GESTÃO PROATIVA DE PERFIS OFENSIVOS: DETECÇÃO DE TENDÊNCIAS EM ATAQUES CIBERNÉTICOS A INSTITUIÇÕES NO BRASIL POR MEIO DA ANÁLISE DE COMUNIDADES DE HACKERS UTILIZANDO REDES COMPLEXAS E ALGORITMOS DE APRENDIZADO DE MÁQUINA[1]

Proactive management of offensive profiles: detecting trends in cyberattacks on institutions in Brazil through the analysis of hacker communities using complex networks and machine learning algorithms

Claudio Henrique Marques de Oliveira[1], Marcelo Ladeira[2], Flávio de Queiroz Guimarães[3]

Resumo: O Twitter, atual "X", é uma das maiores plataformas digitais para a troca de ideias e informações, mas também atrai hackers com intuito de atividades ilegais e danosas. Este estudo propõe uma abordagem aprimorada para detectar perfis ofensivos associados ao hacktivismo, utilizando redes complexas e algoritmos de aprendizado de máquina, com foco em notificadores da plataforma Zone-H que relatam ações hacktivistas no Brasil. Foram identificados os usuários mais atuantes com base em métricas de rede e palavras-chave, e suas postagens foram analisadas por meio de técnicas de agrupamento (clusterização). A principal contribuição reside na identificação de contas alinhadas ao hacktivismo e na avaliação de seu potencial de ameaça, a fim de prevenir ataques cibernéticos, gerando alertas precisos e oportunos.

Palavras-chave: Hacktivismo, Ciberataques, Redes complexas, Aprendizado de máquina, Detecção de ameaças, Análise de comunidades.

Abstract: "X" (Twitter) has established itself as an influential platform for the exchange of ideas and information, but it also attracts hackers engaged in illegal and harmful activities. This study proposes an enhanced approach to detect offensive profiles associated with hacktivism on "X", utilizing complex networks and machine learning algorithms and focusing on notifiers from the Zone-H platform who report hacktivist actions in Brazil. Key users were identified based on network metrics and keywords, and their posts were analyzed using clustering techniques. The main contribution lies in identifying accounts aligned with hacktivism and assessing their threat potential in order to prevent cyberattacks by generating accurate and timely alerts.

Keywords: Hacktivism. Cyberattacks. Complex networks. Machine learning. Threat detection. Community analysis.

1. Mestre. Departamento de Ciências da Computação da Universidade de Brasília, Brasília, DF – Brasil. E-mails: claudio.oliveira@aluno.unb.br
2. Doutor. Departamento de Ciências da Computação da Universidade de Brasília, Brasília, DF – Brasil. E-mail: mladeira@unb.br
3. Mestre. Instituto de Computação da Universidade Federal Fluminense, Niterói, RJ – Brasil. E-mail: flavioqueiroz2004@gmail.com

[1]Este artigo é uma republicação do trabalho originalmente apresentado no XXI Encontro Nacional de Inteligência Artificial e Computacional (ENIAC 2024). Referência original: OLIVEIRA, C. H. M.; LADEIRA, M.; GUIMARÃES, F. Q. Proactive management of offensive profiles: detecting trends in cyberattacks on institutions in Brazil through the analysis of hacker communities using complex networks and machine learning algorithms. Anais do Encontro Nacional de Inteligência Artificial e Computacional (ENIAC), Porto Alegre, p. 695-706, 2024. DOI: https://doi.org/10.5753/eniac.2024.245063. Publicado sob licença CC BY-NC 4.0.

1. INTRODUÇÃO

Utilizando a construção de uma rede de usuários baseada em notificadores listados no Zone-H, site de monitoramento de atividades hacker (Zone-H 2023), é possível identificar e categorizar usuários com comportamentos típicos de hackers atuantes em redes no Brasil. A pesquisa destaca conexões entre usuários por meio de métricas de rede como centralidade, proximidade e intermediação, revelando atores centrais e influentes na comunidade hacker. Postagens desses atores centrais foram analisadas quanto à positividade ou negatividade e submetidas a técnicas de processamento de linguagem natural (NLP) para pré-processamento das postagens e para aprendizado de máquina.

O estudo inova ao explorar um conjunto de dados brasileiro autêntico de perfis de hackers no "X", com foco em *defacements*, um tipo de ataque cibernético em que um invasor modifica a aparência visual de um site ou página da web. Embora a eficácia dos métodos careça de aprofundamento, espera-se que o modelo identifique hackers com maior precisão e avalie proativamente a gravidade de suas intenções (ZHANG et al., 2022). A proposta de estudo é um sistema aprimorado que alerte sobre ameaças cibernéticas iminentes, contribuindo para a proteção contra invasões cibernéticas (HERNANDEZ et al., 2016).

Este artigo está estruturado da seguinte forma: a Seção 2 revisa a literatura relacionada; a Seção 3 detalha a metodologia; a Seção 4 aborda a coleta e o pré-processamento de dados; a Seção 5 analisa os resultados; e a Seção 6 conclui com as principais descobertas e implicações para pesquisas futuras.

2. TRABALHOS RELACIONADOS

O hacktivismo é um fenômeno cultural e social complexo, caracterizado por princípios como liberdade de informação, desconfiança da autoridade e defesa da descentralização. Apresenta características de mérito e competição, valorizando a demonstração de habilidades técnicas e sociais (HIMANEN, 2001). Desafia as noções tradicionais de política digital, combinando elementos de individualismo com experimentos de coletivismo não hierárquico (COLEMAN, 2014). Alsaffar et al. (2019) avaliam o desempenho de vários algoritmos de aprendizado de máquina e de aprendizado profundo na detecção de spam no "X". Gururaj et al. (2021) propõem abordagem similar com foco específico na detecção de usuários maliciosos em redes sociais por meio de aprendizado de máquina. Benjamin e Chen (2015) utilizam modelos de linguagem baseados em redes neurais recorrentes (RNNLMs; *recurrent neural network language models*) para aprender relações semânticas entre termos usados por hackers, sugerindo que esses modelos podem ser ferramentas promissoras na modelagem da linguagem hacker. Khandpur (2018) propõe uma abordagem para detectar ataques cibernéticos utilizando mídias sociais como fonte de dados, destacando a importância de identificar atividades maliciosas em estágios iniciais e apresentando uma metodologia proativa para identificar ameaças cibernéticas. Le Sceller et al. (2017) introduzem o SOund Navigation And Ranging (SONAR, navegação e determinação de distância por som) para detectar eventos de segurança cibernética em tempo real no "X", enfatizando a importância de monitorar esses eventos de maneira prévia.

