Nudges Digitais contra o Extremismo Online: Hiperpersonalização para a Desescalada da Polarização

Wladimir A. P. Neves[1], Angélica Dias[2], Daniel Schneider[2,3]

[1]Programa de Pós-Graduação em Informática (PPGI) - Universidade Federal do Rio de Janeiro (UFRJ) - Rio de Janeiro, RJ - Brasil

[2]Instituto Tércio Pacitti de Aplicações e Pesquisas Computacionais (NCE) - Universidade Federal do Rio de Janeiro (UFRJ) - Rio de Janeiro, RJ - Brasil

[3]Programa de Engenharia de Sistemas e Computação (PESC/COPPE) - Universidade Federal do Rio de Janeiro (UFRJ) - Rio de Janeiro, RJ - Brasil

wape.neves@gmail.com, angelica@nce.ufrj.br, schneider@nce.ufrj.br

Resumo. Este artigo visa planejar intervenções sociotécnicas para mitigar a polarização em plataformas digitais, promovendo a interação intergrupal construtiva em alinhamento ao tema "Colaborar para Transformar" do SBSC 2026. Fundamentado em uma RSL prévia, o estudo parte da premissa de que tratar os usuários de forma homogênea é uma limitação metodológica relevante. Sob o paradigma de *Design Science Research* (DSR), propõe-se o delineamento empírico de um artefato colaborativo que utiliza a hiperpersonalização psicométrica de *Nudges*. O objetivo é respeitar a "latitude de diversidade" de cada usuário e mitigar o Efeito Bumerangue. A estratégia de avaliação empírica, estruturada em um *survey* híbrido com *mockups*, buscará transpor a fragmentação metodológica interdisciplinar em IHC, unificando métricas topológicas de rede computacional com instrumentos psicológicos de polarização afetiva para validar a arquitetura proposta.

Abstract. This paper aims to plan sociotechnical interventions to mitigate polarization on digital platforms, promoting constructive intergroup interaction in alignment with the "Collaborate to Transform" theme of SBSC 2026. Based on a previous SLR, the study assumes that treating users homogeneously is a relevant methodological limitation. Under the Design Science Research (DSR) paradigm, it proposes the empirical design of a collaborative artifact using the psychometric hyper-personalization of Nudges. The goal is to respect each user's "latitude of diversity" and mitigate the Boomerang Effect. The empirical evaluation strategy, structured as a hybrid survey with mockups, will seek to bridge the interdisciplinary methodological fragmentation in HCI by unifying computational network topological metrics with psychological instruments of affective polarization to validate the proposed architecture.

1. Contexto: Colaboração e a Reconstrução do Debate Público

A ascensão das plataformas digitais como esferas primárias do debate público contemporâneo trouxe o desafio sociotécnico da radicalização e fragmentação sociopolítica [Kubin e Von Sikorski 2021], [Tucker et al. 2018]. O ecossistema informacional atual é estruturado por mecanismos latentes que priorizam a homogeneidade, criando Bolhas de Filtro (*Filter Bubbles*) que restringem invisivelmente o horizonte [Mattis et al. 2024).] Esses ambientes fomentam o *bonding social capital*, consolidando Câmaras de Eco (*Echo Chambers*) que fornecem a infraestrutura para a validação coletiva baseada na homofilia [Currin et al. 2022], [Pal et al. 2023]. Em resposta à chamada do SBSC 2026, o fenômeno colaborativo central desta pesquisa é a redução da hostilidade para viabilizar a interação intergrupal, promovendo o *bridging social capital*. O uso de *Digital Nudges* surge como uma ferramenta promissora [Thaler e Sunstein 2008], [Valta e Maier 2025]. Contudo, a premissa de que os usuários formam um bloco homogêneo compromete a eficácia das intervenções, configurando uma limitação metodológica relevante [Tucker et al. 2018], [Biselli et al. 2025].

