aos *preprints* (Príncipe, 2021). Krumholz, Bloom e Ross (2020) destacam, ainda, como vantagens dos *preprints*, que os cientistas podem publicá-los rapidamente e receber *feedback* de seus pares, às vezes, quase instantaneamente, mantêm os cientistas informados sobre o que seus pares estão estudando e os ajudam a desenvolver seu trabalho, são arquivados de forma que possam ser referenciados, estarão sempre disponíveis *online* e, à medida que a Ciência evolui, novas versões do artigo podem ser publicadas, mantendo as versões históricas mais antigas disponíveis, incluindo quaisquer comentários associados feitos sobre elas.

O primeiro servidor de *preprints*, instalado no Laboratório Nacional de Los Alamos, nos Estados Unidos, em 1991, foi originalmente desenvolvido por Paul Ginsparg, para que pesquisadores das Ciências Físicas pudessem compartilhar seus trabalhos de modo mais ágil, antes da revisão por pares, foi lançado posteriormente como *arXiv* e, desde então, é hospedado na *Cornell University* (Krumholz, Bloom & Ross, 2020). Em 1998, mediante uma parceria entre a *Cornell University* e a *Association for Computing Machinery (ACM)*, fundada em 1947, a primeira sociedade científica e educacional, em âmbito internacional, dedicada à Ciência da Computação, foi criado o *Computing Research Repository (CoRR)*, seção de Ciência da Computação no *arXiv*, o *arXiv/CoRR* (ArXiv, [202?]; Halpern, 2000), tornando-se o primeiro repositório de pesquisa em computação de acesso aberto, permanente e *online* e tem sido o servidor de *preprints* amplamente utilizado pelas comunidades, nacional e internacional, da Ciência da Computação para postar manuscritos (Krumholz, Bloom & Ross, 2020), embora, no Brasil, ainda sejam escassas as informações encontradas na literatura quanto à extensão dessa tendência entre a comunidade da área.

Questões recorrentes em defesa dos *preprints* pela comunidade da área têm sido temas de discussão recorrentes e refletidas na literatura da computação, algumas delas selecionadas e apresentadas a seguir.

Segundo Lim (1996), por causa da demora entre a submissão de um artigo de pesquisa e sua publicação, muitos cientistas na vanguarda da pesquisa acham que



as informações publicadas servem mais para validar e certificar a posição de autor da pesquisa ou ter uma função de arquivo, do que ser um veículo para a comunicação de informações atuais e relevância direta para pesquisas em andamento. Le Goues *et al.* (2018) discutem a demora com que se processam as avaliações, bem como o problema das avaliações tendenciosas. Soergel *et al.* (2013) discutem as oportunidades e armadilhas das revisões por pares abertas, como uma solução para a demora e a imparcialidade nas avaliações, e a possibilidade de um sistema de avaliação mais flexível.

LeCun (2012) sugere como tentativa de solução de problemas com a avaliação pelos pares a adoção, pelo *arXiv/CoRR*, de um novo modelo de publicação, o *overlay journal*, Herman *et al.* (2020) e Rousi e Laakso (2022) acreditam que os *overlay journals* prometem mudanças importantes no cenário das publicações científicas, em áreas como a Matemática e a Computação. Sutton e Gong (2017) realizaram pesquisa sobre a popularidade do *arXiv*, medida pela porcentagem de artigos de anais de eventos da Ciência da Computação, anteriormente postados como *preprints* durante um período de 10 anos, e os resultados revelaram que *preprints* aumentaram, no período, cerca de 23%, entre as conferências mais seletivas da área. Polka *et al.* (2021) avaliaram *preprints* posteriormente publicados em periódicos para avaliar o quanto e como os artigos mudaram entre as versões e constataram que foram poucas as mudanças. Klein *et al.* (2019) investigaram os argumentos de editores de que adicionam valor às revisões de *preprint* em suas contrapartidas publicadas e os resultados revelam que o conteúdo do texto dos artigos científicos geralmente muda muito pouco, o que levanta questões sobre as vantagens do investimento nas revisões e as decisões financeiras das bibliotecas relativas à aquisição de publicações acadêmicas. Soderberg, Errington e Nosek (2020) investigaram as causas do ceticismo quanto à credibilidade de *preprints* e os motivos de não serem frequentemente citados. Descobriram que isto se deve à política dos periódicos que, até pouco tempo, não permitia citações de *preprints*, bem como à ausência de reputação do periódico ou do artigo, definida por métricas de citação e de revisão por pares, ambas ausentes em *preprints*.



