O Cenário Atual de Golpes em Redes Sociais: Uma Revisão da Literatura

Anderson Frasão
aacfrasao@inf.ufpr.br
Universidade Federal do Paraná
Curitiba, Paraná, Brasil

Tiago Heinrich
theinric@mpi-inf.mpg.de
Max Planck Institute for Informatics
Saarbrücken, Saarland, Germany

Vinicius Fulber-Garcia
vinicius@inf.ufpr.br
Universidade Federal do Paraná
Curitiba, Paraná, Brasil

Abstract

Social networks were conceived as platforms for interaction, content creation and engagement between users. Over time, browsing these networks has become an integral part of everyday life, culminating in billions of profiles created and accessed regularly on various platforms. However, the ease of access and limited control over content by the maintainers have turned social networks into a favorable environment for fraudulent activities, allowing malicious users to act as scammers in search of illicit benefits. Faced with this scenario, academia and industry have dedicated themselves to analyzing and developing methods to identify, detect, prevent and mitigate scams. However, the diversity of scams, many with unique characteristics, requires specific analysis and solutions for each case. In this context, this paper proposes a systematization of social media scams addressed in the literature between 2014 and 2024, defining categories and subcategories based on their methodological characteristics and objectives. The main result is a taxonomy made up of 5 categories and 19 subcategories.

Keywords

GOLPES, REDE SOCIAL, SEGURANÇA, REVISÃO.

1 Introdução

As redes sociais disponibilizadas por plataformas digitais permitem a interação e o compartilhamento de informações entre usuários, superando barreiras geográficas e promovendo a aproximação entre culturas. Essas redes emergiram como fenômenos da sociedade da informação, crescendo rapidamente graças aos avanços nas tecnologias de comunicação. Exemplos como Facebook, YouTube, TikTok e Instagram destacam-se como redes sociais de alcance global, tornando-se essenciais para a comunicação e o entretenimento em escala mundial. Esse crescimento é também impulsionado pela expansão do acesso tecnológico, especialmente por meio de dispositivos móveis, que consolidaram essas redes como elementos centrais da experiência online [1].

A integração das redes sociais à vida cotidiana é reforçada pelo fato de que elas atendem a necessidades humanas básicas, como comunicação, socialização e expressão. Essas plataformas digitais não apenas conectam pessoas, mas também oferecem meios para criação e compartilhamento de conteúdo, permitindo que os usuários expressem suas opiniões, interesses e emoções. Além disso, as redes sociais têm se consolidado como espaços para a construção de identidades e o engajamento cívico, aproximando os indivíduos de líderes e debates políticos [1].

O ambiente proporcionado pelas redes sociais é altamente favorável à geração de conteúdo direcionado, adaptado aos interesses dos usuários com base em fatores como relações entre seguidores/amigos, conteúdo compartilhado e reações. Nesse contexto, alguns usuários tornam-se influenciadores naturais, desempenhando um papel ativo na disseminação de informações amplamente recebidas por outros usuários e fomentando mecanismos de marketing viral. Esse fenômeno ressalta a importância de compreender a rede de relacionamentos e os padrões de acesso às redes sociais para implementar estratégias eficazes de gerenciamento e distribuição de conteúdo [2].

Nesse contexto, uma das preocupações mais recorrentes está relacionada à privacidade dos usuários, devido à natureza de funcionamento e ao modelo de negócios adotado pelas plataformas de redes sociais. Além disso, essas redes criam um ambiente propício para a realização de diversas atividades fraudulentas, em que criminosos exploram uma variedade de estratégias para obter ganhos ilícitos. Entre as práticas potencialmente empregadas por usuários maliciosos estão a criação de perfis falsos, clonagem de contas e roubo de identidade. Tais atividades podem resultar em golpes que envolvem, por exemplo, o vazamento de informações pessoais, posteriormente usadas para chantagem das vítimas ou para a realização de transações financeiras e negociais fraudulentas. Assim, os prejuízos causados aos usuários não são apenas materiais, mas também emocionais. Embora as plataformas de redes sociais adotem medidas para mitigar essas atividades fraudulentas, ainda enfrentam dificuldades em proporcionar a rápida detecção e prevenção desses problemas [3].

A detecção de golpes em plataformas de redes sociais é uma tarefa desafiadora. Um dos fatores que tornam esse processo particularmente complexo são as diversas técnicas exploradas por criminosos para executar seus ataques. Compreender essas estratégias de maneira holística é ainda mais difícil devido à rapidez com que novas metodologias de golpes surgem e se disseminam, exigindo atenção constante e ferramentas automáticas e adaptativas para detectar padrões associados a atividades fraudulentas [3]. Esse processo inclui o processamento de textos, imagens e vídeos, com a necessidade de diferenciar de forma eficaz atividades legítimas de possíveis fraudes.

Na literatura recente, as discussões sobre golpes estabelecidos em redes sociais oferecem uma visão abrangente das estratégias e métodos fraudulentos utilizados para enganar usuários. Por exemplo, Gupta et al. [4] analisam como campanhas de spam utilizam números de telefone para ampliar o alcance de ataques, criando um canal direto para fraudes. Em Nghiem et al. [5], o foco está nas fraudes de manipulação de mercado, em que sinais sociais e financeiros indicam estratégias fraudulentas de valorização artificial de criptomoedas, conhecidas como pump-and-dump. Já Janetzko et al. [6] destacam o envolvimento de bots em golpes do tipo rug pull, que promovem projetos cripto-fraudulentos seguidos de uma rápida

retirada de fundos. Por sua vez, Chergarova et al. [7] investigam o uso de perfis falsos para persuadir vítimas a investir em criptomoedas. Outros golpes destacados incluem os descritos por Cui et al. [8], que exploram o uso de sentimentos e informações multimodais em notícias falsas, um recurso frequentemente empregado para manipular as emoções dos usuários.

