

Mapeamento Sistemático da Literatura sobre a Caracterização do Usuário do Twitter

João Marcelo Silva de Oliveira¹, Isadora Mendes dos Santos¹, Marcelle Pereira Mota¹

¹Instituto de Ciências Exatas e Naturais – Universidade Federal do Pará (UFPA)
Belém – PA – Brasil

jmarceu@ufpa.br, isadoramsantos@gmail.com, mpmota@ufpa.br

Abstract. *Social media have a vast set of data from their users. Collecting this data, transforming it into information and, later, into knowledge, has unique importance, not only for the companies that own these networks, but for the entire “ecosystem” in these networks. This article presents a systematic mapping of the literature, which aimed to find an answer to the following question: who is the twitter user? The articles were collected from the ACM Digital Library, Science Direct and IEEE Xplorer databases, according to the defined string, using the automatic search method. From the selected articles, 8 user identification categories were removed: individual or organization, multi-network, malicious, health, behavior, demographics, interests and identity. Also, the gray literature was consulted to integrate the result regarding the Twitter user and generated information such as the number of users by gender and the countries with the most Twitter users.*

Resumo. *As redes sociais possuem um vasto conjunto de dados dos seus usuários. Coletar estes dados, transformá-los em informação e, posteriormente, em conhecimento, tem importância ímpar, não apenas para as empresas proprietárias destas redes mas para todo o “ecossistema” nestas redes. Este artigo apresenta um mapeamento sistemático da literatura e teve por objetivo encontrar uma resposta para o seguinte questionamento: quem é o usuário do twitter? Os artigos foram coletados das bases ACM Digital Library, Science Direct e IEEE Xplorer, conforme string definida, utilizando o método de busca automática. Dos artigos selecionados, foram retiradas 8 categorias de identificação de usuários: indivíduo ou organização, multirredes, malicioso, saúde, comportamento, demografia, interesses e identidade. Também a literatura cinzenta foi consultada para integrar o resultado a respeito do usuário do Twitter e gerou informações como a quantidade de usuários por gênero e os países com mais usuários do Twitter.*

1. Introdução

Existe uma considerável diversidade de redes sociais, com variados propósitos e formas de comunicação, tais quais: texto, foto, vídeo, áudio ou mesmo uma combinação destes recursos. Redes sociais populares como Facebook, Instagram, Tiktok e Twitter, por exemplo, possuem bilhões de usuários ao redor do mundo (respectivamente, 2.93 bilhões, 1.38 bilhões, 1 bilhão e 544 milhões [DataReportal 2023]) nas mais diversas idades, classes, etnias, etc. Essa incrível quantidade e diversidade de usuários gera, diariamente, um enorme montante de dados, que são a matéria-prima da informação.

As redes sociais alcançaram uma abrangência tão ampla e plural que não é de se surpreender que se tornaram objeto de estudo e um vasto campo para se obter conhecimento. Tal atenção e interesse nestes dados não são de hoje. O Orkut, por exemplo, já na sua época, foi tema de alguns estudos como [Chen et al. 2009] e [Spertus et al. 2005].

Atualmente, uma rede social bastante conhecida e utilizada é o Twitter [DataReportal 2023]. O Twitter é um serviço de *microblogging* no qual a comunicação se dá, primordialmente, por meio de textos chamados *tweets*. Segundo a própria empresa: “O Twitter é o lugar certo para saber mais sobre o que está acontecendo e sobre o que as pessoas estão falando agora.” [Twitter 2023]. Assim, o Twitter se apresenta como uma rede de informação em tempo real. No entanto, além da pura e simples informação, esta rede social é largamente utilizada para publicação de opinião, ideias, debates, humor, etc.

O Twitter, assim como outras redes sociais, tem sido campo de pesquisas que passam por problemas técnicos da computação [Tang et al. 2021] a questões psicológicas e sociais [Phand and Phand 2017]. Em face deste conjunto heterogêneo de estudos, foi realizado um mapeamento sistemático da literatura a fim de encontrar respostas à seguinte indagação sobre o usuário do Twitter: quem é este usuário? Responder a esta pergunta significa entender o usuário com base em suas características, seus gostos, seus costumes, sua localização, etc. E, a partir de então, em posse de tais dados, produzir conhecimento.

2. Metodologia

O mapeamento sistemático da literatura (MSL) foi formulado com base na estratégia PICOC [Kitchenham and Charters 2007]. Esta estratégia está descrita na Tabela 1. A busca dos estudos foi realizada via estratégia de busca automatizada nas bibliotecas digitais escolhidas.

Tabela 1. Estratégia PICOC

População	Publicações relacionadas com usuários e Twitter
Intervenção	Avaliação de perfil de usuário
Comparação	Não se aplica
Resultado	Identificação do usuário do Twitter
Contexto	A diversidade de redes sociais e diferentes usos/aplicações

2.1. Questão de Pesquisa

A pergunta precursora deste estudo é assim definida: como a literatura aborda o conhecimento a respeito do usuário do Twitter? Com base neste questionamento, a questão de pesquisa deste trabalho foi estabelecida como: quem é o usuário do Twitter?.

2.2. Fontes e Critérios de Pesquisa

As principais fontes de pesquisa para este estudo foram as bibliotecas digitais *The ACM Digital Library*, *IEEE Xplore Digital Library*, e *Science Direct*. Amplamente reconhecidas pela comunidade acadêmica, são algumas das principais bases para trabalhos de computação. A busca dos artigos foi guiada pela seguinte *string*, para todas as fontes:

[Abstract: profile] AND [Abstract: user] AND [[Abstract: “social media”] OR [Abstract: “social network”]] AND [Abstract: twitter]

Houve somente uma adição de configuração relativa à base da *Science Direct*, a aplicação do seguinte filtro:

Subject areas: Computer Science, Social Sciences, Psychology

Após a execução da *string* de busca, as seguintes quantidades de artigos foram retornadas: *The ACM Digital Library* retornou 132 artigos; *IEEE Xplore Digital Library* retornou 168 artigos; *Science Direct* retornou 80 artigos. Um total de 380 artigos.

Também a literatura cinzenta, por meio do site DataReportal [DataReportal 2023], foi consultada como fonte de pesquisa complementar. A definição desta fonte se deu mediante pesquisa manual realizada no buscador Google [Google 2023]. Para tal consulta, a *string* utilizada foi:

[internet] PLUS [stats] PLUS [research]

Esta *string* pretendia encontrar fontes que realizassem pesquisas estatísticas sobre a Internet e retornou diversas empresas e institutos de pesquisas em nível global, tais como: *Internet World Stats*, *Our World in Data* e *Statista*. Dentre os resultados obtidos, o site do Data Reportal foi considerado o mais adequado por entregar, de forma gratuita, relatórios completos e atuais e também por servir como fonte para outros institutos de pesquisa. O relatório utilizado foi o de outubro de 2022. Este relatório foi o escolhido por ser o último com dados estabelecidos e confiáveis do Twitter. No fim de outubro de 2022, uma transação de compra do Twitter foi concluída e, desde então, a rede social passou por diversas mudanças que impactaram na sua estrutura e na confiabilidade dos dados.