A origem das postagens utilizadas não foi abordada em profundidade nesses estudos, que se concentraram em seu conteúdo, sem analisar as características de contas de hackers e usuários comuns. O presente estudo busca preencher essa lacuna ao incorporar uma metodologia para identificar perfis de hackers e atividades correlatas no Brasil, contribuindo para uma compreensão mais abrangente das ameaças cibernéticas divulgadas na plataforma "X" de mídia social no país.

3. MÉTODO

Este estudo adota uma abordagem multifásica para melhorar a detecção de atividades de hackers no "X", selecionando-as conforme informações coletadas do Zone-H e utilizando análise de redes complexas e aprendizado de máquina.

Fase 1 – Coleta inicial dos dados. No repositório Zone-H, são extraídas informações de notificadores e registros de *defacement*, fundamentados na documentação de incidentes correlatos na web, com ênfase no contexto brasileiro. Com base nessas informações, busca-se identificar e registrar, no "X", perfis de usuários associados a atividades ilícitas e perfis ativos na comunidade hacker. Os dados coletados das postagens incluem informações tais como data, texto, nome de usuário, interações e conteúdo multimídia. Eles são processados e armazenados em um banco de dados SQLite. Tal técnica é amplamente estabelecida e documentada na literatura acadêmica (COGBURN & ESPINOZA-VASQUEZ, 2011). Esse método permite a

obtenção de informações públicas disponíveis na plataforma "X", seguindo diretrizes éticas e legais vigentes. Foram coletadas 455.735 postagens no período de 20/10 a 01/12/2023.

Fase 2 – Identificação de atividades de hackers. Técnicas de processamento de linguagem natural (NLP) são usadas para pré-processar o texto contido nas postagens, padronizando o uso de letras minúsculas e maiúsculas, retirando *stopwords* (palavras de parada) e identificando menções de um usuário a outro(s), endereços eletrônicos (URLs) e outros elementos relevantes (MANNING; RAGHAVAN; SCHÜTZE, 2008). As postagens pré-processadas são filtradas utilizando 21 *hashtags*, consideradas as principais relacionadas a *defacement*, resultando em 23.672 postagens. As *hashtags* utilizadas são: *leaked, deface, Anonymous, owned, hack, breach, cyberattack, nofields, hacked, hacking, defacing, leak,* hackeada, *cyberteam, zoneh, BrazilianCyberArmy,* InvasãoEspecial e invasão. Essas postagens são armazenadas em um banco de dados SQLite. O uso de filtros de *hashtags* é uma abordagem estabelecida na literatura (MORSTATTER et al., 2013) e permite focar a coleta em tópicos de interesse específico, proporcionando contexto e compreensão mais profunda das atividades relacionadas a esses tópicos.

Fase 3 – Identificação das comunidades. Com base nas postagens filtradas, realiza-se uma análise de clusterização. O algoritmo de clusterização utilizado é o k-Means. O número de clusters é determinado pelo método do cotovelo, em razão de sua comprovada eficácia na determinação do número de clusters (FORTUNATO, 2010; ROUSSEEUW, 1987). Para cada cluster, são identificados os termos mais frequentes relacionados a *defacement,* e é realizada uma análise de rede complexa com o objetivo de identificar os principais usuários. As postagens dos usuários que mencionam perfis identificados como potenciais atores de ameaças cibernéticas são coletadas se não o tiverem sido. Métricas de rede como centralidade, proximidade e intermediação são aplicadas para identificar e destacar atores de ameaça. A identificação de usuários-chave desempenha papel crucial na análise de mídias sociais e na compreensão da dinâmica das comunidades *online* (WASSERMAN; FAUST, 1994). A etapa de análise de rede envolve a aplicação de ferramenta de análise de redes para calcular métricas, tais como proximidade e intermediação, que fornecem percepções valiosas sobre a estrutura e a influência na rede (NEWMAN, 2010).

Fase 4 – Construção e análise da rede de interações. O cluster mais apropriado (com maior quantidade de usuários e maior frequência de termos relacionados à cibersegurança) é selecionado para análise detalhada, seguido da construção de uma rede de interações que conecta os diversos usuários abrangidos por aquele cluster. Chouchani e Abed (2020) apresentam uma revisão comparativa abrangente das abordagens para agrupamento de atores de redes sociais em comunidades de interesse que ajuda a contextualizar a construção de redes de interações. Destaca-se a relevância da segmentação de comunidades e grupos de interesse em ambientes *online* (CHOUCHANI; ABED, 2020). Bellaby (2021) contribui propondo um *framework* ético para operações de *hacking*. A construção de redes de interações é baseada em menções entre usuários e desempenha papel crucial na revelação de padrões e na compreensão da estrutura das interações dos usuários envolvidos em discussões relacionadas a *hacking* e atividades cibernéticas. A construção da rede de interações requer as seguintes tarefas:

i. Coleta de Dados de Menções. Do banco de dados coletados, extraem-se as menções direcionadas ou recebidas por usuários identificados. Incluem-se postagens que incorporam o símbolo "@" seguido do nome de usuário do indivíduo mencionado;
ii. Construção da Rede de Menções. Com os dados de menções supracitados, procede-se à construção da rede. Nessa rede, os nós representam os usuários do "X", enquanto as arestas representam as menções, estabelecendo uma conexão direcional entre o usuário que fez a menção e o usuário mencionado. A direção é importante para entender quem inicia a comunicação e quem recebe atenção (MAHARANI et al., 2018).
iii. Análise da Rede de Menções. As análises são conduzidas para identificar usuários com alta centralidade de grau, ou seja, aqueles que são frequentemente mencionados e/ou mencionam muitos outros usuários. Essa identificação pode indicar influência ou importância na rede (BARABÁSI, 2016; FREEMAN, 1979). Além disso, técnicas de detecção de comunidades podem ser aplicadas para identificar grupos densamente conectados na rede, que podem representar hackers ou colaboradores (CLAUSET et al., 2004; FORTUNATO, 2010). As representações gráficas ajudam a identificar padrões visuais proeminentes, grupos e usuários de forma intuitiva (HANSEN; SHNEIDERMAN; SMITH, 2011). Essa metodologia, baseada na análise de redes sociais, encontra respaldo na literatura acadêmica, na qual estudos anteriores reconheceram a utilidade da análise de redes para compreender comunidades *online* (KNOKE; YANG, 2008; SCOTT, 2017).