De forma geral, a análise de redes sociais apresenta-se como uma alternativa eficaz para investigar e monitorar atividades suspeitas em plataformas digitais, contribuindo para a identificação e prevenção de golpes online [9]. Além disso, essas análises possibilitam uma compreensão mais aprofundada dos variados mecanismos envolvidos em atividades fraudulentas, permitindo o projeto de soluções de detecção e mitigação de golpes.

Dada a diversidade de golpes existentes em redes sociais, aliada à heterogeneidade metodológica e operacional envolvida em sua execução, torna-se essencial a construção de um arcabouço teórico que organize e sistematize informações, servindo de base para o desenvolvimento de novas propostas na área. Nesse cenário, as contribuições deste artigo são: **(i)** a identificação de golpes praticados em redes sociais; e **(ii)** a apresentação sistemática de estudos realizados em 10 anos de pesquisa, compreendendo o período de 2014 até 2024, que analisam os golpes identificados por meio de uma taxonomia com 5 categorias e 19 sub-categorias.

O restante deste artigo está organizado da seguinte forma: A Seção 2 apresenta aspectos motivadores para a pesquisa relacionada a golpes em redes sociais. A Seção 3 apresenta a metodologia de varredura da literatura, além de sistematizar e discutir os resultados da revisão, destacando os golpes encontrados através de uma estrutura taxonômica. Por fim, a Seção 4 traz as considerações finais e expectativas futuras em relação ao trabalho.

2 Golpes em Redes Sociais: Traços e Consequências

Os golpes online têm se tornado cada vez mais prevalentes, especialmente com o crescimento das grandes plataformas de mídia, como o YouTube, TikTok, WhatsApp e Instagram. Essas redes sociais são exploradas de diferentes maneiras com o instituindo de gerar ganhos pessoais a usuários maliciosos de forma ilícita, a custo de perdas financeiras e danos emocionais causados a usuários legítimos.

Entre os golpes realizados nessas plataformas, destacam-se aqueles que utilizam os campos de comentários para atrair vítimas com promessas de prêmios e oportunidades de investimentos. Um estudo sistemático e em larga escala coletou dados de 8,8 milhões de comentários em 20 canais do YouTube ao longo de seis meses, identificando 206 mil comentários fraudulentos de 10 mil contas [10]. A análise dos dados revelou as campanhas fraudulentas, a dinâmica dos comentários e as técnicas de evasão empregadas pelos golpistas. Além disso, a interação com 50 golpistas forneceu detalhes sobre suas táticas de engenharia social e preferências de pagamento. Com isso, foi possível analisar transações em *blockchains* públicas, revelando roubos de milhões de dólares.

Em outra ocorrência, na Austrália, a vitimização por fraudes online afetou mais de 1,2 milhão de cidadãos entre 2010 e 2011, resultando em perdas de aproximadamente US$ 1,4 bilhão [11]. Mais da metade das vítimas foi contatada por diferentes canais na internet ou por e-mail. Além das perdas financeiras, as vítimas enfrentaram sérios problemas psicológicos, emocionais, sociais e até físicos. Os desafios para responder à vitimização por fraudes online incluem a necessidade de serviços de apoio específicos para ajudar as vítimas a lidar com as consequências desse crime. Isso também é um problema nas Filipinas e em outros países do sudeste asiático, onde o Grupo de Combate ao Crime Cibernético da Polícia Nacional das Filipinas (PNP-ACG) relatou um crescimento expressivo nas queixas de fraudes online, passando de números de dois dígitos em 2013 para três dígitos em 2017 [12].

Além disso, com o crescimento das compras online, consumidores na Malásia enfrentam golpes em que vendedores falsos atraem compradores por meio de diferentes canais, incluindo redes sociais, para pagar por produtos que nunca são entregues ou que diferem do anunciado [13] – evidenciando a necessidade de identificação das ofertas fraudulentas. Além disso, plataformas de anúncios classificados, como o *Craigslist*, são vulneráveis a fraudes, com golpistas se aproveitando da ausência de encontros presenciais entre compradores e vendedores [14]. Golpistas utilizam números de telefone VoIP para ocultar suas identidades e imagens copiadas de outros sites para anunciar produtos inexistentes. Além disso, certas palavras-chave comuns em golpes são empregadas para enganar os usuários. Embora existam orientações de prevenção, elas não apresentam taxas de sucesso absolutas devido à variedade dos golpes aplicados.

Nesse contexto, um estudo recente indica que mensagens com instruções demasiadamente detalhadas para a prevenção de golpes são ineficazes [15]. As vítimas relataram dificuldades em aplicar o conhecimento sobre fraudes às suas experiências reais. A categorização e sistematização das atividades fraudulentas é irrelevante para os criminosos, que buscam obter ganhos financeiros por qualquer meio possível. A recomendação, então, é simplificar as mensagens de prevenção para o público geral, focando em evitar ações de alto nível que são objetivos comuns a maioria dos golpes, como a transferência de valores e o vazamento de dados pessoais.