2. Motivação e Revisão da Literatura: O Paradoxo e as Novas Fronteiras

A motivação advém de uma Revisão Sistemática da Literatura (RSL) prévia na interseção entre IHC, Computação e Ciências Sociais. A RSL revelou o "Paradoxo Interface vs. Estrutura": intervenções na interface preservam a agência, mas falham contra o viés de confirmação [Mattis et al. 2024], já intervenções de *backend* dissolvem bolhas, mas esbarram em manipulação e *Dark Patterns* [Özdemir 2020], [Meske e Amojo 2020]. Para superar esse paradoxo empiricamente, o planejamento foca em duas barreiras cruciais:

2.1. A "Latitude de Diversidade" e o Risco do Efeito Bumerangue

O problema primário estabelecido é que forçar usuários polarizados a consumirem mensagens da ideologia oposta frequentemente falha em promover a interação intergrupal construtiva. A literatura demonstra que a exposição frontal e abrupta a visões opostas muitas vezes engatilha o Efeito Bumerangue (*backfire effect*), onde o viés de confirmação atua e o indivíduo interpreta o *nudge* não como um convite ao diálogo, mas como uma ameaça à sua identidade social, radicalizando-o ainda mais [Bail et al. 2018]. Estudos apontam que a aceitação de conteúdos diversos possui uma estreita "latitude de diversidade" [Mattis et al. 2024]. Se a exposição algorítmica for muito intensa, gera-se uma sobrecarga cognitiva que provoca reatividade, prejudicando a desescalada da polarização. O mapeamento exato desta latitude suportada por cada indivíduo é essencial para evitar tal resistência.

2.2. A Hiperpersonalização pelo Perfil Psicométrico

A RSL demonstrou que diferentes subgrupos reagem de forma assimétrica às mesmas intervenções de diversidade [Tucker et al. 2018], [Kubin e Von Sikorski 2021]. O *digital nudge* contemporâneo possui natureza computacional imaterial, capaz de processar vastas quantidades de dados em tempo real para personalizar dinamicamente a arquitetura de escolha [Bartosiak 2022]. Trabalhos emergentes destacam a relevância de utilizar traços psicométricos — mensuráveis por meio de instrumentos validados, como a Escala de Necessidade de Cognição (*Need for Cognition Scale - NFC*) ou estilos de tomada de decisão — para criar intervenções de precisão [Biselli et al. 2025]. É nessa granularidade do comportamento humano que a tecnologia deve atuar: um *nudge* padronizado para um perfil moderado pode, inadvertidamente, radicalizar usuários com perfis mais extremos, reforçando a necessidade de calibração sistêmica.

3. Problema e Objetivos

Com base nas motivações expostas, e adotando o paradigma de *Design Science Research* (DSR), o problema que guiará esta pesquisa é subdividido em duas questões centrais: (QP1 - Desenho) Como projetar uma arquitetura sociotécnica que hiperpersonalize a injeção algorítmica de diversidade respeitando a latitude psicométrica do usuário? e (QP2 - Avaliação) Como correlacionar métricas topológicas de rede e escalas psicológicas para avaliar a mitigação da polarização afetiva sem desencadear o Efeito Bumerangue?

O objetivo central desta trilha de pesquisa é planejar e avaliar um protótipo de sistema colaborativo baseado na simbiose de *Nudges* e *Boosts* metacognitivos, desenhado para atenuar câmaras de eco. Os objetivos específicos alinham-se às três dimensões de contribuição da pesquisa: **(i) Contribuição Conceitual:** Estruturar o mapeamento da "latitude de diversidade" individual por meio de perfis psicométricos; **(ii) Contribuição Tecnológica:** Planejar a injeção calibrada de moderação algorítmica por meio de uma arquitetura sociotécnica; **(iii) Contribuição Metodológica:** Superar a fragmentação metodológica interdisciplinar, unificando indicadores computacionais e psicológicos na avaliação de resultados empíricos.

4. Solução Proposta: A Tríade Sociotécnica Hiperpersonalizada

A arquitetura de solução proposta neste plano empírico assenta-se em um modelo sistêmico dividido em três camadas estruturais de intervenção, projetadas para operar de forma orquestrada e hiperpersonalizada:

- No nível do *Backend* (Algoritmo): O sistema utilizará *Random Dynamical Nudges* (RDN) para quebrar a homofilia injetando opiniões moderadas [Currin et al. 2022], [Yu et al. 2024]. A inovação abandona a probabilidade genérica: o RDN será parametrizado pelo perfil psicométrico, inserindo diversidade estritamente dentro da "latitude" suportada pelo usuário [Pal et al. 2023], mitigando a sobrecarga cognitiva e prevenindo o Efeito Bumerangue.