2.2.1. Critérios de Inclusão e Exclusão

Os critérios de seleção, que caracterizam os estudos para inclusão ou exclusão dos artigos [Felizardo et al. 2017], que foram utilizados, são: para inclusão, foram considerados trabalhos que focavam no usuário ‘pessoa física’ do Twitter. Trabalhos que: (1) não respondiam, de alguma forma, a QP01, (2) tratavam de usuários institucionais, como organizações e marcas, (3) possuíam amostra limitada - poucas pessoas, ambientes específicos, etc., (4) eram inconclusivos ou revisões e (5) não possuíam o Twitter em seu corpo, acabaram por serem excluídos. O critérios de seleção estão expostos na Tabela 2.

Tabela 2. Critérios de inclusão e exclusão

Critérios de Inclusão
CI01 - Artigos que tratam do usuário do Twitter.
Critérios de Exclusão
CE01 - Artigos que não respondem a QP1
CE02 - Artigos que tratam de organizações e marcas
CE03 - Artigos com amostra limitada
CE04 - Artigos inconclusivos, revisões
CE05 - Artigos que não possuam o Twitter no seu escopo

2.3. Extração de Dados

Um processo iterativo foi aplicado para incluir ou excluir artigos de acordo com os critérios de seleção. Na primeira fase, os artigos foram incluídos ou excluídos com base na análise do título, resumo e conclusão [Brereton et al. 2007]. Na segunda fase, os papéis duplicados foram removidos. Na terceira fase, estudos que geravam dúvidas foram reavaliados após leitura completa dos textos. Na fase final, os dados foram extraídos, a análise e a interpretação dos dados realizadas. Os dados extraídos dos artigos para a criação da tabela de referência foram: fonte de pesquisa, título, nome dos autores, ano de publicação, resumo e endereço eletrônico onde estão armazenados. Dos 380 artigos retornados após a busca, 105 foram selecionados para compor a coleção final para o estudo. Valor que representa 27,63% do conjunto universo original.

2.4. Criação das Categorias

Os artigos foram separados e agrupados em oito categorias. Cada artigo foi enquadrado em apenas uma categoria, aquela que mais representava o foco principal do artigo. As seguintes categorias foram criadas: indivíduo ou organização, multirredes, malicioso, saúde, comportamento, demografia, interesses e identidade.

Artigos que tinham como foco diferenciar se o usuário do Twitter era uma pessoa física ou uma organização foram catalogados em ‘indivíduo ou organização’. Artigos que pretendiam identificar um mesmo usuário em várias redes sociais distintas, por quaisquer meios que fossem, como, por exemplo, relacionamento entre contas, comparação de atributos ou características extraídas de conteúdos, foram agrupados em ‘multirredes’.

Na categoria ‘maliciosos’ foram incluídos os estudos que se dedicavam a identificar se um usuário era um *bot*, um *fake*, um clone, etc. Também foram incluídos nesta categoria artigos: que buscavam reconhecer perfis terroristas, com comportamentos abusivos, com linguagem abusiva e discurso de ódio, entre outros.

A categoria ‘saúde’ engloba artigos que tratavam de estudos a respeito da saúde física ou psicológica do usuário do Twitter, de identificação e classificação de personalidade, de perfis suicidas, de perfis depressivos, estudos para reconhecimento de estresse, entre outros.

Os artigos que buscavam estudar o comportamento dos usuários do Twitter, principalmente nas questões de influência e autoridade, foram catalogados em ‘comportamento’. Aqueles artigos que se propunham a inferir dados demográficos do usuário do Twitter como gênero, localização, nacionalidade, etc., foram para o grupo ‘demografia’.

Os artigos que buscavam identificar os interesses: dos usuários que usavam o Twitter de forma passiva, implícitos e explícitos dos usuários, ao longo do tempo de uso da rede social, para fins de recomendação de conteúdo e outros estão categorizados como ‘interesses’. Por fim, o artigo que estudava como o usuário do Twitter se identifica socialmente foi colocado em ‘identidade’. A categorização dos artigos pode ser visualizada na Tabela 3. A lista de referência dos artigos pode ser consultada no anexo.

Tabela 3. Categorização

Categorias	Artigos
Indivíduo ou Organização	S01, S02
Multirredes	S03, S04, S05, S06, S07, S08, S09, S10, S11, S12, S13, S14, S15
Maliciosos	S16, S17, S18, S19, S20, S21, S22, S23, S24, S25, S26, S27, S28, S29, S30, S31, S32, S33, S34, S35, S36, S37, S38, S39, S40, S41
Saúde	S42, S43, S44, S45, S46, S47, S48, S49, S50, S51, S52, S53, S54, S55, S56, S57, S58, S59, S60, S61, S62
Comportamento	S63, S64, S65, S66, S67, S68, S69, S70, S71, S72, S73, S74
Demografia	S75, S76, S77, S78, S79, S80, S81, S82, S83, S84, S85, S86, S87, S88, S89, S90, S91, S92, S93, S94, S95, S96, S97
Interesse	S98, S99, S100, S101, S102, S103, S104
Identidade	S105

3. Resultados e Discussões

3.1. Características Gerais

Durante o processo de aplicação do protocolo não foram encontrados estudos que respondessem, de forma direta (com dados demográficos, estatísticos ou quaisquer outras formas de caracterização do usuário), à questão principal desta pesquisa: quem é o usuário do Twitter?. Porém, foram encontrados artigos que respondiam a ela de forma indireta. Por exemplo, se um determinado artigo pretendia inferir o gênero do usuário do Twitter [Ma et al. 2014] e obtinha sucesso, indiretamente, este artigo respondia à questão principal. Desta forma, a seleção e os demais passos continuaram sendo executados.

Finalizada a análise dos dados, foi observado que alguns assuntos eram mais regulares que outros e, a partir de então, as categorias foram criadas e os artigos categorizados. Nota-se que as três maiores frequências de pesquisa estão nas categorias malicioso, demografia e saúde, com 26, 23 e 21 artigos, respectivamente. Em oposição, as três menores frequências estão nas categorias indivíduo ou organização, interesses e identidade, com 2, 7 e 1 artigo, respectivamente, conforme a Figura 1.

Porém, tentar responder de forma indireta à questão guia do MSL acrescenta uma camada de complexidade indesejada. Dessa forma, encontrar respostas diretas seria o ideal. Logo, a literatura cinzenta, que pode ser definida como “...um tipo de literatura que não foi sujeita a mecanismos de controle de qualidade ...” [Kamei et al. 2021] ou uma “...literatura que não pode ser obtida através dos canais normais de publicação...” [Petticrew 2006], foi consultada.

3.2. Das categorias

Na categoria malicioso, a mais popular dentre as categorias, há artigos que se propuseram a detectar perfis falsos no Twitter [Kaubiyal and Jain 2019], a identificar se uma conta é ou não uma conta *bot* (abreviatura de *robot*) [Siddiqui et al. 2017] e a descobrir perfis terroristas [Saidi et al. 2022], entre outros estudos. Tais dados parecem apontar para uma maior preocupação em garantir a segurança, de uma forma geral, nas redes sociais e a partir delas como, por exemplo, o combate à desinformação [Ruffo et al. 2023].