O mercado de criptomoedas, que emergiu como uma plataforma financeira alternativa na última década, ainda enfrenta a falta de regulamentações legais adequadas, tornando-se um atrativo para golpistas que exploram essas brechas para obter lucros ilícitos [5]. Análises de dados públicos e de *blockchain* revelaram que entidades associadas a sites fraudulentos e contas de mídia social enganosas atuam de forma coordenada para roubar criptomoedas [16]. Esses grupos utilizam atividades fabricadas no *blockchain* para aparentar legitimidade, além de diversificarem os métodos para lavar os lucros.

Dado o potencial prejudicial e a recorrência de golpes online, especialmente aqueles realizados em redes sociais, torna-se evidente que esses golpes configuram uma verdadeira pandemia virtual. É fundamental direcionar esforços para uma compreensão aprofundada desse fenômeno, de modo a viabilizar o desenvolvimento de soluções eficazes para prevenir, detectar e mitigar essas atividades fraudulentas.

3 Uma Análise Abrangente dos Golpes em Redes Sociais

Nesta seção, a metodologia, os resultados e as discussões da revisão da literatura sobre golpes em redes sociais são apresentados. Os

trabalhos analisados foram classificados de acordo com as características dos golpes considerados, sendo organizados em categorias e subcategorias dispostas em uma estrutura taxonômica.

Particularmente, a Subseção 3.1 apresenta a metodologia empregada para realizar a busca e seleção de trabalhos relevantes na literatura. Já a Subseção 3.2 define os golpes identificados durante o processo de revisão, analisando-os com base nos métodos, técnicas e alvos destacados em pesquisas recentes. O objetivo dessa análise é oferecer uma compreensão abrangente das categorias de golpes que afetam usuários legítimos, destacando seus métodos e abordagens mais comuns. Por fim, a Subseção 3.3 faz uma discussão crítica dos resultados obtidos.

3.1 Metodologia

Na literatura, existem discussões voltadas aos mais diversos tópicos relacionados a golpes em redes sociais. Levando em consideração a existência desses trabalhos, uma revisão sistemática foi realizada. O objetivo da revisão é analisar a literatura existente sobre detecção de fraudes em redes sociais, focando em identificar as principais técnicas, limitações e desafios relacionados a golpes em redes sociais. Para alcançar esse objetivo, a revisão foi estruturada em quatro etapas:

(1) Definição dos critérios de seleção dos trabalhos;
(2) Busca, análise e seleção dos trabalhos;
(3) Extração dos dados relevantes;
(4) Tratamento e correlação dos dados.

Os critérios de inclusão e exclusão foram definidos para considerar publicações realizadas entre 2014 e 2024, abrangendo artigos de periódicos e conferências que tratam de golpes em redes sociais de forma direta ou indireta. A busca foi conduzida nas bases IEEE Xplore, Springer, ACM Digital Library, ScienceDirect, Taylor & Francis, Sage Journals, Google Scholar, utilizando termos como "social media fraud", "social media fake news", "sock puppets", "cryptocurrency fraud", "fraud detection in social networks" e "social media scams".

A extração de dados incluiu informações como título, categorização e subcategorização do artigo, ano de publicação, plataforma de publicação e tipos de golpes abordados. Esses dados foram organizados em planilhas para facilitar a análise comparativa. Durante o processo, foram identificadas limitações, como a ausência de informações detalhadas em alguns artigos e a restrição a determinadas bases de dados, o que pode ter resultado na exclusão de estudos potencialmente relevantes.

3.2 Resultados da Revisão

A revisão da literatura, conduzida conforme a metodologia descrita na Seção 3.1, resultou na seleção de 23 artigos relevantes. Nesses trabalhos, foram identificadas 5 categorias e 19 subcategorias de golpes executados em redes sociais. A Tabela 1 apresenta uma visão geral das principais categorias, suas subcategorias, os objetivos dos ataques e as referências correspondentes na literatura. Além disso, para facilitar a compreensão e oferecer uma visão abrangente das categorias e subcategorias identificadas, a Figura 1 apresenta uma ilustração em formato de árvore taxonômica das mesmas. A seguir, as subcategorias consideradas de golpes em redes sociais são detalhadas.

Advanced-fee scam (golpe de taxa antecipada) é uma fraude na qual os golpistas induzem as vítimas a pagar antecipadamente por bens, serviços ou oportunidades que não existem. No contexto de criptomoedas, esses golpes frequentemente envolvem promessas de investimentos lucrativos ou de *tokens* a preços reduzidos. Phillips and Wilder [16] analisam como golpistas operam em páginas visualmente semelhantes, mas aparentemente desconexas, promovidas por contas maliciosas em redes sociais. Essas páginas enganam as vítimas para que enviem criptomoedas com a promessa de retornos elevados; no entanto, os golpistas simplesmente embolsam os fundos recebidos. Gupta et al. [4] destacam que campanhas de *spam* frequentemente utilizam números de telefone para enganar usuários em redes sociais. Essas campanhas podem estar associadas a golpes de taxa antecipada, nos quais os golpistas usam números de telefone para estabelecer confiança e persuadir as vítimas a realizar pagamentos antecipados.

Pump-and-Dump (bombardeamento e despejo) é um esquema de manipulação de mercado que ocorre tanto em mercados tradicionais quanto no de criptomoedas. Nesse esquema, os organizadores adquirem uma grande quantidade de um ativo de baixo valor e, em seguida, o promovem intensamente para inflacionar artificialmente seu preço (fase de *pump*). Uma vez que o preço atinge um nível desejado, esses indivíduos vendem suas participações com lucro (fase de *dump*), causando uma queda rápida no valor do ativo e deixando outros investidores com prejuízo. Hamrick et al. [17] analisam a prevalência de esquemas de *pump-and-dump* no mercado de criptomoedas, especialmente em plataformas como Discord e Telegram, identificando milhares de sinais de *pump* e destacando que moedas com menor capitalização de mercado são mais suscetíveis a essa manipulação. Nghiem et al. [5] investigam como as mídias sociais, particularmente o Twitter, são utilizadas para coordenar manipulações de mercado em criptomoedas. Já em Mirtaheri et al. [18], é proposta uma abordagem para identificar fraudes de *pump-and-dump* combinando dados de mercado e sinais sociais, demonstrando que é possível detectar essas fraudes em tempo real, permitindo intervenções mais eficazes.