4. IMPLEMENTAÇÃO E AVALIAÇÃO

Essa seção apresenta detalhes sobre como a metodologia foi implementada e testada, a discussão dos desafios enfrentados durante a implementação e como foram superados.

4.1 COLETA DE DADOS DO ZONE-H

Os dados foram coletados do site Zone-H, conhecido por registrar atividades de *defacement*. Os dados extraídos incluíam informações como data e hora, notificador, tipos de *defacement*, domínio, sistema operacional e URL do *defacer*. Esses dados foram armazenados em um *data frame* para análise, conforme ilustrado na Figura 1.

4.2 EXTRAÇÃO DE NOTIFICAÇÕES E BUSCA CORRESPONDENTE NO "X"

Com base nos dados coletados do Zone-H, os notificadores foram extraídos (Figura 2) e usados como referência para buscar postagens ou nomes de usuários relacionados no "X". As postagens que apresentam semelhanças com o nome do notificador são baixadas e armazenadas no banco de dados para filtragem.

	Time	Notifier	H	M	R	L	Special	Domain	OS	View URL
0	23:41:08	Gab	1	0	0	Brazil	0	esthosting.com.br	Linux	/mirror/id/40979686
1	00:23:56	Clash Hackers	0	0	0	Brazil	0	silvam.pompeumg.com.br/cl.html	Linux	/mirror/id/40979416
2	23:02:40	Clash Hackers	0	0	0	Brazil	0	rdev.epimaringa.com.br/cl.html	Linux	/mirror/id/40979370
3	22:07:39	Clash Hackers	0	0	0	Brazil	0	teste.epimaringa.com.br/cl.html	Linux	/mirror/id/40979350
4	09:01:53	Plastyne	0	0	0	Brazil	0	palinialves.com.br/play.txt	Linux	/mirror/id/40974065
...	...	...	...	...	...	...	...	...	...	...
995	2023/05/22	VandaTheGod	1	0	1	Brazil	0	diaseproenca.com.br	Linux	/mirror/id/40609325
996	2023/05/21	B1G0D1N	1	0	0	Brazil	0	tatianacapanema.com.br	Linux	/mirror/id/40608842
997	2023/05/14	diparis	0	0	0	Brazil	0	glpi.cenciseg.com.br/glpi/	Linux	/mirror/id/40591452
998	2023/05/13	Rxc404	1	0	0	Brazil	0	cetri.com.br	Linux	/mirror/id/40591087
999	2023/05/13	Rxc404	1	0	0	Brazil	0	explosaodeleads.com.br	Linux	/mirror/id/40591085

1000 rows × 10 columns

Figura 1. *Data frame* do Zone-H.

	Notifier	Counts
0	VandaTheGod	63
1	ProtoWave Reloaded	40
2	Junin-CLS	32
3	B1G0D1N	30
4	Tux Society	29
...	...	...
195	B0yzTeam	2
196	terezinha security	2
197	rtax	2
198	Finistro	2
199	abeille23	2

200 rows × 2 columns

```
hashtags
#CyberAttack              4214
#cyberattack              3685
#cybersecurity            3552
#Hacked                   3296
#Anonymous                2698
                          ...
#OSINT                     113
#aleistercrowley           113
#CyberSecurityAwareness    113
#cloudsecurity             112
#Israeli                   111
Name: count, Length: 200, dtype: int64
```

Figura 2. Notificadores e número de *defaces* e *hashtags* frequentes.

4.3 COLETA DE DADOS DO "X"

Com a coleta de dados no "X", postagens relevantes foram identificadas e foram extraídos dados como data, texto, nome de usuário, nome exibido, número de comentários, repostagens, curtidas, *links* e outros metadados relevantes.

4.4 FILTRAGEM DOS DADOS DO "X" E COLETA DE *HASHTAGS* COMUNS

Os dados foram filtrados para identificar *hashtags* comuns e palavras-chave associadas ao *hacking* e às atividades cibernéticas. Uma nova coleta de dados foi realizada com base nessas *hashtags*, permitindo expandir o escopo da busca previamente realizada (Figura 3).

4.5 IDENTIFICAÇÃO DE USUÁRIOS RELEVANTES E COLETA DE POSTAGENS

Foi realizada uma análise com os dados coletados para identificar os principais usuários mencionados em relação às 21 *hashtags* predeterminadas. Posteriormente, as postagens desses usuários foram coletadas para análise adicional, particularmente aquelas que apresentaram uma alta incidência de termos relacionados ao *hacking*.

4.6. ANÁLISE DE CLUSTERS

Após o pré-processamento das postagens, os dados foram clusterizados com o algoritmo k-Means para auxiliar na identificação de padrões e agrupamentos (HASTIE, 2009). O número de clusters foi determinado com a aplicação dos métodos do cotovelo (Figura 4) e da silhueta (Figura 5).

O valor do Silhouette Score varia de -1 a 1. Um valor alto indica que o objeto está bem combinado com seu próprio cluster e mal combinado com clusters vizinhos. Se a maioria dos objetos tiver um valor alto, então a configuração dos clusters é apropriada. Se muitos pontos mostrarem um valor baixo ou negativo, a configuração dos clusters pode ter espaço para melhoria. Foram realizados três testes com 3, 4 e 5 clusters, como mostrado na Figura 5, em que o silhouette score para 3 clusters é maior do que para as demais configurações de clusters.