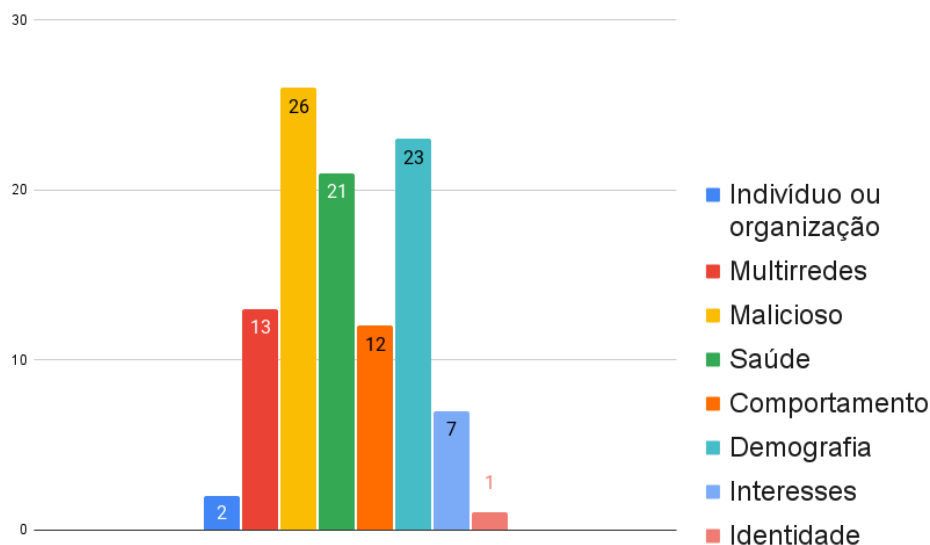


Figura 1. Artigos por categoria

Os estudos da segunda categoria com mais artigos, a categoria demografia, pretendem resolver questões como a de inferir a localização geográfica dos usuários das redes sociais [Zhong et al. 2020] e outras informações como idade e gênero [Ikeda et al. 2013]. De modo diferente da categoria malicioso, da qual se pode insinuar ter um viés de controle, a categoria demografia aparenta ser mais mercadológica.

Já a categoria saúde, a terceira com mais artigos, possui estudos que se propõem a coletar dados para reconhecimento de estresse [Meshram et al. 2020], por exemplo, e estudos cujo objeto central é a personalidade, visando a diversos fins: para inferir profissões e atributos relacionados à personalidade [Wagner et al. 2013], relacionar a personalidade ao tipo de usuário [Quercia et al. 2011] ou para prever a personalidade de um usuário a partir de suas informações disponíveis [Golbeck et al. 2011], entre outros. Esta categoria também apresenta um traço mercadológico.

O conjunto final de artigos ficou restrito a um intervalo de publicação, não intencional, de 12 anos, de 2011 a 2022 (contados em intervalos fechados) e, ao se observar a distribuição dos artigos no decorrer do anos, alguns dados chamam atenção. Verifica-se, por exemplo, o crescimento na quantidade de estudos da categoria malicioso a partir de 2017. Talvez um reflexo das eleições norte-americanas de 2016 [Chatfield et al. 2017], a partir da qual o termo se popularizou e ficou em evidência.

Na categoria demografia, regular de 2012 a 2016, há uma redução de papéis a partir de 2016. Neste ano, o Parlamento Europeu adotou o Regulamento Geral sobre a Proteção de Dados (GDPR). A União Europeia estabeleceu para seus estados-membros que estes somente poderiam realizar comércio e/ou serviço (que incluíssem dados pessoais) com outros países cuja legislação fosse minimamente equiparada com a deles.

Observa-se que os dois únicos estudos da categoria indivíduo ou organização posuem entre si um hiato de 3 anos. Um estudo de 2017 e outro de 2021. Também que a categoria saúde possui uma quantidade baixa de artigos por ano, porém frequente, tendo

ficado sem publicações apenas em 2018, mas com crescimento em 2022.

Dos dados do relatório do Data Reportal, é possível verificar as porcentagens de usuários multirredes do Twitter. Ou seja, usuários do Twitter que utilizam, concomitantemente ao Twitter, outras redes sociais. Esta categoria, apesar de apresentar um baixo volume de artigos por ano, é uma categoria que se mostra regular durante os anos.

3.3. Dos usuários

Destacam-se, do relatório do Data Reportal, as seguintes informações: a predileção do Twitter, por gênero, Figura 2, na qual se verifica que a maioria dos usuários do Twitter são pessoas jovens, faixa etária de 16 a 24 anos. E a porcentagem de atividades realizadas no Twitter, na qual, a principal atividade realizada, por 61,2% dos usuários, é a de “atualizar-se com notícias e eventos atuais”, ou seja, a busca por informações.

Ao se considerar estas informações e relacioná-las com o mundo real, uma conclusão possível, e amplamente observada nos últimos anos, é a ligação desta rede social com a profusão de *fake news* [Balestrucci and De Nicola 2020]. Outra conclusão que divide o mesmo grau de possibilidade é uma resposta para o porquê da quantidade de artigos na classificação ‘maliciosos’. É aceitável inferir que tais artigos podem ser tentativas de identificação e combate ao mal das *fake news* [Pizarro 2020].

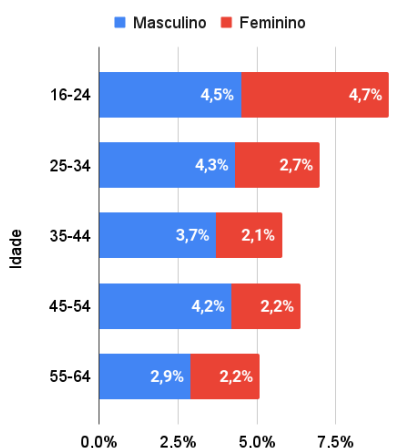


Figura 2. Predileção por gênero

Ainda segundo o Data Reportal, os Estados Unidos é o país com mais usuários no mundo¹, seguido por Japão, Índia, Indonésia e Brasil, que ocupa a quinta colocação do *ranking*. Quanto ao sexo, há uma predominância de usuários do sexo masculino em relação ao sexo feminino.

4. Limitações da Pesquisa

Duas limitações observadas nesta pesquisa, que se referem à metodologia, são: a falta da definição das ameaças à validade do estudo, que não foi realizada, principalmente, porque este mapeamento foi desenvolvido, majoritariamente, por apenas uma pessoa - condição

¹Os números podem não representar usuários únicos ou correspondência à base total de usuários ativos. Há de se considerar duplicidade de contas e contas falsas.

que por si própria também é uma limitação da pesquisa; e o modesto refinamento da *string* de busca devido à falta de um artigo de controle que respondesse de forma direta, em algum nível, a questão guia deste estudo. Pode-se ainda citar o fato de a categorização ser qualitativa e à venda do Twitter que gerou uma série de alterações que impactaram na confiabilidade dos dados.

5. Pesquisas Relacionadas

Conjuntos de dados de alcance mundial apresentam grandes oportunidades para pesquisadores. Com dados de redes sociais mundiais, como o Twitter e o Facebook, que são alimentados de forma orgânica por seus usuários, não é diferente.