Rug Pull (puxão de tapete) ocorre quando os criadores de um projeto (geralmente de *token* ou moedas digitais) promovem massivamente sua nova iniciativa para atrair investidores e inflar artificialmente o valor. Quando o projeto atinge uma valorização significativa, esses criadores retiram abruptamente todo o valor dos investidores ao venderem suas participações ou encerrar o projeto, fazendo com que o valor do ativo caia drasticamente. Janetzko et al. [6] investigam o papel de *bots* em golpes de *promote-hit-and-run* (promoção e fuga), com foco em *rug pulls*. Nesse contexto, foi realizada uma análise de *bots* utilizados para manipular o interesse e o valor de criptomoedas. Utilizando dados do Twitter para observar a atuação de *bots* em 27 casos específicos, o estudo revela que esses *bots* são empregados para gerar uma percepção falsa de legitimidade e volume, atraindo investidores antes da retirada repentina dos fundos pelos golpistas.

Fake News (notícias falsas) referem-se a informações intencionalmente fabricadas, distorcidas ou enganosas, disseminadas como se fossem fatos, geralmente visando manipular a opinião pública, confundir ou influenciar decisões e comportamentos de grandes grupos de pessoas. Em sua essência, *Fake News* difere da desinformação não intencional, pois é criada e promovida com o propósito

Categoria	Sub-Categoria	Objetivo	Referência
Cryptocurency e Golpe de mercado	Advanced-fee Scam	Convencer a vítima a pagar uma "taxa antecipada"para receber um prêmio, herança ou outra recompensa, que na realidade não existe.	[4, 16]
	Pump-and-Dump	Inflar artificialmente o valor de ações ou criptomoedas para vendê-las em alta, deixando os novos investidores com prejuízo quando o preço despenca.	[5, 17, 18]
	Rug Pulls	Criar um projeto financeiro, geralmente uma criptomoeda, e, após atrair investidores, abandonar o projeto e ficar com o capital investido.	[6]
Enganação digital	Fake News	Manipular a percepção pública com informações falsas, influenciando opiniões e comportamentos em benefício próprio ou de terceiros.	[8]
	Fake Shopping	Atrair pessoas para comprar produtos falsos ou inexistentes em plataformas de e-commerce, levando-as a perder dinheiro.	[13]
	Giveaway Scam	Atração com promessas de prêmios ou sorteios falsos que, para serem reivindicado, exigem dados pessoais ou uma taxa, visando roubo de informações ou dinheiro.	[7, 10, 19, 20]
	Phishing	Coletar informações pessoais (como senhas, números de cartão) por meio de mensagens enganosas que imitam comunicações legítimas.	[16, 21, 22]
	Tech Support Scam	Convencer a vítima de que seu dispositivo possui problemas de segurança, oferecendo "suporte técnico"falso para roubar dados ou cobrar por serviços desnecessários.	[4, 23, 24]
Manipulação Emocional	Gaslighting	Manipular a vítima para que duvide da própria percepção ou sanidade, controlando seu comportamento e gerando dependência emocional.	[9]
	LoveGuru	Explorar vítimas emocionalmente vulneráveis, muitas vezes cobrando por "consultoria amorosa"ou "conselhos de relacionamento", geralmente com promessas de resolver problemas sentimentais.	[4]
	Romance Baiting	Conquistar emocionalmente a vítima com a intenção de extrair dinheiro ou informações confidenciais, muitas vezes sob o pretexto de uma relação romântica.	[25, 26]
	Social Engineering	Obter informações confidenciais, manipular ou persuadir pessoas a realizarem ações específicas por meio da exploração de emoções e relações pessoais.	[27]
Perfis Falsos	Catfishing	Enganar alguém ao criar um perfil falso, geralmente para conquistar a confiança da vítima e conseguir favores emocionais, financeiros ou informações sensíveis.	[9]
	Honey Trap	Induzir a vítima a revelar informações confidenciais ou comprometer-se em uma situação embaraçosa, geralmente usando a atração romântica ou sexual como isca.	[9]
	Identity Theft	Obter e explorar informações pessoais de forma fraudulenta para ganhos financeiros, manipulação social ou outros usos ilícitos.	[28]
Spam	HashJacker	Explorar hashtags populares para promover golpes, geralmente redirecionando para sites de phishing ou de vendas fraudulentas.	[29]
	Scam by Bots	Automatizar interações online para influenciar opiniões, inflar seguidores, manipular discussões ou até enganar pessoas em transações.	[9, 30]
	Social Spamming	Divulgar conteúdo em massa nas redes sociais com links de phishing ou publicidade enganosa, visando conversões ou coleta de dados.	[9, 31, 32]
	Sockpuppets Scam	Criar contas falsas para manipular discussões, promover desinformação ou dar apoio falso em debates, geralmente para influenciar opiniões.	[9]

Tabela 1: Categorização de Golpes em Redes Sociais

de enganar, normalmente por razões políticas, econômicas ou sociais. Cui et al. [8] abordam a detecção de notícias falsas nas mídias sociais, enfatizando a importância de considerar os comentários dos usuários e os sentimentos presentes, propondo um *framework* para detecção de *Fake News* que analisa não só o conteúdo da notícia, mas também os comentários e sentimentos dos usuários em relação a ela.