O valor do silhouette score médio mais alto entre essas configurações é para 3 clusters, sugerindo que a configuração com 3 clusters é a mais apropriada. Contudo, foi escolhida a configuração de 4 clusters, pois, em uma análise manual, mostrou melhor distribuição de grupos de hackers. Essa decisão foi baseada na observação qualitativa dos dados, em que a configuração de 4 clusters proporcionou uma separação mais intuitiva e relevante dos diferentes grupos de hackers, permitindo melhor interpretação e análise das atividades desses grupos. A escolha por 4 clusters facilitou a identificação de padrões e comportamentos específicos dentro de cada grupo, o que é crucial para a análise de segurança cibernética.

4.7. ANÁLISE TEMPORAL, IDENTIFICAÇÃO DE TERMOS COMUNS E SENTIMENTO DE CLUSTER

Foi realizada uma análise dos termos das postagens em cada cluster com o objetivo de identificar tendências e padrões temporais

	text	hashtags
0	Join us! #OpNewBlood #Anonymous #ExpectUs	[#OpNewBlood, #Anonymous, #ExpectUs]
1	Ghosts of Palestine targeted major websites of...	[#Cti, #Threatintel, #Israel]
2	Ghosts of Palestine is targeting Rafael's Iron...	[#Threatintel, #Israel]
3	Several #Zionist websites were taken offline b...	[#Zionist, #OpIsrael, #OpIsraelv2, #AlaqsaStorm]
4	Malek Team Hacked Ono Academic College! Hebre...	[]
...	...	...
95	Do you remember when you joined X? I do! #MyXA...	[#MyXAnniversary]
96	Israeli government educational portal has been...	[#TangoDown, #OpIsrahell]
97	Ministry Of Industries: https:///industry./go...	[#MTB]
98	Sri Lanka Government websites dropped. Nation...	[#MTB]
99	Websites belonging to the Israeli government #...	[#TangoDown, #OpIsrahell]

100 rows × 2 columns

Figura 3. Extração de *hashtags* e filtragem com *hashtags* relevantes.

relevantes ao longo do período de estudo. Observam-se flutuações no volume de postagens, o que é essencial para entender o impacto de eventos ou atividades específicos em determinados períodos (LIN 2009; RODRIGUES 2021). Os termos relacionados à cibersegurança mais comuns nos clusters foram *cyber, attack, website, Russian, cybersecurity, ciberattack, Israel, security* e

```
Para n_clusters = 3, o silhouette score médio é: 0.010558376277538322

Para n_clusters = 4, o silhouette score médio é: 0.010416846371843945

Para n_clusters = 7, o silhouette score médio é: 0.0352063713265274
```

Figura 5. Silhouette score.

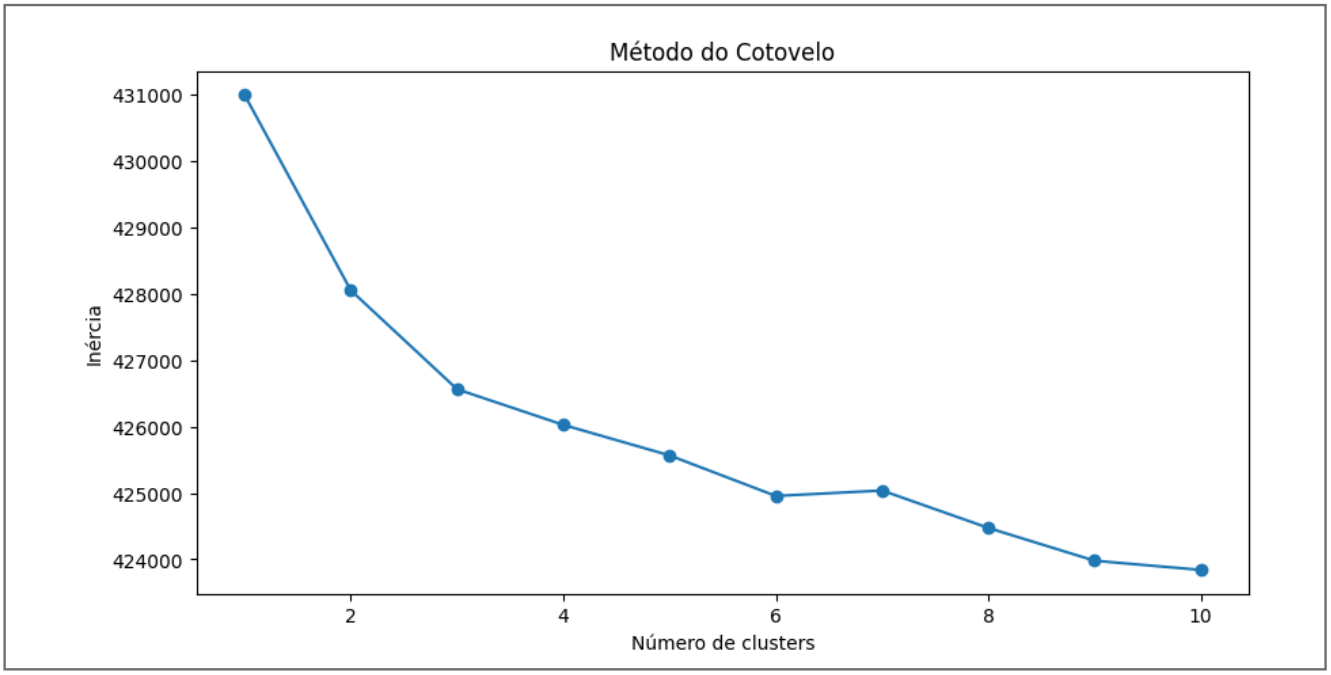

Figura 4. Gráfico de variância intra-cluster em relação ao número de clusters.