Diversos estudos que utilizam o Twitter como fonte de dados objetivaram caracterizar ou agrupar, de alguma forma, os usuários dessa rede. Há artigos que utilizaram os dados do Twitter para encontrar possíveis diferenças e semelhanças no uso desta plataforma em diversos países [Poblete et al. 2011]. Outros propuseram uma estrutura para agrupar usuários semelhantes a fim de formar uma comunidade visando a “identificar, extrair e caracterizar componentes proeminentes de tópicos e opiniões que as pessoas compartilham...”, neste caso, “...durante eventos de crise no Twitter” [Gupta et al. 2012]. Há outros, por exemplo, que visavam à recomendação de amizade baseada na correspondência de perfis [Choumane et al. 2017] e ainda outros que utilizaram entrevistas com intuito de verificarem a perspectiva de uma parcela específica de usuários [Brough et al. 2020]. Todos estes casos trazem, de algum modo, um processo de identificação e caracterização do usuário do Twitter que, no entanto, não corresponde diretamente ao objetivo deste estudo. Assim, no processo deste mapeamento, não foram encontrados estudos que se propusessem a gerar uma caracterização ampla do usuário do twitter.

6. Considerações Finais

Este estudo apresentou o mapeamento sistemático da literatura que pretendia responder à pergunta: quem é o usuário do Twitter? Verificou-se que, do modo como o mapeamento foi conduzido, não foram encontrados, nas bases científicas utilizadas, artigos que respondessem de forma direta à pergunta principal deste trabalho. No entanto, alguns resultados relevantes foram encontrados na literatura cinzenta. É importante salientar que boa parte dos esforços para identificar o usuário do Twitter, e não apenas do Twitter, mas das redes sociais em geral, tem inspirações, pretensões e finalidades diversas. Para trabalhos futuros, há o intuito de pesquisas, complementares a esta, que busquem responder a perguntas como: Qual a faixa etária do usuário que semeia desinformação? O usuário do Twitter age da mesma forma em diferentes redes sociais? E, há como criar alguma persona ou personas dos usuários do twitter, agrupando-os por comportamentos distintos?

Referências

- Abbasi, R., Rehman, G., Lee, J., Riaz, F. M., and Luo, B. (2017). Discovering temporal user interest on twitter by using semantic based dynamic interest finding model (tut). In *2017 IEEE 2nd Information Technology, Networking, Electronic and Automation Control Conference (ITNEC)*, pages 743–747.
- Abrol, S., Khan, L., and Thuraisingham, B. (2012a). Tweecalization: Efficient and intelligent location mining in twitter using semi-supervised learning. In *8th International*

Conference on Collaborative Computing: Networking, Applications and Worksharing (CollaborateCom), pages 514–523.

Abrol, S., Khan, L., and Thuraisingham, B. (2012b). Tweeque: Spatio-temporal analysis of social networks for location mining using graph partitioning. In *2012 International Conference on Social Informatics*, pages 145–148.

Agarwal, S. and Mehta, S. (2020). Effective influence estimation in twitter using temporal, profile, structural and interaction characteristics. *Information Processing Management*, 57(6):102321.

Ahmad, N. and Siddique, J. (2017). Personality assessment using twitter tweets. *Procedia Computer Science*, 112:1964–1973. Knowledge-Based and Intelligent Information Engineering Systems: Proceedings of the 21st International Conference, KES-20176-8 September 2017, Marseille, France.

Ahmad, W. and Ali, R. (2018). Understanding the users personal attributes selection tendency across social networks. In *2018 3rd International Conference On Internet of Things: Smart Innovation and Usages (IoT-SIU)*, pages 1–6.

Al Maruf, H., Meshkat, N., Ali, M. E., and Mahmud, J. (2015). Human behaviour in different social medias: A case study of twitter and disqus. In *2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, pages 270–273.

Al-Qurishi, M., Hossain, M. S., Alrubaian, M., Rahman, S. M. M., and Alamri, A. (2018). Leveraging analysis of user behavior to identify malicious activities in large-scale social networks. *IEEE Transactions on Industrial Informatics*, 14(2):799–813.

Al-Zoubi, A. M., Faris, H., Alqatawna, J., and Hassonah, M. A. (2018). Evolving support vector machines using whale optimization algorithm for spam profiles detection on online social networks in different lingual contexts. *Knowledge-Based Systems*, 153:91–104.

Alami, S. and Elbeqqali, O. (2015). Cybercrime profiling: Text mining techniques to detect and predict criminal activities in microblog posts. In *2015 10th International Conference on Intelligent Systems: Theories and Applications (SITA)*, pages 1–5.

Alharthi, R., Alhothali, A., and Moria, K. (2019). Detecting and characterizing arab spammers campaigns in twitter. *Procedia Computer Science*, 163:248–256. 16th Learning and Technology Conference 2019 Artificial Intelligence and Machine Learning: Embedding the Intelligence.

Alp, Z. Z. and Ögüdücü, c. G. (2016). Influential user detection on twitter: Analyzing effect of focus rate. In *Proceedings of the 2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, ASONAM '16*, page 1321–1328. IEEE Press.

Alvari, H., Beigi, G., Sarkar, S., Ruston, S. W., Corman, S. R., Davulcu, H., and Shakerian, P. (2020). A feature-driven approach for identifying pathogenic social media accounts. In *2020 3rd International Conference on Data Intelligence and Security (ICDIS)*, pages 26–33.

- Alvari, H., Sarkar, S., and Shakarian, P. (2019a). Detection of violent extremists in social media. In *2019 2nd International Conference on Data Intelligence and Security (ICDIS)*, pages 43–47.
- Alvari, H., Shaabani, E., Sarkar, S., Beigi, G., and Shakarian, P. (2019b). Less is more: Semi-supervised causal inference for detecting pathogenic users in social media. In *Companion Proceedings of The 2019 World Wide Web Conference, WWW '19*, page 154–161, New York, NY, USA. Association for Computing Machinery.
- Ardehaly, E. M. and Culotta, A. (2017). Co-training for demographic classification using deep learning from label proportions. In *2017 IEEE International Conference on Data Mining Workshops (ICDMW)*, pages 1017–1024.
- Balestrucci, A. and De Nicola, R. (2020). Credulous users and fake news: a real case study on the propagation in twitter. In *2020 IEEE Conference on Evolving and Adaptive Intelligent Systems (EAIS)*, pages 1–8.
- Bin Tareaf, R., Alhosseini, S. A., and Meinel, C. (2019). Facial-based personality prediction models for estimating individuals private traits. In *2019 IEEE Intl Conf on Parallel Distributed Processing with Applications, Big Data Cloud Computing, Sustainable Computing Communications, Social Computing Networking (ISPA/BDCloud/SocialCom/SustainCom)*, pages 1586–1594.
- Biswas, K., Shivakumara, P., Pal, U., Chakraborti, T., Lu, T., and Ayub, M. N. B. (2022). Fuzzy and genetic algorithm based approach for classification of personality traits oriented social media images. *Knowledge-Based Systems*, 241:108024.
- Brereton, P., Kitchenham, B. A., Budgen, D., Turner, M., and Khalil, M. (2007). Lessons from applying the systematic literature review process within the software engineering domain. *Journal of Systems and Software*, 80(4):571–583. Software Performance.
- Brough, M., Literat, I., and Ikin, A. (2020). “good social media?”: Underrepresented youth perspectives on the ethical and equitable design of social media platforms. *Social Media + Society*, 6(2):2056305120928488.
- Burger, J. D., Henderson, J., Kim, G., and Zarrella, G. (2011). Discriminating gender on twitter. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing, EMNLP '11*, page 1301–1309, USA. Association for Computational Linguistics.
- Castellini, J., Poggioni, V., and Sorbi, G. (2017). Fake twitter followers detection by denoising autoencoder. In *Proceedings of the International Conference on Web Intelligence, WI '17*, page 195–202, New York, NY, USA. Association for Computing Machinery.
- Chatfield, A. T., Reddick, C. G., and Choi, K. P. (2017). Online media use of false news to frame the 2016 trump presidential campaign. In *Proceedings of the 18th Annual International Conference on Digital Government Research, dg.o '17*, page 213–222, New York, NY, USA. Association for Computing Machinery.
- Chen, J., Liu, Y., and Zou, M. (2016). Home location profiling for users in social media. *Information Management*, 53(1):135–143.