Fake Shopping (loja falsa) refere-se a uma forma de golpe no qual uma loja ou perfil/página de compra simula ser legítima para atrair consumidores e realizar transações, mas, na realidade, é falsa ou enganosa. Essas lojas falsas online podem exibir produtos atrativos e ofertar preços tentadores para convencer o consumidor a realizar uma compra, mas não realizam a entrega do produto ou entregam um produto de qualidade inferior ao anunciado. Mokhsin et al. [13] investigam os fatores que influenciam o comportamento dos consumidores em relação às compras online, com um foco específico em sua vulnerabilidade a golpes e fraudes digitais. Analisando dados de 201 consumidores, variáveis como confiança em plataformas de redes sociais, experiência de compra e conhecimento sobre fraudes que afetam o consumidor são averiguadas.

Giveaway Scam (golpe de sorteio) é um tipo de golpe que envolve falsos sorteios ou promoções, muitas vezes com promessas de prêmios em criptomoedas, como Bitcoin ou Ethereum, para atrair vítimas e enviar dinheiro ou informações pessoais aos golpistas. Esse tipo de golpe é recorrente em redes sociais, como o YouTube, Twitter, Instagram e Facebook, onde os criminosos aproveitam a popularidade e a credibilidade de figuras públicas ou marcas conhecidas para enganar os usuários. Vakilinia [19] e Li et al. [10] exploram a execução de golpes em plataformas de mídia como o YouTube e outras redes sociais, onde golpistas utilizam transmissões ao vivo e comentários falsos para enganar os usuários, promovendo

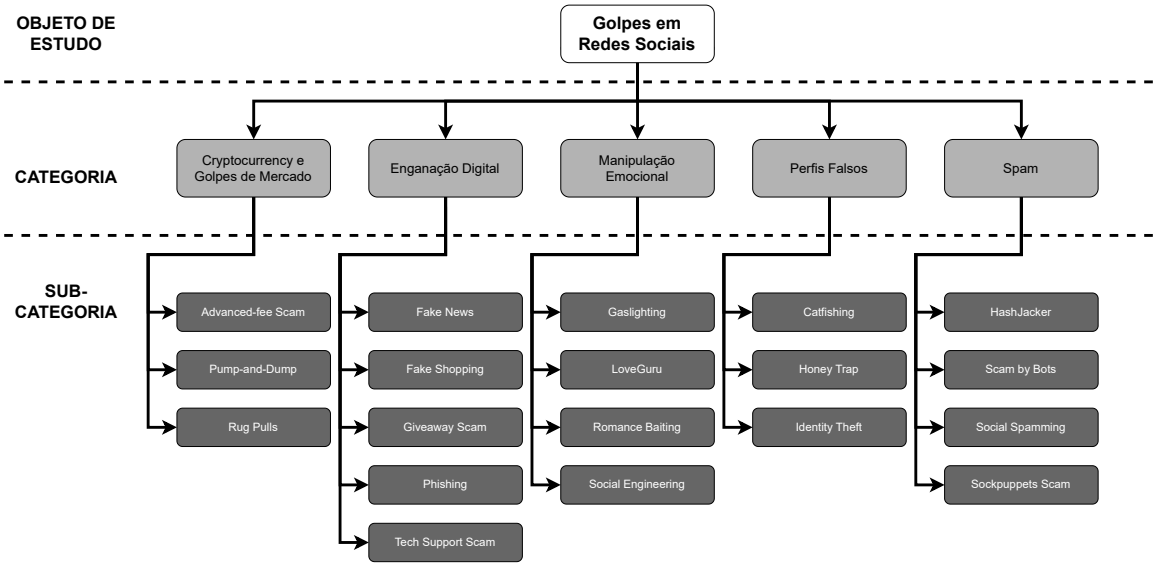

Figura 1: Árvore Taxonômica dos Golpes em Redes Sociais

falsos sorteios de criptomoedas. Os estudos analisam os métodos de criação de engajamento falso e as táticas usadas para aumentar a credibilidade do golpe. Liu et al. [20] realizam uma análise de ponta a ponta dos golpes de sorteio, examinando desde as técnicas de atração de vítimas até a medição das taxas de conversão e o lucro gerado pelos golpistas. Usando dados de diversas plataformas e *blockchain*, o estudo identifica o valor total perdido pelas vítimas e os principais pontos de vulnerabilidade nas plataformas digitais. Chergarova et al. [7] investigam a criação de perfis falsos para induzir vítimas a investir em criptomoedas. A pesquisa examina o perfil dos golpistas e a estrutura de engenharia social utilizada para enganar indivíduos.

Phishing (roubo de dados) é uma técnica de golpe que utiliza engenharia social para enganar vítimas e obter informações confidenciais, como senhas e dados bancários, por meio de mensagens ou sites que imitam entidades confiáveis. Os ataques podem ocorrer via e-mail, redes sociais ou até aplicativos, explorando a psicologia humana, como medo ou urgência, para induzir a vítima a clicar em *links* maliciosos e fornecer informações pessoais. Ataques comuns incluem *phishing* por e-mail, *spear phishing* (direcionado) e *whale phishing* (alvos de alto perfil). Ding et al. [21] propõem um método que combina busca em mecanismos de pesquisa, regras heurísticas e regressão logística para identificar páginas de *phishing*, aumentando a eficiência na filtragem de páginas legítimas e a eficácia na detecção de fraudes. Frauenstein and Flowerday [22] exploram como o comportamento de usuários em redes sociais, acostumados a cliques frequentes, os torna mais vulneráveis a ataques de *phishing*, destacando a importância de entender a psicologia do usuário para mitigar esses riscos. Phillips and Wilder [16] buscam identificar padrões em sites de golpes com criptomoedas via análise de dados de *blockchains* para rastrear o fluxo de fundos e identificar campanhas coordenadas de *phishing*.