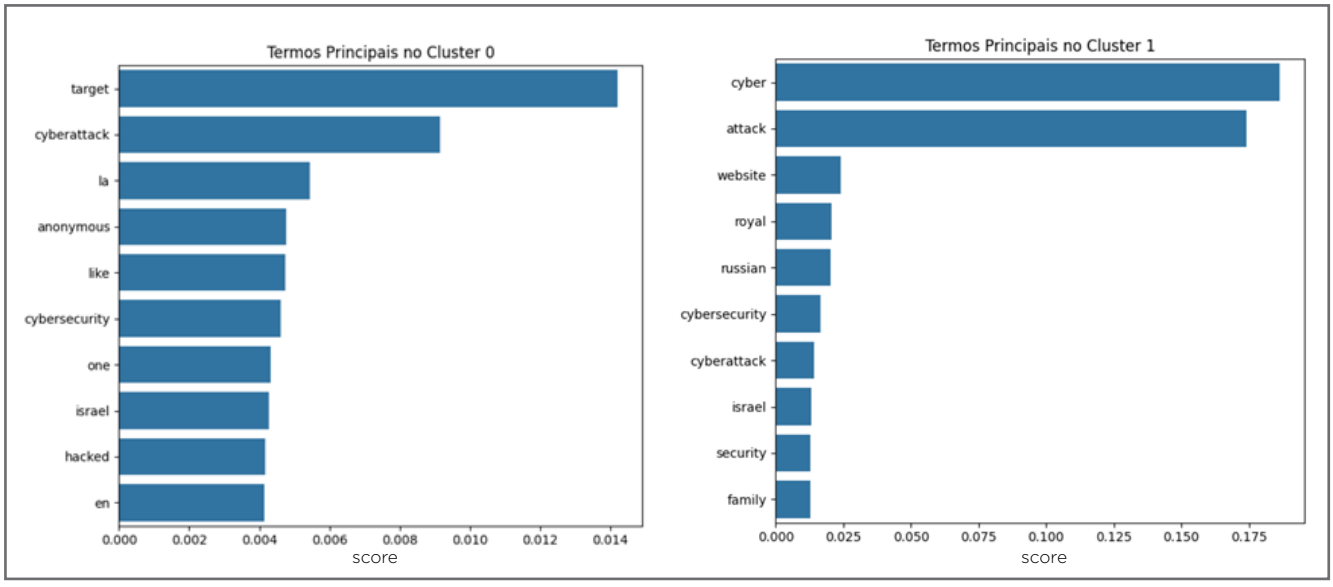

Figura 6. Gráfico de distribuição de postagens dos clusters 0 e 1.

family (Figura 6). Isso sugere uma concentração significativa de discussões sobre questões de cibersegurança e possíveis ataques cibernéticos, refletindo os interesses predominantes na comunidade de hackers e hacktivistas. A análise da distribuição de postagens por cluster revela que certos clusters têm um número significativamente maior de postagens, indicando tópicos de maior interesse durante o período de estudo.

Finalmente, a avaliação do sentimento médio das postagens de cluster revelou nuances sobre o tom emocional das conversas. Por exemplo, os clusters 1 e 0 mostram sentimentos médios de 0.03 e 0.04, respectivamente. Embora esses valores sejam relativamente neutros, é importante notar que qualquer desvio significativo de 0 pode indicar uma tendência emocional nas discussões do cluster. Essa métrica, além de ser uma ferramenta crucial para perceber o ambiente emocional predominante (PANG; LEE, 2008), pode ajudar na categorização e identificação dos temas que geram respostas mais positivas ou negativas na comunidade (BOLLEN et al., 2011 HERNANDEZ-SUAREZ et al., 2018; HERNANDEZ et al., 2016).

4.8 CONSTRUÇÃO E ANÁLISE DAS REDES DE MENÇÕES DO CLUSTER

Nesta etapa do estudo, a análise foca na construção e análise de redes de menções, com ênfase nas interações de usuário para usuário. O processo é crucial para entender a estrutura e os padrões de comunicação entre os participantes nas discussões relacionadas ao hacktivismo e atividades cibernéticas. Para melhor processamento, os usuários mais ativos de cada cluster foram selecionados e os usuários mais mencionados por eles foram extraídos. O objetivo é identificar possíveis canais de disseminação de informações relevantes para esses contextos. A título de exemplo, a Figura 7 apresenta os usuários mais ativos no cluster 1, e a Figura 8 apresenta os usuários mais mencionados pelos usuários mais ativos no cluster 1.

Durante esta análise, foram identificados atores de ameaça ativos que não haviam sido coletados inicialmente (Figura 9). São usuários que tinham sido coletados na Fase 1, mas foram excluídos na Fase 2, ou usuários antigos com alta probabilidade de serem hacktivistas, mencionados em postagens recentes, mas que não postaram durante o período de coleta inicial. Não foram descobertos novos usuários no cluster 3.

Finalmente, são geradas redes de menções com base nas postagens no "X" contidas em cada cluster. A estrutura de cada rede de menções sugere a possibilidade de existência de diferentes subgrupos ou de tópicos sendo discutidos, com alguns indivíduos atuando como pontes entre diferentes