- No nível do Processo (Interação e Regulação Emocional): O sistema introduzirá "fricções cognitivas" nas interações de compartilhamento para regulação emocional [Suzuki e Inaba 2025]. O mecanismo de espera (*Await*) substituirá a resposta imediata (Sistema 1) pela deliberação reflexiva (Sistema 2), reduzindo a agressividade que inviabiliza o diálogo [Konstantinou e Karapanos 2023].
- No nível da Interface (Transparência e *Boosting*): Para evitar *Dark Patterns* e preservar a agência do usuário, aplicar-se-ão ferramentas de *Boosting* [Grüne-Yanoff e Hertwig 2016]. A interface explicará ativamente a diversificação da dieta informacional promovendo o letramento [Pimentel et al. 2023], com personalização baseada na necessidade de cognição do indivíduo [Biselli et al. 2025].

Para fins de validação empírica e em alinhamento com a fase de construção do *Design Science Research* (DSR), a operacionalização deste sistema dar-se-á através do desenvolvimento de um protótipo funcional de rede social simulada (um *web app*). O fluxo operacional ocorrerá em três etapas integradas: (1) no acesso, o usuário responde a um instrumento psicométrico breve; (2) o sistema converte os traços mapeados em um limiar numérico de tolerância no *backend*; e (3) durante a navegação, o algoritmo consulta esse limiar em tempo real para calibrar a injeção de conteúdos divergentes e a ativação de fricções.

5. Métodos de Avaliação: Superando a Fragmentação Interdisciplinar

Considerando a etapa de avaliação do ciclo de *Design Science Research* (DSR), esta pesquisa adotará um delineamento quase-experimental acoplado a um levantamento crítico (*Survey* híbrido). O objetivo é avaliar o protótipo e calibrar a "latitude de diversidade" suportada por diferentes subgrupos ideológicos. O experimento dividirá os participantes em três braços: um Grupo Controle (navegação livre), um Grupo de Intervenção Genérica (exposto a *nudges* padronizados) e um Grupo de Intervenção Hiperpersonalizada (exposto à Tríade Sociotécnica). A simulação ocorrerá por meio de *mockups* interativos emulando uma rede social, onde os participantes serão expostos a manchetes polarizantes e terão métricas comportamentais (cliques, tempo de tela, compartilhamento) monitoradas.

O diferencial metodológico deste projeto será a transposição da fragmentação interdisciplinar entre Ciência da Computação e Psicologia Social, uma limitação histórica [Duane et al. 2025], [Valta e Maier 2025]. Para estruturar a hiperpersonalização, a pesquisa propõe a unificação de métricas quantitativas — como a variação matemática na distância de picos de opiniões em simulação de redes complexas (o $\Delta peak$, fundamental nas intervenções RDN [Pal et al. 2023] — com instrumentos validados da Psicologia. Especificamente, adotar-se-á a Escala de Necessidade de Cognição (*Need for Cognition Scale - NFC*) para mapear o perfil psicométrico, e a escala do Termômetro de Sentimentos (*Feeling Thermometer*) para mensurar a variação da polarização afetiva intergrupal.

Este cruzamento permitirá o uso de testes estatísticos de correlação para diferenciar claramente a "polarização ideológica" (quando há apenas uma discordância racional de propostas políticas) da "polarização afetiva" (o sentimento hostil e a animosidade direcionada), [Kubin e Von Sikorski 2021]. Reconhece-se, contudo, as ameaças à validade externa neste estágio, visto que o engajamento em um ambiente simulado de *mockup* pode diferir do comportamento orgânico em plataformas reais, exigindo cautela na generalização. Apenas mitigando a polarização afetiva é que as arquiteturas de escolha poderão voltar a inovar e conectar cidadãos de forma construtiva.

Conforme as diretrizes exigidas, declara-se que o modelo NotebookLM, de Inteligência Artificial Generativa, foi utilizado na pesquisa.

Referências

Alatawi, F., et al. (2021) "A Survey on Echo Chambers on Social Media: Description, Detection and Mitigation", Preprint / Artigo de Revisão.

Bail, C. A., Argyle, L. P., Brown, T. W., Bumpus, J. P., Chen, H., Fallin Hunzaker, M. B., Lee, J., Mann, M., Merhout, F. e Volfovsky, A. (2018) "Exposure to opposing views on social media can increase political polarization", Proceedings of the National Academy of Sciences.