- Chen, W.-Y., Chu, J.-C., Luan, J., Bai, H., Wang, Y., and Chang, E. Y. (2009). Collaborative filtering for orkut communities: Discovery of user latent behavior. In *Proceedings of the 18th International Conference on World Wide Web, WWW '09*, page 681–690, New York, NY, USA. Association for Computing Machinery.
- Choumane, A., Al Abidin Ibrahim, Z., and Chebaro, B. (2017). Correspondência de perfis em redes sociais com base em semelhanças semânticas e relacionamentos comuns. In *Proceedings of the International Conference on Compute and Data Analysis, ICCDA '17*, page 14–18, Nova York, NY, EUA. Association for Computing Machinery.
- Coletto, M., Lucchese, C., and Orlando, S. (2018). Do violent people smile: Social media analysis of their profile pictures. In *Companion Proceedings of the The Web Conference 2018, WWW '18*, page 1465–1468, Republic and Canton of Geneva, CHE. International World Wide Web Conferences Steering Committee.
- Correa, D., Sureka, A., and Sethi, R. (2012). Whacky! - what anyone could know about you from twitter. In *2012 Tenth Annual International Conference on Privacy, Security and Trust*, pages 43–50.
- Das, U., Narayanan, A., Gupta, A., Bagga, O. S., and Chopra, S. (2018). Social champion identification for ngos. In *2018 2nd International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC)I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC), 2018 2nd International Conference on*, pages 361–366.
- DataReportal (2023). Digital 2022 - october global. Último acesso em 01 de junho de 2023.
- Efstathiades, H., Antoniadis, D., Pallis, G., and Dikaiakos, M. D. (2015). Identification of key locations based on online social network activity. In *2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, pages 218–225.
- Elmendili, F., Chaoui, H., and Bouzekri El Idrissi, Y. E. (2019). *Social Network's Security Related to Healthcare*, pages 91–113.
- Faralli, S., Stilo, G., and Velardi, P. (2017). Automatic acquisition of a taxonomy of microblogs users' interests. *Journal of Web Semantics*, 45:23–40.
- Felizardo, K. R., Nakagawa, E. Y., Fabbri, S. C. P. F., and Ferrari, F. C. (2017). *Revisão sistemática da literatura em engenharia de software: teoria e prática*. Elsevier.
- Fernández, D., Moctezuma, D., and Siordia, O. S. (2016). Features combination for gender recognition on twitter users. In *2016 IEEE International Autumn Meeting on Power, Electronics and Computing (ROPEC)*, pages 1–6.
- Firdaus, S. N., Ding, C., and Sadeghian, A. (2021). Retweet prediction based on topic, emotion and personality. *Online Social Networks and Media*, 25:100165.
- Gao, M., Lim, E.-P., Lo, D., Zhu, F., Prasetyo, P. K., and Zhou, A. (2015). Cnl: Collective network linkage across heterogeneous social platforms. In *2015 IEEE International Conference on Data Mining*, pages 757–762.
- Golbeck, J., Robles, C., Edmondson, M., and Turner, K. (2011). Predicting personality from twitter. In *2011 IEEE Third International Conference on Privacy, Security, Risk*

and Trust and 2011 IEEE Third International Conference on Social Computing, pages 149–156.

Google (2023). Pesquisa realizada em 20 de abril de 2023.

Gupta, A., Joshi, A., and Kumaraguru, P. (2012). Identifying and characterizing user communities on twitter during crisis events. In *Proceedings of the 2012 Workshop on Data-Driven User Behavioral Modelling and Mining from Social Media*, DUBMMSM '12, page 23–26, New York, NY, USA. Association for Computing Machinery.

Gutierrez, F. J. and Poblete, B. (2015). Sentiment-based user profiles in microblogging platforms. In *Proceedings of the 26th ACM Conference on Hypertext & Social Media*, HT '15, page 23–32, New York, NY, USA. Association for Computing Machinery.

Han Veiga, M. and Eickhoff, C. (2016). A cross-platform collection of social network profiles. In *Proceedings of the 39th International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '16, page 665–668, New York, NY, USA. Association for Computing Machinery.

Harrison, J., Bell, E., Corley, C., Dowling, C., and Cowell, A. (2015). Assessment of user home location geoinference methods. In *2015 IEEE International Conference on Intelligence and Security Informatics (ISI)*, pages 148–150.

Hayama, T. (2022). Analyzing features of passive twitter's users to estimate passive twitter-user's interests. In *IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology*, WI-IAT '21, page 476–481, New York, NY, USA. Association for Computing Machinery.

He, S., Wang, H., and Jiang, Z. H. (2014a). Identifying user behavior on twitter based on multi-scale entropy. In *Proceedings 2014 IEEE International Conference on Security, Pattern Analysis, and Cybernetics (SPAC)*, pages 381–384.

He, S., Wang, H., and Jiang, Z. H. (2014b). Identifying user behavior on twitter based on multi-scale entropy. In *Proceedings 2014 IEEE International Conference on Security, Pattern Analysis, and Cybernetics (SPAC)*, pages 381–384.

Heidari, M., Jones, J. H., and Uzuner, O. (2020). Deep contextualized word embedding for text-based online user profiling to detect social bots on twitter. In *2020 International Conference on Data Mining Workshops (ICDMW)*, pages 480–487.

Herzig, J., Mass, Y., and Roitman, H. (2014). An author-reader influence model for detecting topic-based influencers in social media. In *Proceedings of the 25th ACM Conference on Hypertext and Social Media*, HT '14, page 46–55, New York, NY, USA. Association for Computing Machinery.

Huang, W., Weber, I., and Vieweg, S. (2014). Inferring nationalities of twitter users and studying inter-national linking. In *Proceedings of the 25th ACM Conference on Hypertext and Social Media*, HT '14, page 237–242, New York, NY, USA. Association for Computing Machinery.

Humadde, H. S., Abdul-Hassan, A. K., and Mahdi, B. S. (2019). Proposed user identification algorithm across social network using hybrid techniques. In *2019 2nd Scientific Conference of Computer Sciences (SCCS)*, pages 158–161.