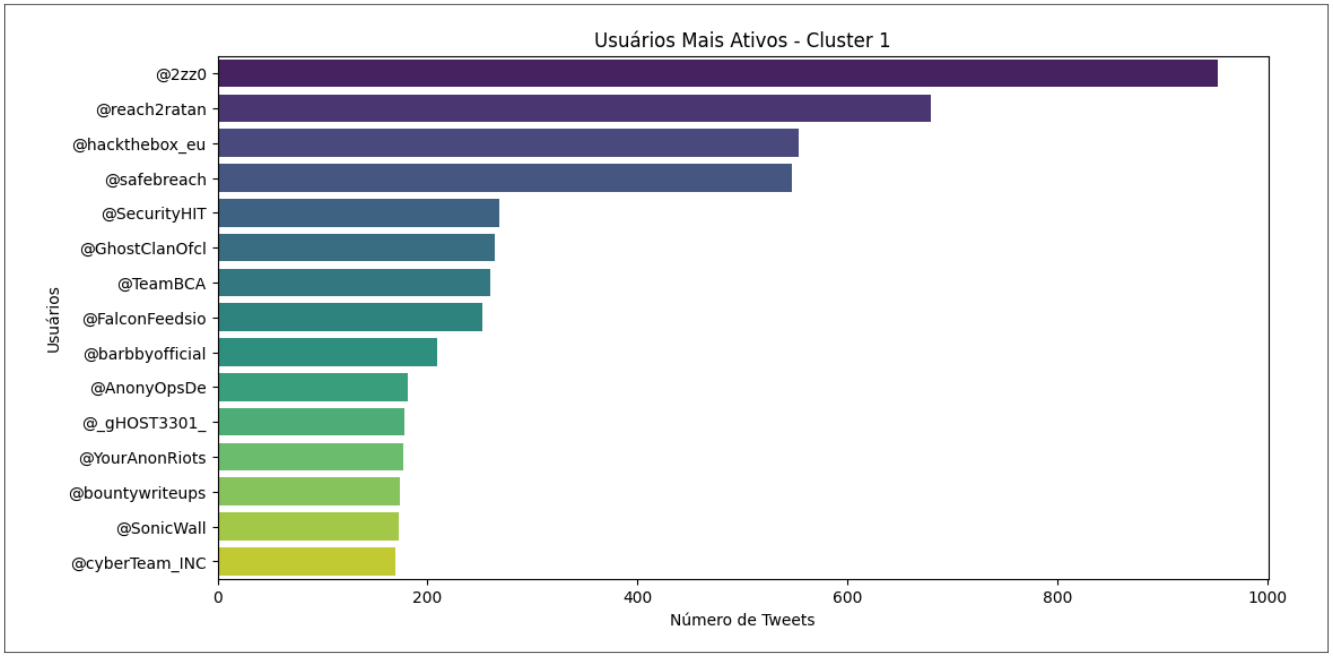

Figura 7. Usuários mais ativos por número de postagens no cluster 1.

subcomunidades. A identificação de atores-chave e a análise de redes são componentes essenciais na análise de ameaças cibernéticas (ROMAGNA, 2020). A importância de identificar atores de ameaça em atividades hacktivistas também é destacada na literatura (BENJAMIN; CHEN, 2012). A Figura 10 apresenta a rede de menções obtida para o cluster 1, sendo o grafo gerado com os dados das postagens que registram as interações entre os usuários. Observa-se uma densidade significativa de conexões, indicando um alto grau de interação entre os usuários, bem como a presença de um núcleo central com vários nós altamente interconectados, sugerindo a existência de indivíduos ou entidades influentes dentro do grupo. Essa estrutura densa pode indicar um grupo de interesse, em que os membros compartilham informações e interagem frequentemente.

```
====================================
Cluster 1:
@BleepinComputer: 47
@MysteriousTeam0: 44
@TheHackersNews: 43
@Land2Cyber: 34
@MichelleRagusa: 29
@_barbby: 25
@OstermanRsch: 24
@MysteriousT34m0: 21
@DarkReading: 20
@aavivi: 19
@AWS: 18
@defcon: 17
@LulzSecSL: 16
@KamikazeJapan5: 16
@YourAnonNews: 16
@Sprek3rsSec: 14
@anonbarbby: 14
@Hornetsecurity: 14
@TweetBrookcourt: 14
@ArmisSecurity: 13
----------------------------------------
Menções em string para o Cluster 1: @reach2ratan, @BleepinComputer,
====================================
```

Figura 8. Usuários mais mencionados pelos usuários mais ativos no cluster 1.

5. CONCLUSÕES

Neste estudo, a intricada rede de hacktivismo e atividades cibernéticas no "X" foi investigada com o emprego de análise de redes complexas e aprendizado de máquina. Isso permitiu uma incursão nas camadas prevalecentes da comunicação *online* entre indivíduos envolvidos em *hacking*, evidenciando uma rede complexa de interações, influências e intenções.

A segmentação dos dados revelada pela análise de clusters, especialmente em nuvens de palavras e na distribuição de sentimentos, não apenas destacou a frequência da linguagem hacker, mas também expôs nuances significativas nas comunicações dentro dessa comunidade. A coexistência de sentimentos tanto negativos quanto positivos sugere uma comunidade heterogênea e multidimensional, a ser pesquisada e detalhada no futuro. Outra evidência foi a visualização da rede de menções, que destacou a vasta e densa rede de indivíduos envolvidos em práticas de *hacking*, com padrões de interconexão que refletem um ecossistema robusto de colaboração e compartilhamento de informações.

```
Cluster 0 - Usuários não existente na coleta inicial (Primeiros 40):
EPSLinares, 0xWORD, CyberHunterSec, 0xWord, rootedcon, DrGiammattei, Singularity_Ex, FundacionINCYDE, hack, gmail
NavajaNegra_AB, eldpit, geeks_academy, zendalibros, ivoox, Gwalrock, inetum_es, AlightSolutions, MiolnirST, WatchGuardSpain
Women4Cyber_SP, 123emprende, AntonioCortesB, _CARITAS, chema_garabito, ssantosv, perezreverte, C1b3rWall, Spreaker, elpais_tec
HazzimIO, MPguatemala, GuatemalaGob, PlexusTech_, oricio_org, fundacionfulgenciomeseguer, martrudix, mikiminoru, TwitchES, c1b3rwall
----------------------------------------
Cluster 1 - Usuários não existente na coleta inicial (Primeiros 40):
Land2Cyber, aavivi, anonbarbby, KamikazeJapan5, cybersaiyanIT, darkstar7471, RoadRunnerHacks, Sprek3rsSec, hacktivistlink, JTSEC13
hpylarinos, _leHACK_, itzikkotler, AWS, zekeriufunet, GITEX_GLOBAL, CertBros, RecordedFuture, Jenny_Radcliffe, Microsoft
TAG_Cyber, securelink, GhostCodin, AnonDragonNeb, MysteriousTeam0, _TheGhostSquad, dilagrafie_, Google, szymex73, 21y4d
mrb3n813, HagueSabastian, snyksec, Infosecurity, TheCyberGeek19, idekCTF, _kavigihan, EuroInformation, KatzcyPlayCyber, Cero_On3
----------------------------------------
Cluster 2 - Usuários não existente na coleta inicial (Primeiros 40):
eric_jeanjean, RollingStones, officialKeef, MysteriousTeamO, AfricaCyberMag, hackyourjob, OVHcloud_FR, ovh_support_fr, icann_fr, ICANN
----------------------------------------
Cluster 3 - Usuários não existente na coleta inicial (Primeiros 40):

----------------------------------------
```

Figura 9. Usuários ativos coletados após análise da rede de menções.