Tech Support Scam (golpe de suporte técnico) é um golpe em que criminosos se passam por representantes de empresas de tecnologia para enganar usuários. Geralmente, utilizam redes sociais e anúncios pagos para disseminar mensagens alarmistas com números de telefone [23]. Assim, as vítimas são levadas a acreditar que seus dispositivos estão comprometidos, ligam para os golpistas e, durante a interação, são persuadidas a pagar por serviços falsos, fornecer informações pessoais ou instalar software malicioso. Além disso, Rauti and Leppänen [24] mostram que, durante a interação, é comum que os golpistas implantem ou explorem ferramentas que permitem acesso remoto aos dispositivos, ampliando os danos potenciais. Gupta et al. [4] analisam como as campanhas de ataque utilizam redes sociais para publicar mensagens alarmistas com números de telefone, visando enganar usuários. Foi identificado padrão de abuso, destacando a persistência e repetição desses números em várias plataformas. O estudo demonstra que, ao correlacionar dados entre redes sociais, é possível detectar e mitigar essas campanhas de forma eficaz, reduzindo o impacto financeiro e a disseminação do golpe.

Gaslighting (manipulação psicológica) é um golpe que busca desestabilizar a vítima, fazendo-a questionar sua memória, percepção e julgamento, enquanto cria dependência emocional no manipulador. Por meio da distorção da realidade, negação de eventos e rebaixamento das percepções da vítima, o manipulador semeia dúvidas e confusão, gerando impactos como ansiedade, depressão e isolamento social. Rawat et al. [9] investigam como o *gaslighting* é explorado como uma tática psicológica utilizada em ambientes online para desorientar e manipular indivíduos ou grupos, facilitando ações de manipulação social e controle psicológico em redes sociais.

LoveGuru (golpe do guru do amor) consiste em campanhas de *spam* que promovem serviços de concelhos românticos em redes sociais, utilizando números de telefone para atrair vitimas. As mensagens, repetitivas e distribuídas em grande escala por *bots* ou

perfis comprometidos, são postadas em plataformas como Facebook e Twitter. Prometendo soluções para problemas amorosos, a campanha pode levar a prejuízos financeiros e fraudes. Gupta et al. [4] analisam campanhas com astrólogos prometendo soluções para relacionamentos, usando os casos para ilustrar os métodos e impactos de *spam* baseado em números de telefone, bem como as limitações das redes sociais na detecção e mitigação dessas atividades, apontando para a necessidade de melhores medidas e colaboração entre plataformas.

Romance Baiting (golpes de romance) é uma estratégia de golpe online em que criminosos criam perfis falsos para estabelecer relacionamentos afetivos com vítimas em plataformas digitais. O golpista constrói uma conexão emocional profunda usando narrativas persuasivas, como emergências fictícias, para manipular sentimentos e obter vantagens financeiras. A prática envolve engajamento prolongado, uso de linguagem afetiva e táticas psicológicas que criam a ilusão de autenticidade. As vítimas, confiando na relação, acabam transferindo dinheiro, sofrendo prejuízos financeiros e emocionais. Anesa [25] analisam como o *romance baiting* opera por meio de estratégias linguísticas e psicológicas usadas por golpistas para manipular emocionalmente suas vítimas em golpes românticos online. A pesquisa explora exemplos reais de interação, identificando táticas de persuasão, construção de confiança e apelo emocional. O trabalho também investiga como essas práticas afetam as vítimas e sugere a importância de entender esses métodos para desenvolver iniciativas preventivas e educacionais contra esse tipo de golpe. Cross [26] exploram como as fraudes de relacionamento evoluíram, integrando esquemas de investimentos fraudulentos, particularmente em criptomoedas. Assim, os golpistas usam a fachada de um relacionamento romântico para enganar as vítimas, incentivando-as a investir em oportunidades de criptomoedas falsas. O estudo analisa as razões subjacentes ao sucesso dessa combinação de fraudes e como isso tem distorcido o conceito de fraude romântica tradicional.

Social Engineering (engenharia social) é o uso de manipulação psicológica para enganar pessoas e obter informações confidenciais, acesso não autorizado ou levá-las a realizar ações específicas. Esses golpes exploram vulnerabilidades humanas, como falta de atenção ou confiança excessiva. Yu et al. [27] analisam como a inteligência artificial está revolucionando os ataques de engenharia social, tornando-os mais eficazes e perigosos. Além disso, os autores categorizam as técnicas de ataques, avaliam os riscos crescentes associados a essas práticas e propõem estratégias para mitigar essas ameaças, como detecção automatizada, educação do usuário e autenticação robusta.

Catfishing (golpe de identidade falsa) é uma prática fraudulenta em que indivíduos criam perfis falsos em redes sociais para enganar ou manipular vítimas, com objetivos que podem variar entre ganhos financeiros, comprometer a vítima emocionalmente, incomodar ou realizar fraudes. Usando perfis falsos com informações enganosas, como fotos e biografias que não lhes pertencem, os fraudadores criam relacionamentos fictícios para extorquir, chantagear ou obter vantagens. Rawat et al. [9] exploram o uso de técnicas de análise de redes sociais para identificar padrões de comportamento associados ao *catfishing*, permitindo uma melhor detecção e prevenção de atividades fraudulentas em plataformas digitais.