Essas questões sugerem temas também a serem investigados entre as comunidades nacionais da Ciência da Computação, visando obter um quadro mais fiel do seu movimento em torno dos *preprints*. Distribuída entre vários Programas de Pós-graduação, a comunidade da área já faz uso dessa prática e conta com um número expressivo de manuscritos no *arXiv/CoRR*, necessitando de aprofundamento em estudos para compreender melhor o seu comportamento e o perfil da sua produção em *preprints*

3 CONSIDERAÇÕES FINAIS

A produção, disseminação e uso da informação são fundamentais para o desenvolvimento das comunidades científicas. Os avanços tecnológicos e o advento da Internet, junto com o Acesso Aberto aceleraram o processo de comunicação científica e tornaram as pesquisas mais visíveis, uma vez que permitem o acesso remoto a bancos de dados e repositórios, como o *arXiv*.

Os *preprints* têm sido vistos como uma tendência positiva para a comunicação científica, pois são acessíveis rápida e gratuitamente, permitem a interação dos cientistas não só com seus pares, mas com pesquisadores de outras áreas e mesmo com não acadêmicos e, sobretudo, democratizam o conhecimento. Deve-se lembrar, no entanto, que por serem postados em repositórios abertos e de rápida publicação, é necessária uma avaliação da qualidade, complexidade e seriedade dos trabalhos. O *preprint* não se propõe substituir os artigos científicos avaliados por pares, mas já demonstrou que é uma peça importante na forma como se faz e se comunica Ciência.

REFERÊNCIAS

ArXiv. ([202?]) *Computer Science Archive*. Recuperado de:

<https://arxiv.org/archive/cs>



Halpern, J. Y. (2000). CORR: a computing research repository. *ACM Journal of Computer Documentation*, 24(2), 41-48.

<https://dl.acm.org/doi/10.1145/337271.337274>

Herman, E., Akeroyd, J., Bequet, G., Nicholas, D., & Watkins, A. (2020). The changed – and changing – landscape of serials publishing: Review of the literature on emerging models. *Learned Publishing*, 33(3), 213-229.

Recuperado de: <https://onlinelibrary.wiley.com/doi/10.1002/leap.1288>

Klein, M., Broadwell, P., Farb, S. E., & Grappone, T. (2019). Comparing published scientific journal articles to their pre-print version. *International Journal of Digital Library*, 20, 335–350. Recuperado de:

<https://link.springer.com/article/10.1007/s00799-018-0234-1>

Krumholz, H., Bloom, T., & Ross, J. (2020). Preprints can fill a void in times of rapidly changing Science. *STAT Newsletter*. Recuperado de:

<https://www.statnews.com/2020/01/31/preprints-fill-void-rapidly-changing-science>

LeCun, Y. (2012). A new publishing model in Computer Science. *Pamphlets*.

Recuperado de: <http://yann.lecun.com/ex/pamphlets/publishing-models.html>



Le Goues, B. Y., Apel, S., Berger, E., Khurshid, S., & Smaragdakis, Y. (2018).

Effectiveness of anonymization in double-blind review. *Communications of the ACM*, 61 (6), 30-33. Recuperado de:

<https://cacm.acm.org/magazines/2018/6/228027-effectiveness-of-anonymization-in-double-blind-review/fulltext>

Lim, E. (1996). Preprint Servers: a new model for scholarly publishing? *Australian Academic Research Libraries*, Perth, 27 (1), 21-30. Recuperado de:

<https://www.tandfonline.com/doi/abs/10.1080/00048623.1996.10754952>

Polka, J. K., Dey, G., Palfy, M., Nanni, F., Brierley, L., Fraser, N., & Coates, J. A. (2021). Preprints in motion: tracking changes between posting and journal publication. *BioRxiv: the preprint server for biology*. Recuperado de:

<https://www.biorxiv.org/content/10.1101/2021.02.20.432090v1>

Príncipe, E. (2021). Prática da ciência aberta: os preprints em movimento. *Páginas a&b*, Coimbra, s. 3, nº especial ConfOA, 59-70. Recuperado de:

<https://ojs.letras.up.pt/index.php/paginasueb/article/view/10177/9677>

Rousi, A. M., & Laakso, M. (2022). Overlay journals: a study of the current landscape. *Journal of Librarianship and Information Science*. Recuperado de:

<https://arxiv.org/ftp/arxiv/papers/2204/2204.03383.pdf>



Soderberg, C. K., Errington, T. M., & Nosek, B. A. (2020). Credibility of preprints: an interdisciplinary survey of researchers. *Royal Society Open Science*, Charlottesville, 7(10). Recuperado de:
<https://royalsocietypublishing.org/doi/10.1098/rsos.201520>

Soergel, D., Saunders, A., & McCallum, A. (2013). Open scholarship and peer review: a time for experimentation. *In: International Conference On Machine Learning, 30th, Atlanta, 2013. Proceedings of the ICML Workshop on Peer Reviewing and Publishing Models (PEER)*. IMLR: W&CP v. 28. Recuperado de:
<https://openreview.net/forum?id=xf0zSBd2iufMg>

Sutton, C., & Gong, L. (2017). *Popularity of arXiv.org within Computer Science*. School of Informatics, University of Edinburgh, UK. Recuperado de:
<https://arxiv.org/pdf/1710.05225.pdf>