

Tipo do Artigo (Revisão)

Arquitetura Algorítmica, Polarização Afetiva e Erosão da Confiança Social: uma Revisão Crítica do Paradigma das Bolhas de Filtro

Rogério Ferreira Gonçalves^{1*}.

¹ Centro Universitário Evangélico de Goianésia (UNIEGO).

* Autor correspondente: rogerio.goncalves@docente.uniego.edu.br Telefone: 55 (62)984642280.

Resumo

Este artigo examina criticamente a hipótese de que a curadoria algorítmica de conteúdo em plataformas digitais produz isolamento informacional, amplificação de câmaras de eco e, consequentemente, polarização afetiva e erosão da confiança institucional. Por meio de revisão crítica de literatura em ciência da computação social, psicologia política e comunicação, contrapõe-se o modelo intuitivo, difundido em obras de divulgação científica, aos achados experimentais, que revelam efeitos mais modestos e heterogêneos do que a narrativa popular sugere. Foram analisados estudos experimentais, quase-experimentais e revisões sistemáticas publicados entre 2016 e 2025, priorizando desenhos com maior força inferencial sobre estudos meramente correlacionais. Os resultados indicam que a prevalência de bolhas de filtro estritas é fraca e dependente de operacionalização metodológica; que os efeitos polarizadores detectados decorrem primariamente da amplificação seletiva de conteúdo emocionalmente negativo e não da mera homogeneidade informacional; e que a exposição cruzada a conteúdo ideologicamente oposto pode paradoxalmente intensificar a polarização. Conclui-se que a compreensão psicossocial rigorosa do fenômeno exige um modelo de interação entre arquitetura algorítmica, heurísticas cognitivas e contexto social, em substituição ao modelo determinista de causalidade linear.

Palavras-chave: polarização afetiva; bolhas de filtro; câmaras de eco; capitalismo de vigilância; confiança social; psicologia política.

1. INTRODUÇÃO

A hipótese de que a personalização algorítmica de conteúdo fragmenta a esfera pública e intensifica a polarização social tornou-se um dos diagnósticos mais recorrentes sobre os efeitos psicossociais da comunicação digital contemporânea (Pariser, 2012; Sunstein, 2001; Zuboff, 2019). Essa hipótese tem sido tratada de forma predominantemente especulativa em produções de divulgação científica, que apresentam o fenômeno como fato empírico consolidado, quando a literatura acadêmica revela um quadro mais heterogêneo e contestado (Bruns, 2019; Ross Arguedas et al., 2022).

Diante da lacuna entre o discurso popular e a evidência científica revisada por pares, o presente estudo teve como objetivo sistematizar criticamente as evidências disponíveis sobre três proposições interligadas: a existência e prevalência de bolhas de filtro e câmaras de eco; a relação entre exposição algorítmica e polarização afetiva; e o papel dessas dinâmicas na erosão da confiança institucional.

2. MATERIAL E MÉTODOS

Trata-se de revisão crítica de literatura, seguindo os princípios de transparência metodológica descritos por Grant e Booth (2009) para revisões com finalidade de síntese teórica. A busca foi realizada nas bases Web of Science, Scopus e Google Scholar, restringindo-se a artigos publicados entre 2016 e 2025, em periódicos com revisão por pares. Foram priorizados estudos experimentais, quase-experimentais, revisões sistemáticas e meta-análises sobre estudos meramente correlacionais, em razão da fragilidade inferencial destes últimos para sustentar afirmações causais recorrentes na literatura de divulgação científica. Os termos de busca combinaram, em português e inglês, os descritores "filter bubble", "echo chamber", "affective polarization", "algorithmic curation" e "trust erosion".

Não foram utilizados dados envolvendo seres humanos ou animais, tampouco ferramentas de inteligência artificial generativa para análise, interpretação ou geração de dados, dispensando-se, portanto, aprovação de comitê de ética e a respectiva declaração..

3. RESULTADOS

3.1 Prevalência empírica da bolha de filtro

O conceito de bolha de filtro, cunhado por Pariser (2012), postula isolamento cognitivo progressivo decorrente da personalização algorítmica. Flaxman et al. (2016), analisando histórico de navegação de 50.000 usuários, encontraram evidência mista, com efeitos de magnitude modesta. Bakshy et al. (2015) constataram que a escolha individual do usuário exerce impacto maior sobre a exposição a conteúdo diverso do que a curadoria algorítmica. A revisão do Reuters Institute (Ross Arguedas et al., 2022) estimou entre 6% e 8% a proporção do público do Reino Unido em câmaras de eco estritamente delimitadas.

3.2 Exposição cruzada e conteúdo emocional

Bail et al. (2018), em experimento controlado, verificaram que republicanos expostos a conteúdo de bot liberal por um mês tornaram-se mais conservadores, resultado que desafia a premissa de que diversidade informacional funcionaria como antídoto automático à polarização. Milli et al. (2023) e Piccardi et al. (2024), em experimentos de campo, demonstraram que o mecanismo relevante é a amplificação algorítmica seletiva de conteúdo com raiva e animosidade intergrupal, e não a homogeneidade informacional em si.

3.3 Modelagem computacional dos mecanismos de retroalimentação

Chavalarias et al. (2023) demonstraram, por simulação calibrada com dados reais, que sistemas otimizados para engajamento produzem sobre-exposição a conteúdo negativo e concentração de poder social em usuários mais tóxicos. Germano et al. (2022) forneceram suporte empírico à hipótese de que o mesmo parâmetro algorítmico de personalização eleva simultaneamente engajamento, desinformação e polarização.