Honey Trap (armadilha afetiva) é uma tática de manipulação em que um indivíduo é atraído ou enganado por meio de sedução para obter informações, favores ou vantagens. Comumente usada em espionagem, investigações políticas ou situações pessoais, envolve conquistar a confiança ou criar vulnerabilidade emocional na vítima com intenções estratégicas, como acesso a segredos ou chantagem. Rawat et al. [9] apresentam *honey trap* como uma técnica investigativa usada para obter informações ou para comprometer alvos por meio de interações românticas ou sexuais, muitas vezes simuladas. Essa prática é empregada tanto no mundo físico quanto em plataformas digitais, onde perfis falsos podem ser criados para atrair alvos, coletar informações sensíveis ou manipulá-los.

Para diferenciação, *Romance Baiting* prioriza vantagens financeiras por meio de relacionamentos artificiais e exploração emocional prolongada. Já *Catfishing* engloba diferentes objetivos, que vão desde entretenimento e manipulação até fraudes financeiras. Por fim, *Honey Trap* é frequentemente usado em contextos de espionagem ou manipulação tática, com objetivos claros, como obtenção de informações ou controle de alvos.

Identity Theft (roubo de identidade) é o uso ilegal de informações pessoais, como nome, CPF, dados bancários e senhas, para obter benefícios financeiros ou cometer crimes. Isso pode ocorrer por meio de *phishing*, clonagem de cartão, *hacking* ou coleta de documentos descartados de forma inadequada. As consequências incluem dívidas fraudulentas, prejuízos ao histórico de crédito e até processos judiciais contra a vítima. Irshad and Soomro [28] exploram como o roubo de identidade evoluiu com o crescimento das redes sociais, que se tornaram um ambiente propício para crimes devido ao compartilhamento de informações pessoais. Os autores descrevem os métodos usados pelos criminosos e detalham os impactos do crime em plataformas populares como Facebook, Instagram e Twitter. Além disso, o artigo apresenta estatísticas sobre o aumento do roubo de identidades em redes sociais, enfatizando suas consequências financeiras e emocionais para as vítimas.

Scam by Bots (golpes realizados por robôs) consistem em golpes viabilizados por programas automatizados usados por fraudadores para enganar, manipular ou roubar informações. Os *bots* podem simular interações humanas, criar perfis falsos, distribuir links maliciosos, realizar ataques, como tentativa de acesso a contas, ou manipulação de mercados financeiros. Comuns em redes sociais, os *bots* enviam conteúdo fraudulento em massa, coletam dados pessoais e pressionam vítimas a compartilhar informações sensíveis. Ferrara et al. [30] exploram a evolução dos *bots*, destacando como eles podem imitar o comportamento humano de forma sofisticada para manipular discursos políticos, influenciar mercados financeiros, disseminar desinformação e roubar dados pessoais. O artigo também aborda métodos de detecção, como aprendizado de máquina e análise de redes, para diferenciar *bots* de usuários humanos, ressaltando os desafios de acompanhar sua crescente complexidade. Rawat et al. [9] concentram-se na aplicação da análise de redes sociais para investigar atividades suspeitas e criminosas em redes sociais. Utilizando métricas como centralidade e densidade, identificam contas falsas e nós influentes em atividades como recrutamento terrorista, radicalização e fraudes.

HashJacker (sequestro de *hashtags*) é a prática oportunista ou antiética de explorar *hashtags* populares para atrair atenção ou gerar engajamento. Essas *hashtags* são geralmente associadas a

eventos, campanhas ou tópicos virais, com o objetivo de promover conteúdos não relacionados, como propagandas, spam ou interesses pessoais. Embora seja uma estratégia para aumentar a visibilidade, o *hashjacking* é frequentemente mal visto por usuários e pode prejudicar a reputação de quem a utiliza, além de gerar ruído em discussões legítimas. Jain et al. [29] objetivam detectar e analisar o sequestro de *hashtags* no Twitter, através da identificação de palavras-chave mais relevantes em *tweets* relacionados a *hashtags* populares, categorizadas em temas como tecnologia, entretenimento, política e marcas, e correlacionando *tweets* que usam as mesmas com essas áreas temáticas, buscando definir uma reputação destes.

Sockpuppets Scam (golpe de marionetes) consiste no uso de perfis falsos ou identidades fictícias criadas por uma pessoa para fingir múltiplas vozes em discussões online, usadas para manipular opiniões, fraudar avaliações ou gerar engajamento artificial. Essas práticas, comuns em redes sociais, fóruns e avaliações, podem ser identificadas por padrões de postagens, conexões entre contas e uso do mesmo IP. Embora amplamente considerada antiética, e em alguns casos ilegais, *sockpuppets* são usados para enganar, distorcer debates ou ganhar vantagens indevidas em plataformas digitais. Rawat et al. [9] descreve *sockpuppets* como identidades falsas criadas em redes sociais para enganar outros usuários. Essas contas são utilizadas para manipular debates online, influenciar opiniões públicas, burlar restrições de plataformas e promover ações fraudulentas, incluindo votos múltiplos em enquetes e campanhas de propaganda. O trabalho aborda como técnicas de análise de redes sociais podem ajudar a identificar *sockpuppets* ao mapear conexões e padrões de comportamento, como frequência de interações, centralidade e uso de táticas coordenadas. Além disso, práticas como ataques *Sybil*, que envolvem múltiplos *sockpuppets* para manipular sistemas, são detalhadas como ameaças críticas em plataformas digitais.