- Huynh, H. N., Legara, E. F., and Monterola, C. (2015). A dynamical model of twitter activity profiles. In *Proceedings of the 26th ACM Conference on Hypertext & Social Media*, HT '15, page 49–57, New York, NY, USA. Association for Computing Machinery.
- Ikeda, K., Hattori, G., Ono, C., Asoh, H., and Higashino, T. (2013). Twitter user profiling based on text and community mining for market analysis. *Knowledge-Based Systems*, 51:35–47.
- Jain, P., Kumaraguru, P., and Joshi, A. (2013). @i seek 'fb.me': Identifying users across multiple online social networks. In *Proceedings of the 22nd International Conference on World Wide Web*, WWW '13 Companion, page 1259–1268, New York, NY, USA. Association for Computing Machinery.
- Jain, P., Kumaraguru, P., and Joshi, A. (2015). Other times, other values: Leveraging attribute history to link user profiles across online social networks. In *Proceedings of the 26th ACM Conference on Hypertext & Social Media*, HT '15, page 247–255, New York, NY, USA. Association for Computing Machinery.
- Kamei, F., Wiese, I., Lima, C., Polato, I., Nepomuceno, V., Ferreira, W., Ribeiro, M., Pena, C., Cartaxo, B., Pinto, G., and Soares, S. (2021). Grey literature in software engineering: A critical review. *Information and Software Technology*, 138:106609.
- Kantepe, M. and Ganiz, M. C. (2017). Preprocessing framework for twitter bot detection. In *2017 International Conference on Computer Science and Engineering (UBMK)*, pages 630–634.
- Karamitsos, I., Mohasseb, A., and Kanavos, A. (2022). A graph mining method for characterizing and measuring user engagement in twitter. In *2022 17th International Workshop on Semantic and Social Media Adaptation Personalization (SMAP)*, pages 1–6.
- Kasbekar, P., Potika, K., and Pollett, C. (2020). Find me if you can: Aligning users in different social networks. In *2020 IEEE Sixth International Conference on Big Data Computing Service and Applications (BigDataService)*, pages 46–53.
- Kashyap, R. and Nahapetian, A. (2014). Tweet analysis for user health monitoring. In *2014 4th International Conference on Wireless Mobile Communication and Healthcare - Transforming Healthcare Through Innovations in Mobile and Wireless Technologies (MOBIHEALTH)*, pages 348–351.
- Kaubiyal, J. and Jain, A. K. (2019). A feature based approach to detect fake profiles in twitter. In *Proceedings of the 3rd International Conference on Big Data and Internet of Things*, BDIOT 2019, page 135–139, New York, NY, USA. Association for Computing Machinery.
- Kim, S. M., Paris, C., Power, R., and Wan, S. (2017). Distinguishing individuals from organisations on twitter. *WWW '17 Companion*, page 805–806, Republic and Canton of Geneva, CHE. International World Wide Web Conferences Steering Committee.
- Kitchenham, B. and Charters, S. (2007). *Guidelines for Performing Systematic Literature Reviews in Software Engineering*.

- Kostakos, P., Pandya, A., Kyriakouli, O., and Oussalah, M. (2018). Inferring demographic data of marginalized users in twitter with computer vision apis. In *2018 European Intelligence and Security Informatics Conference (EISIC)*, pages 81–84.
- Kotzias, D., Lappas, T., and Gunopulos, D. (2016). Home is where your friends are: Utilizing the social graph to locate twitter users in a city. *Information Systems*, 57:77–87.
- Kumar, S., Hu, X., and Liu, H. (2014). A behavior analytics approach to identifying tweets from crisis regions. In *Proceedings of the 25th ACM Conference on Hypertext and Social Media, HT '14*, page 255–260, New York, NY, USA. Association for Computing Machinery.
- Kwon, S., Liang, P., Tandon, S., Berman, J., Chang, P.-j., and Gilbert, E. (2018). Tweety holmes: A browser extension for abusive twitter profile detection. In *Companion of the 2018 ACM Conference on Computer Supported Cooperative Work and Social Computing, CSCW '18*, page 17–20, New York, NY, USA. Association for Computing Machinery.
- Li, R., Wang, S., Deng, H., Wang, R., and Chang, K. C.-C. (2012). Towards social user profiling: Unified and discriminative influence model for inferring home locations. In *Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '12*, page 1023–1031, New York, NY, USA. Association for Computing Machinery.
- Ma, X., Tsuboshita, Y., and Kato, N. (2014). Gender estimation for sns user profiling using automatic image annotation. In *2014 IEEE International Conference on Multimedia and Expo Workshops (ICMEW)*, pages 1–6.
- Malhotra, A., Totti, L., Meira Jr., W., Kumaraguru, P., and Almeida, V. (2012). Studying user footprints in different online social networks. In *2012 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, pages 1065–1070.
- Masood, M. A. and Abbasi, R. A. (2021). Using graph embedding and machine learning to identify rebels on twitter. *Journal of Informetrics*, 15(1):101121.
- Mbarek, A., Jamoussi, S., and Hamadou, A. B. (2022). An across online social networks profile building approach: Application to suicidal ideation detection. *Future Generation Computer Systems*, 133:171–183.
- McGee, J., Caverlee, J., and Cheng, Z. (2013). Location prediction in social media based on tie strength. In *Proceedings of the 22nd ACM International Conference on Information & Knowledge Management, CIKM '13*, page 459–468, New York, NY, USA. Association for Computing Machinery.
- Merler, M., Cao, L., and Smith, J. R. (2015). You are what you tweet... pic! gender prediction based on semantic analysis of social media images. In *2015 IEEE International Conference on Multimedia and Expo (ICME)*, pages 1–6.
- Meshram, S., Babu, R., and Adhikari, J. (2020). Detecting psychological stress using machine learning over social media interaction. In *2020 5th International Conference on Communication and Electronics Systems (ICCES)*, pages 646–649.

- NAJI, M., Najima, D., Hasnae, R., and AJHOUN, R. (2019). Customized data extraction and processing for the prediction of baby blues from social media. In *2019 1st International Conference on Smart Systems and Data Science (ICSSD)*, pages 1–6.
- Narayanan, A., Garg, A., Arora, I., Sureka, T., Sridhar, M., and Prasad, H. (2018). Ironsense: Towards the identification of fake user-profiles on twitter using machine learning. In *2018 Fourteenth International Conference on Information Processing (ICINPRO)*, pages 1–7.
- Nargundkar, A. and Rao, Y. S. (2016). Influencerank: A machine learning approach to measure influence of twitter users. In *2016 International Conference on Recent Trends in Information Technology (ICRTIT)*, pages 1–6.
- Nie, Y., Jia, Y., Li, S., Zhu, X., Li, A., and Zhou, B. (2016). Identifying users across social networks based on dynamic core interests. *Neurocomputing*, 210:107–115. SI:Behavior Analysis In SN.
- Pakaya, F. N., Ibrohim, M. O., and Budi, I. (2019). Malicious account detection on twitter based on tweet account features using machine learning. In *2019 Fourth International Conference on Informatics and Computing (ICIC)*, pages 1–5.
- Pathak, A., Madani, N., and Joseph, K. (2021). A method to analyze multiple social identities in twitter bios. 5(CSCW2).
- Petticrew, M., R.-H. (2006). *Systematic reviews in the social sciences : a practical guide*. BLACKWELL PUBLISHING, page 90.
- Pham, P., Nguyen, L. T., Vo, B., and Yun, U. (2022). Bot2vec: A general approach of intra-community oriented representation learning for bot detection in different types of social networks. *Information Systems*, 103:101771.
- Phand, S. A. and Phand, J. A. (2017). Twitter sentiment classification using stanford nlp. In *2017 1st International Conference on Intelligent Systems and Information Management (ICISIM)*, pages 1–5.
- Piao, G. (2021). A simple language independent approach for distinguishing individuals on social media. HT '21, page 251–256, New York, NY, USA. Association for Computing Machinery.
- Pizarro, J. (2020). Profiling bots and fake news spreaders at pan'19 and pan'20 : Bots and gender profiling 2019, profiling fake news spreaders on twitter 2020. In *2020 IEEE 7th International Conference on Data Science and Advanced Analytics (DSAA)*, pages 626–630.
- Poblete, B., Garcia, R., Mendoza, M., and Jaimes, A. (2011). Do all birds tweet the same? characterizing twitter around the world. In *Proceedings of the 20th ACM International Conference on Information and Knowledge Management, CIKM '11*, page 1025–1030, New York, NY, USA. Association for Computing Machinery.
- Pratama, R. P. and Maharani, W. (2021). Predicting big five personality traits based on twitter user using random forest method. In *2021 International Conference on Data Science and Its Applications (ICoDSA)*, pages 110–117.
- Preotiuc-Pietro, D., Carpenter, J., Giorgi, S., and Ungar, L. (2016). Studying the dark triad of personality through twitter behavior. In *Proceedings of the 25th ACM Interna-*