O presente estudo foi eficaz em identificar novos atores de ameaça com os dados analisados e mostrou a importância crucial da coleta contínua de novos usuários potencialmente maliciosos antes que os perfis sejam apagados da rede social. Para trabalhos futuros, é pertinente desenvolver uma rotina de coleta com palavras-chave dinâmicas criadas com base na análise de atores de ameaça já identificados em coleções anteriores, bem como coletar novos usuários detectados como potenciais ameaças na rede social e realizar sua análise.

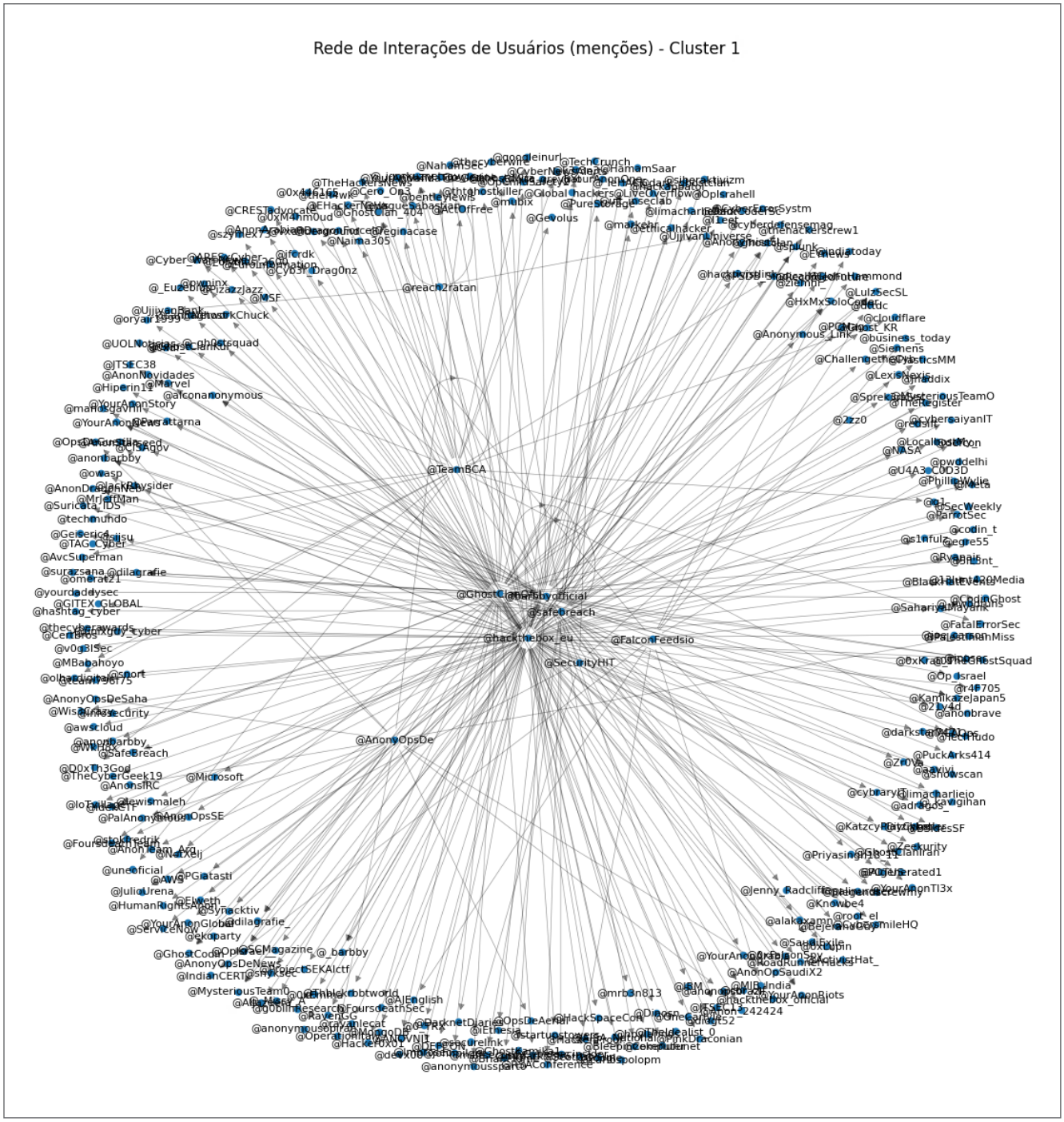

Figura 10. Gráfico da rede de interações entre usuários no cluster 1.

REFERÊNCIAS

ALSAFFAR, D.; ALFAHHAD, A.; ALQHTANI, B.; ALAMRI, L.; ALANSARI, S.; ALQAHTANI, N.; ALBOANEEN, D. A. Machine and deep learning algorithms for Twitter spam detection. *International Conference on Advanced Intelligent Systems and Informatics*, Springer, Cham, p. 483-491, 2019.

BARABÁSI, A.-L. *Network Science*. Cambridge University Press, 2016.

BELLABY, R.W. An Ethical Framework for Hacking Operations. *Ethic Theory Moral Prac*, 24, 231–255 (2021). https://doi.org/10.1007/s10677-021-10166-8

BENJAMIN, V.; CHEN, H. Securing cyberspace: Identifying key actors in hacker communities. *2012 IEEE International Conference on Intelligence and Security Informatics*. Washington, DC, USA, 2012. p. 24-29. https://doi.org/10.1109/ISI.2012.6283296

BENJAMIN, V.; CHEN, H. Developing understanding of hacker language through the use of lexical semantics. *IEEE International Conference on Intelligence and Security Informatics (ISI)*, 2015. p. 79-84.