Bartosiak, M. (2022) "Not So Digital After All? A Look at the Nature of Digital Nudging through the Prism of the Digital Object Concept", Proceedings of the 55th Hawaii International Conference on System Sciences.

Biselli, T., Hartwig, K. e Reuter, C. (2025) "Mitigating Misinformation Sharing on Social Media through Personalised Nudging", Proceedings of the ACM on Human-Computer Interaction.

Chaudhuri, S. e Bose, I. (2025) "Do nudges work? Examining the role of digital nudges to curb the spread of misinformation on social messaging platforms", Journal of Management Information Systems.

Currin, C. B., Vera, S. V. e Khaledi-Nasab, A. (2022) "Depolarization of echo chambers by random dynamical nudge", Scientific Reports.

Duane, J., Ericson, J. e McHugh, P. (2025) "Digital nudges: a systematic narrative review and taxonomy", Behaviour & Information Technology.

Grüne-Yanoff, T. e Hertwig, R. (2016) "Nudge Versus Boost: How Coherent are Policy and Theory?", Minds and Machines.

Konstantinou, L. e Karapanos, E. (2023) "Nudging for Online Misinformation: a Design Inquiry", CSCW '23 Companion.

Konstantinou, L. e Karapanos, E. (2025) "How Do Users Perceive Nudges Against Online Misinformation? A Repertory-Grid Study", International Journal of Human–Computer Interaction.

Kubin, E. e Von Sikorski, C. (2021) "The role of (social) media in political polarization: a systematic review", Annals of the International Communication Association.

Mattis, N., Groot Kormelink, T., Masur, P. K., Möller, J. e van Atteveldt, W. (2024) "Nudging News Readers: A Mixed-Methods Approach to Understanding When and How Interface Nudges Affect News Selection", Digital Journalism.

Mattis, N., Masur, P. K., Möller, J. e van Atteveldt, W. (2024) "Nudging towards news diversity: A theoretical framework for facilitating diverse news consumption through recommender design", New Media & Society.

Meiske, B., Álvarez-Benjumea, A., Andrighetto, G. e Polizzi, E. (2024) "Nudging punishment against sharing of fake news", European Economic Review.

Meske, C. e Amojo, I. (2020) "Ethical Guidelines for the Construction of Digital Nudges", Proceedings of the 53rd Hawaii International Conference on System Sciences.

Özdemir, Ş. (2020) "Digital nudges and dark patterns: The angels and the archfiends of digital communication", Digital Scholarship in the Humanities.

Pal, R., Kumar, A. e Santhanam, M. S. (2023) "Depolarization of opinions on social networks through random nudges", Physical Review E.

Pennycook, G. e Rand, D. G. (2022) "Nudging Social Media toward Accuracy", The Annals of the American Academy of Political and Social Science.

Pimentel, A. P., Motta, C., Correia, A. e Schneider, D. (2023) "Agenda of Solutions to Mitigate the Challenge of Polarization of Extreme Positions in Social Media Environments", Proceedings of the 26th IEEE International Conference on Computer Supported Cooperative Work in Design.

Suzuki, H. N. e Inaba, M. (2025) "Digital Nudges Using Emotion Regulation to Reduce Online Disinformation Sharing", Computers in Human Behavior.

Thaler, R. H. e Sunstein, C. R. (2008) "Nudge: Improving Decisions about Health, Wealth, and Happiness", Yale University Press.

Thornhill, C., Meeus, Q., Peperkamp, J. e Berendt, B. (2019) "A Digital Nudge to Counter Confirmation Bias", Frontiers in Big Data.

Tucker, J. A., Guess, A., Barberá, P., Vaccari, C., Siegel, A., Sanovich, S., Stukal, D. e Nyhan, B. (2018) "Social Media, Political Polarization, and Political Disinformation: A Review of the Scientific Literature", Hewlett Foundation.

Valta, M. e Maier, C. (2025) "Digital Nudging: A Systematic Literature Review, Taxonomy, and Future Research Directions", The DATA BASE for Advances in Information Systems.

Windasari, N. A., Kurnia, S. e Santoso, A. J. (2025) "The Use of Digital Nudges to Mitigate Online Incivility: A Systematic Literature Review", Information & Management.

Yu, X., Haroon, M., Menchen-Trevino, E. e Wojcieszak, M. (2024) "Nudging recommendation algorithms increases news consumption and diversity on YouTube", PNAS Nexus.