4. DISCUSSÃO

Os resultados revisados sustentam três correções ao paradigma popular sobre bolhas de filtro. Primeiro, a prevalência de bolhas de filtro estritas é fraca e metodologicamente contingente, contrariando a narrativa de isolamento informacional generalizado (Bruns, 2019; Ross Arguedas et al., 2022). Segundo, os efeitos polarizadores identificados decorrem da amplificação emocional seletiva promovida pelos algoritmos de recomendação, e não do isolamento informacional per se (Milli et al., 2023; Piccardi et al., 2024), o que desloca o foco explicativo da homogeneidade de conteúdo para a valência emocional do conteúdo amplificado.

Terceiro, e mais contraintuitivo, a exposição cruzada a conteúdo ideologicamente oposto pode intensificar, e não reduzir, a polarização afetiva (Bail et al., 2018), invalidando soluções simplistas baseadas exclusivamente em diversificação informacional como política corretiva. Chatterjee et al. (2023) reforçam que essa relação não é uniforme entre países, sendo mediada por fatores institucionais locais, o que contraindica generalizações universais frequentes em textos de divulgação científica sobre o tema.

Como limitação, a natureza não sistemática desta revisão crítica não permite estimativas quantitativas agregadas de tamanho de efeito, recomendando-se que pesquisas futuras conduzam meta-análises específicas sobre a interação entre conteúdo emocional, tempo de exposição e polarização afetiva.

3. CONCLUSÕES

A revisão crítica demonstra que a narrativa amplamente difundida sobre bolhas de filtro e erosão da confiança institucional como consequência direta da arquitetura algorítmica carece de sustentação empírica uniforme na literatura revisada por pares. Os mecanismos psicossociais envolvidos são mais complexos e contingentes do que sugere o modelo popular, sendo necessário adotar um modelo de interação multivariada entre arquitetura algorítmica, heurísticas cognitivas e contexto institucional, em substituição ao determinismo tecnológico predominante no discurso público sobre o tema.

Declaração de conflito de interesse: Os autores devem declarar a existência ou inexistência de conflitos de interesse de natureza financeira, institucional ou pessoal que possam influenciar a interpretação dos resultados apresentados. Na ausência de conflitos, deve ser incluída uma declaração explícita informando que não há conflito de interesse.

REFERÊNCIAS

- Bail, C. A., Argyle, L. P., Brown, T. W., Bumpus, J. P., Chen, H., Hunzaker, M. B. F., Lee, J., Mann, M., Merhout, F., & Volfovsky, A. (2018). Exposure to opposing views on social media can increase political polarization. *Proceedings of the National Academy of Sciences*, 115(37), 9216–9221. <https://doi.org/10.1073/pnas.1804840115>
- Bakshy, E., Messing, S., & Adamic, L. A. (2015). Exposure to ideologically diverse news and opinion on Facebook. *Science*, 348(6239), 1130–1132. <https://doi.org/10.1126/science.aaa1160>
- Bruns, A. (2019). *Are filter bubbles real?* Polity Press.
- Chatterjee, S., Kar, A. K., & Dwivedi, Y. K. (2023). Global heterogeneity in the relationship between misinformation, hate speech, and polarization: A configurational analysis. *Government Information Quarterly*, 40(2), 101799. <https://doi.org/10.1016/j.giq.2022.101799>
- Chavalarias, D., Bouchaud, P., & Panahi, M. (2023). Can engagement-optimized recommender systems increase enmity, radicalization and social distrust? *arXiv preprint*. <https://doi.org/10.48550/arXiv.2304.06616>
- Flaxman, S., Goel, S., & Rao, J. M. (2016). Filter bubbles, echo chambers, and online news consumption. *Public Opinion Quarterly*, 80(S1), 298–320. <https://doi.org/10.1093/poq/nfw006>
- Germano, F., Gómez, V., & Sobbrío, F. (2022). The economics of digital platforms and personalization: Engagement, misinformation, and polarization. *Information Economics and Policy*, 61, 100977. <https://doi.org/10.1016/j.infoecopol.2022.100977>
- Grant, M. J., & Booth, A. (2009). A typology of reviews: An analysis of 14 review types and associated methodologies. *Health Information & Libraries Journal*, 26(2), 91–108. <https://doi.org/10.1111/j.1471-1842.2009.00848.x>
- Milli, S., Carroll, M., Wang, Y., Pandey, S., Zhao, S., & Dragan, A. D. (2023). Engagement, user satisfaction, and the amplification of divisive content on social media. *arXiv preprint*. <https://doi.org/10.48550/arXiv.2305.16941>
- Pariser, E. (2012). *The filter bubble: How the new personalized web is changing what we read and how we think*. Penguin Books.
- Piccardi, T., Saveski, M., Jia, C., Hancock, J., Bernstein, M. S., & Jurafsky, D. (2024). Social media algorithms can shape affective polarization via emotional contagion. *Proceedings of the National Academy of Sciences*, 121(35), e2318934121. <https://doi.org/10.1073/pnas.2318934121>
- Ross Arguedas, A., Robertson, C. T., Fletcher, R., & Nielsen, R. K. (2022). Echo chambers, filter bubbles, and polarisation: A literature review. Reuters Institute for the Study of Journalism.
- Sunstein, C. R. (2001). *Republic.com*. Princeton University Press.
- Zuboff, S. (2019). *The age of surveillance capitalism: The fight for a human future at the frontier of power*. PublicAffairs.

NOTA DE INFORMAÇÃO: As declarações, opiniões e dados contidos em todas as publicações são de exclusiva responsabilidade dos autores e colaboradores individuais, e não da instituição mantenedora deste periódico e/ou dos editores. A EM SAÚDE e/ou os editores se eximem de qualquer responsabilidade por danos a pessoas ou bens resultantes de ideias, métodos, instruções ou produtos mencionados no conteúdo.