- tional on Conference on Information and Knowledge Management, CIKM '16*, page 761–770, New York, NY, USA. Association for Computing Machinery.
- Quercia, D., Kosinski, M., Stillwell, D., and Crowcroft, J. (2011). Our twitter profiles, our selves: Predicting personality with twitter. In *2011 IEEE Third International Conference on Privacy, Security, Risk and Trust and 2011 IEEE Third International Conference on Social Computing*, pages 180–185.
- Roedler, R., Kergl, D., and Rodosek, G. D. (2017). Content driven profile matching across online social networks. In *Proceedings of the 2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2017, ASONAM '17*, page 1049–1055, New York, NY, USA. Association for Computing Machinery.
- Ruffo, G., Semeraro, A., Giachanou, A., and Rosso, P. (2023). Studying fake news spreading, polarisation dynamics, and manipulation by bots: A tale of networks and language. *Computer Science Review*, 47:100531.
- Sahoo, S. R. and Gupta, B. (2019). Hybrid approach for detection of malicious profiles in twitter. *Computers Electrical Engineering*, 76:65–81.
- Saidi, F., Trabelsi, Z., and Thangaraj, E. (2022). A novel framework for semantic classification of cyber terrorist communities on twitter. *Engineering Applications of Artificial Intelligence*, 115:105271.
- Shah, C., Thacker, C., and Patel, Y. (2022). Multi-label personality prediction on twitter data using machine learning. In *2022 International Mobile and Embedded Technology Conference (MECON)*, pages 140–144.
- Shinde, S. and Mane, S. B. (2021). Malicious profile detection on social media: A survey paper. In *2021 9th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO)*, pages 1–5.
- Si, H., Zhou, J., Chen, Z., Wan, J., Xiong, N. N., Zhang, W., and Vasilakos, A. V. (2019). Association rules mining among interests and applications for users on social networks. *IEEE Access*, 7:116014–116026.
- Siddiqui, H., Healy, E., and Olmsted, A. (2017). Bot or not. In *2017 12th International Conference for Internet Technology and Secured Transactions (ICITST)*, pages 462–463.
- Sowmya, P. and Chatterjee, M. (2020). Detection of fake and clone accounts in twitter using classification and distance measure algorithms. In *2020 International Conference on Communication and Signal Processing (ICCSP)*, pages 0067–0070.
- Spertus, E., Sahami, M., and Buyukkokten, O. (2005). Evaluating similarity measures: A large-scale study in the orkut social network. In *Proceedings of the Eleventh ACM SIGKDD International Conference on Knowledge Discovery in Data Mining, KDD '05*, page 678–684, New York, NY, USA. Association for Computing Machinery.
- Srivastava, D. K. and Roychoudhury, B. (2020). Words are important: A textual content based identity resolution scheme across multiple online social networks. *Knowledge-Based Systems*, 195:105624.

- Suman, C., Naman, A., Saha, S., and Bhattacharyya, P. (2021). A multimodal author profiling system for tweets. *IEEE Transactions on Computational Social Systems*, 8(6):1407–1416.
- Sumner, C., Byers, A., Boochever, R., and Park, G. J. (2012). Predicting dark triad personality traits from twitter usage and a linguistic analysis of tweets. In *2012 11th International Conference on Machine Learning and Applications*, volume 2, pages 386–393.
- Suriakala, M. and Revathi, S. (2018). Privacy protected system for vulnerable users and cloning profile detection using data mining approaches. In *2018 Tenth International Conference on Advanced Computing (ICoAC)*, pages 124–132.
- Tang, C., Wang, B., Luo, Z., Wu, H., Dasan, S., Fu, M., Li, Y., Ghosh, M., Kabra, R., Navadiya, N. K., Cheng, D., Dai, F., Channapattan, V., and Mishra, P. (2021). Forecasting sql query cost at twitter. In *2021 IEEE International Conference on Cloud Engineering (IC2E)*, pages 154–160.
- Tundis, A. and Mühlhäuser, M. (2017). A multi-language approach towards the identification of suspicious users on social networks. In *2017 International Carnahan Conference on Security Technology (ICCSST)*, pages 1–6.
- Twitter (2023). About twitter. Último acesso em 01 de junho de 2023.
- Utami, E., Hartanto, A. D., Adi, S., Oyong, I., and Raharjo, S. (2022). Profiling analysis of disc personality traits based on twitter posts in bahasa indonesia. *Journal of King Saud University - Computer and Information Sciences*, 34(2):264–269.
- Utami, E., Iskandar, A. F., Hartanto, A. D., and Raharjo, S. (2021). Disc personality classification using twitter: Usability testing. In *2021 IEEE 5th International Conference on Information Technology, Information Systems and Electrical Engineering (ICITISEE)*, pages 180–185.
- Velayutham, T. and Tiwari, P. K. (2017). Bot identification: Helping analysts for right data in twitter. In *2017 3rd International Conference on Advances in Computing, Communication Automation (ICACCA) (Fall)*, pages 1–5.
- Volkova, S., Bachrach, Y., and Van Durme, B. (2016). Mining user interests to predict perceived psycho-demographic traits on twitter. In *2016 IEEE Second International Conference on Big Data Computing Service and Applications (BigDataService)*, pages 36–43.
- Wagner, C., Asur, S., and Hailpern, J. (2013). Religious politicians and creative photographers: Automatic user categorization in twitter. In *2013 International Conference on Social Computing*, pages 303–310.
- Wandabwa, H. M., Naeem, M. A., Mirza, F., and Pears, R. (2021). Topical affinity in short text microblogs. *Information Systems*, 96:101662.
- Wang, T., Brede, M., Ianni, A., and Mentzakis, E. (2017). Detecting and characterizing eating-disorder communities on social media. In *Proceedings of the Tenth ACM International Conference on Web Search and Data Mining, WSDM '17*, page 91–100, New York, NY, USA. Association for Computing Machinery.