BOLLEN, J.; MAO, H.; ZENG, X. Twitter mood predicts the stock market. *Journal of Computational Science*, v. 2, n. 1, p. 1-8, 2011. https://doi.org/10.1016/j.jocs.2010.12.007

CHOUCHANI, N.; ABED, M. Online social network analysis: Detection of communities of interest. *Journal of Intelligent Information Systems*, v. 54, n. 1, p. 5-21, 2020. https://doi.org/10.1007/s10844-018-0522-7

COGBURN, D. L.; ESPINOZA-VASQUEZ, F. K. From Networked Nominee to Networked Nation: Examining the Impact of Web 2.0 and Social Media on Political Participation and Civic Engagement in the 2008 Obama Campaign. *Journal of Political Marketing*, v. 10, n. 1-2, p. 189-213, 2011. https://doi.org/10.1080/15377857.2011.540224

COLEMAN, G. *Hacker, Hoaxer, Whistleblower, Spy*: The Many Faces of Anonymous. Verso Books, 2014.

FORTUNATO, S. Community detection in graphs. *Physics Reports*, v. 486, n. 3-5, p. 75-174, 2010. https://doi.org/10.1016/j.physrep.2009.11.002

FREEMAN, L. C. Centrality in social networks: Conceptual clarification. *Social Networks*, v. 1, n. 3, p. 215-239, 1979. https://doi.org/10.1016/0378-8733(78)90021-7

GURURAJ, H. L.; TANUJA, U.; JANHAVI, V.; RAMESH, B. Detecting malicious users in the social networks using machine learning approach. *International Journal of Social Computing and Cyber-Physical Systems*, v. 2, n. 3, p. 229-243, 2021. https://doi.org/10.1504/IJSCCPS.2021.117959

HANSEN, D.; SHNEIDERMAN, B.; SMITH, M. A. *Analyzing Social Media Networks with NodeXL*: Insights from a Connected World. Morgan Kaufmann, 2011.

HASTIE, T.; TIBSHIRANI, R.; FRIEDMAN, J. *The Elements of Statistical Learning*: Data Mining, Inference, and Prediction. Springer, 2009.

HERNANDEZ, A.; SANCHEZ, V.; SANCHEZ, G.; PEREZ, H.; OLIVARES, J.; TOSCANO, K.; MARTINEZ, V. Security attack prediction based on user sentiment analysis of Twitter data. *IEEE International Conference on Industrial Technology (ICIT)*, 2016. p. 610-617.

HERNANDEZ-SUAREZ, A.; SANCHEZ-PEREZ, G.; TOSCANO-MEDINA, K.; MARTINEZ-HERNANDEZ, V.; PEREZ-MEANA, H.; OLIVARES-MERCADO, J.; SANCHEZ, V. Social sentiment sensor in Twitter for predicting cyber-attacks using l1 regularization. *Sensors Journal*, v. 18, n. 5, p. 1-17, 2018. https://doi.org/10.3390/s18051380

HIMANEN, P. *The Hacker Ethic and the Spirit of the Information Age*. Random House, 2001.

KHANDPUR, R. P. *Augmenting Dynamic Query Expansion in Microblog Texts*. 2018.

KNOKE, D.; YANG, S. *Social Network Analysis*. 2nd ed. SAGE Publications, 2008.

LE SCELLER, Q.; KARBAB, E. B.; DEBBABI, M.; IQBAL, F. Sonar: Automatic detection of cyber security events over the Twitter stream. *12th International Conference on Availability, Reliability and Security (ACM)*, 2017. p. 23-34.

LIN, Y.-R.; SUNDARAM, H.; DE CHOUDHURY, M.; KELLIHER, A. Temporal patterns in social media streams: Theme discovery and evolution using joint analysis of content and context. *2009 IEEE International Conference on Multimedia and Expo*, New York, NY, USA, 2009. p. 1456-1459 https://doi.org/10.1109/ICME.2009.5202777

MAHARANI, W.; ADIWIJAYA; GOZALI, A. A. (2015). Degree centrality and eigenvector centrality in twitter. *Proceedings of 2014 8th International Conference on Telecommunication Systems Services and Applications*, *TSSA 2014*. https://doi.org/10.1109/TSSA.2014.7065911.

MANNING, C. D.; RAGHAVAN, P.; SCHÜTZE, H. *Introduction to Information Retrieval*. Cambridge University Press, 2008.

MORSTATTER, F.; PFEFFER, J.; LIU, H.; CARLEY, K. M. Is the sample good enough? Comparing data from Twitter's streaming API with Twitter's firehose. *Seventh International Conference on Weblogs and Social Media (ICWSM 2013)*, 2013. p. 400-408.

NEWMAN, M. E. J. *Networks*: An Introduction. Oxford University Press, 2010.

PANG, B.; LEE, L. Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval*, v. 2, n. 1-2, p. 1-135, 2008. https://doi.org/10.1561/1500000011

RODRIGUES, A. P.; FERNANDES, R.; BHANDARY, A.; SHENOY, A. C.; SHETTY, A.; ANISHA, M. Real-Time Twitter Trend Analysis Using Big Data Analytics and Machine Learning Techniques. *Wireless Communications and Mobile Computing*, 2021. ID 39203252.

ROMAGNA, M. *Hacktivism*: Conceptualization, techniques, and historical view. 2020.

ROUSSEEUW, P. J. Silhouettes: A Graphical Aid to the Interpretation and Validation of Cluster Analysis. *Journal of Computational and Applied Mathematics*, v. 20, p. 53-65, 1987. https://doi.org/10.1016/0377-0427(87)90125-7

SANTA, F.; HENRIQUES, R.; TORRES-SOSPEDRA, J.; PEBESMA, E. A Statistical Approach for Studying the Spatio-Temporal Distribution of Geolocated Tweets in Urban Environments. *Sustainability*, v. 11, n. 3, 595, 2019. https://doi.org/10.3390/su11030595

SCOTT, J. *Social Network Analysis*. SAGE Publications, 2017.

WASSERMAN, S.; FAUST, K. *Social Network Analysis*: Methods and Applications. Cambridge: Cambridge University Press, 1994.

ZHANG, Z.; NING, H.; SHI, F.; FARHA, F.; XU, Y.; XU, J.; ZHANG, F.; CHOO, K-K. R. Artificial intelligence in cyber security: research advances, challenges, and opportunities. *Artif Intell Rev*, v. 55, p. 1029-1053, 2022. https://doi.org/10.1007/s10462-021-09976-0

ZONE-H. *Zone-H – Unrestricted information*. 2023. Disponível em: http://www.zone-h.org/. Acesso em: 20 out. 2023.