- Wang, X., Zhang, H., Wang, Z., Qiao, Y., Ma, J., and Dai, H. (2021). Con&net: A cross-network anchor link discovery method based on embedding representation. *ACM Trans. Knowl. Discov. Data*, 16(2).
- Wickramaratna, N. C., Jayasiriwardena, T. D., Wijesekara, M., Munasinghe, P. B., and Ganegoda, G. U. (2020). A framework to detect twitter platform manipulation and computational propaganda. In *2020 20th International Conference on Advances in ICT for Emerging Regions (ICTer)*, pages 214–219.
- Wijesekara, M. and Ganegoda, G. U. (2020). Source credibility analysis on twitter users. In *2020 International Research Conference on Smart Computing and Systems Engineering (SCSE)*, pages 96–102.
- Yamaguchi, Y., Amagasa, T., and Kitagawa, H. (2013). Landmark-based user location inference in social media. In *Proceedings of the First ACM Conference on Online Social Networks, COSN '13*, page 223–234, New York, NY, USA. Association for Computing Machinery.
- Yamaguchi, Y., Amagasa, T., Kitagawa, H., and Ikawa, Y. (2014). Online user location inference exploiting spatiotemporal correlations in social streams. In *Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management, CIKM '14*, page 1139–1148, New York, NY, USA. Association for Computing Machinery.
- Yazdavar, A. H., Al-Olimat, H. S., Ebrahimi, M., Bajaj, G., Banerjee, T., Thirunarayan, K., Pathak, J., and Sheth, A. (2017). Semi-supervised approach to monitoring clinical depressive symptoms in social media. In *Proceedings of the 2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2017, ASONAM '17*, page 1191–1198, New York, NY, USA. Association for Computing Machinery.
- Yu, Z., Yi, F., Lv, Q., and Guo, B. (2018). Identifying on-site users for social events: Mobility, content, and social relationship. *IEEE Transactions on Mobile Computing*, 17(9):2055–2068.
- Zarrinkalam, F., Kahani, M., and Bagheri, E. (2018). Mining user interests over active topics on social networks. *Information Processing Management*, 54(2):339–357.
- Zengin Alp, Z. and Şule Gündüz Öğüdücü (2018). Identifying topical influencers on twitter based on user behavior and network topology. *Knowledge-Based Systems*, 141:211–221.
- Zhao, J., Gou, L., Wang, F., and Zhou, M. (2014). Pearl: An interactive visual analytic tool for understanding personal emotion style derived from social media. In *2014 IEEE Conference on Visual Analytics Science and Technology (VAST)*, pages 203–212.
- Zhong, T., Wang, T., Wang, J., Wu, J., and Zhou, F. (2020). Multiple-aspect attentional graph neural networks for online social network user localization. *IEEE Access*, 8:95223–95234.

Anexo

Tabela 4. Lista dos Artigos Selecionados

Id	Referência	Id	Referência	Id	Referência
S01	[Piao 2021]	S02	[Kim et al. 2017]	S03	[Wang et al. 2021]
S04	[Roedler et al. 2017]	S05	[Han Veiga and Eickhoff 2016]	S06	[Jain et al. 2015]
S07	[Jain et al. 2013]	S08	[Srivastava and Roychoudhury 2020]	S09	[Nie et al. 2016]
S10	[Kasbekar et al. 2020]	S11	[Humadde et al. 2019]	S12	[Ahmad and Ali 2018]
S13	[Gao et al. 2015]	S14	[Malhotra et al. 2012]	S15	[Correa et al. 2012]
S16	[Kaubiya and Jain 2019]	S17	[Castellini et al. 2017]	S18	[Pham et al. 2022]
S19	[Shinde and Mane 2021]	S20	[Heidari et al. 2020]	S21	[Sowmya and Chatterjee 2020]
S22	[Pakaya et al. 2019]	S23	[Narayanan et al. 2018]	S24	[Siddiqui et al. 2017]
S25	[Velayutham and Tiwari 2017]	S26	[Kantepe and Ganiz 2017]	S27	[Al-Zoubi et al. 2018]
S28	[Alvari et al. 2019b]	S29	[Alvari et al. 2020]	S30	[Sahoo and Gupta 2019]
S31	[Elmendili et al. 2019]	S32	[Suriakala and Revathi 2018]	S33	[Al-Qurishi et al. 2018]
S34	[Alami and Elbeqqali 2015]	S35	[Alharthi et al. 2019]	S36	[Coletto et al. 2018]
S37	[Alvari et al. 2019a]	S38	[Saidi et al. 2022]	S39	[Tundis and Mühlhäuser 2017]
S40	[Masood and Abbasi 2021]	S41	[Kwon et al. 2018]	S42	[Wang et al. 2017]
S43	[Yazdavar et al. 2017]	S44	[Preotiuc-Pietro et al. 2016]	S45	[Gutierrez and Poblete 2015]
S46	[Mbarek et al. 2022]	S47	[Biswas et al. 2022]	S48	[Ahmad and Siddique 2017]
S49	[Shah et al. 2022]	S50	[Pratama and Maharani 2021]	S51	[Utami et al. 2021]
S52	[Utami et al. 2022]	S53	[Meshram et al. 2020]	S54	[Bin Tareaf et al. 2019]
S55	[NAJI et al. 2019]	S56	[Al Maruf et al. 2015]	S57	[Zhao et al. 2014]
S58	[Wagner et al. 2013]	S59	[Sumner et al. 2012]	S60	[Quercia et al. 2011]
S61	[Golbeck et al. 2011]	S62	[Kashyap and Nahapetian 2014]	S63	[Huynh et al. 2015]
S64	[Firdaus et al. 2021]	S65	[Karamitsos et al. 2022]	S66	[Alp and Ögüdücü 2016]
S67	[Herzig et al. 2014]	S68	[Agarwal and Mehta 2020]	S69	[Zengin Alp and Şule Gündüz Ögüdücü 2018]
S70	[Das et al. 2018]	S71	[Nargundkar and Rao 2016]	S72	[Wijesekara and Ganegoda 2020]
S73	[Wickramarathna et al. 2020]	S74	[He et al. 2014a]	S75	[He et al. 2014b]
S76	[Yamaguchi et al. 2014]	S77	[Huang et al. 2014]	S78	[Kumar et al. 2014]
S79	[Yamaguchi et al. 2013]	S80	[McGee et al. 2013]	S81	[Li et al. 2012]
S82	[Burger et al. 2011]	S83	[Chen et al. 2016]	S84	[Kotzias et al. 2016]
S85	[Ikeda et al. 2013]	S86	[Suman et al. 2021]	S87	[Zhong et al. 2020]
S88	[Kostakos et al. 2018]	S89	[Yu et al. 2018]	S90	[Ardehaly and Culotta 2017]
S91	[Fernández et al. 2016]	S92	[Efstathiades et al. 2015]	S93	[Harrison et al. 2015]
S94	[Merler et al. 2015]	S95	[Ma et al. 2014]	S96	[Abrol et al. 2012b]
S97	[Abrol et al. 2012a]	S98	[Wandabwa et al. 2021]	S99	[Zarrinkalam et al. 2018]
S100	[Faralli et al. 2017]	S101	[Si et al. 2019]	S102	[Abbasi et al. 2017]
S103	[Hayama 2022]	S104	[Volkova et al. 2016]	S105	[Pathak et al. 2021]