Social Spamming (*spamming* social) é a prática de enviar mensagens, postagens ou interações não solicitadas em redes sociais, geralmente para promover produtos, serviços ou ideias de forma intrusiva. Exemplos incluem comentários promocionais, mensagens diretas indesejadas, marcações excessivas, perfis falsos e compartilhamento de *links* maliciosos. Além de atrapalhar a experiência do usuário, o social spam pode representar riscos de segurança, como roubo de dados ou instalação de *malware*, e prejudicar a reputação de marcas e indivíduos. Alom et al. [31] e Rawat et al. [9] usam técnicas de análise de dados para identificar e combater *spams* e atividades suspeitas em redes sociais. O primeiro foca no Twitter, combinando textos de tweets e metadados de usuários em um modelo de *deep learning* para classificar contas como *spammers* ou não. Já o segundo utiliza a análise de redes sociais para mapear conexões entre usuários e identificar padrões de comportamento associados a atividades maliciosas. [32] foca especificamente em fraudes financeiras no Instagram e propõe o uso de um modelo de aprendizado de máquina ajustado para identificar comentários fraudulentos ou spam em tempo real.

3.3 Discussão

A compreensão profunda dos golpes em redes sociais é essencial para a busca de alternativas de proteção para indivíduos e instituições. Com o aumento do uso das plataformas de redes sociais, golpistas recorrem a técnicas cada vez mais sofisticadas para roubar dados pessoais, cometer fraudes financeiras e manipular psicologicamente as vítimas. A conscientização sobre esses golpes e seus potenciais danos é um passo crucial para a prevenção, tornando os usuários mais atentos a práticas como *phishing*, perfis falsos e *links* suspeitos. Também, compreender as estratégias empregadas nesses golpes permite que as plataformas de redes sociais implementem medidas de segurança mais efetivas, contribuindo para a criação de um ambiente digital mais seguro e confiável.

Um problema recorrente nas discussões sobre golpes em redes sociais apresentadas na literatura consiste no desbalanceamento de eventos, na alta volatilidade do mercado e na dificuldade de acesso a dados de canais privados [5, 17, 18]. Além disso, a heterogeneidade dos dados entre plataformas também impacta negativamente as possibilidades e o esforço necessários para sua avaliação.

A constante evolução dos ataques pelos golpistas, o disfarce de identidade, a falta de ações regulatórias e a dificuldade na coleta de dados são algumas das limitações observadas por [10, 19, 20]. Uma das dificuldades apontadas por Ferrara et al. [30] é compreender como os *bots*, originários das primeiras formas de inteligência artificial (como os *chatbots*), se tornaram complexos e, em muitos casos, perigosos. Por sua vez, Ding et al. [21] enfrentaram desafios relacionados às técnicas de ofuscação e à necessidade de métodos rápidos e precisos, mesmo utilizando abordagens baseadas em palavras-chave e aprendizado de máquina.

Irshad and Soomro [28] destacam obstáculos como a evolução sofisticada do roubo de identidade, a falta de proteção adequada e a escassez de dados precisos sobre os métodos dos criminosos. Para Anesa [25], a principal dificuldade reside em compreender como os golpistas manipulam suas vítimas, explorando erros de julgamento para enganá-las mesmo diante de sinais evidentes. Já Yu et al. [27] apontam desafios relacionados à complexidade de medir os impactos de tecnologias como grandes modelos de linguagem e ao desenvolvimento de estratégias defensivas eficazes. No contexto de Erben and Waldis [32], as limitações de integração com redes sociais e os problemas de falsos positivos emergem como grandes dificuldades, enquanto Cui et al. [8] enfrentam desafios na integração de dados heterogêneos e na criação de modelos semânticos complexos para detectar *fake news*.

Com isso, é possível destacar três grandes desafios no contexto de golpes em redes sociais: (i) a coleta e o tratamento de amostras dos golpes; (ii) o uso de aprendizado de máquina; e (iii) a antecipação de novos golpes. A coleta sistemática de amostras de interações e transações nas redes sociais possibilita a criação de conjuntos de dados robustos, que podem ser analisados para identificar novas ameaças e comportamentos de golpistas, oferecendo uma base sólida para a detecção de fraudes. Para tal, é essencial aprimorar as ferramentas de acesso e integração com as plataformas de redes sociais. Além disso, embora já existam modelos de aprendizado de máquina, a evolução na detecção e classificação de golpes demanda a integração de dados não estruturados, como imagens e vídeos, bem como a adoção de abordagens de aprendizado contínuo, em que os modelos se atualizem automaticamente com novos padrões de fraudes. Por fim, com um maior volume de dados e modelos consistentes para processá-los, é possível antecipar e identificar padrões de golpes emergentes, permitindo que as plataformas de

redes sociais se preparem com sistemas de segurança mais proativos e estratégias de mitigação eficientes, antes que os ataques se tornem públicos e disseminados.

4 Conclusão

As redes sociais surgiram como plataformas para interação, criação de conteúdo e engajamento entre usuários, tornando-se parte integrante do cotidiano de bilhões de pessoas ao redor do mundo. No entanto, essas redes também são exploradas por usuários mal-intencionados para conduzir atividades fraudulentas, aproveitando vulnerabilidades humanas e tecnológicas. Os golpes realizados por meio das redes sociais podem resultar não apenas em perdas financeiras, mas também em graves danos emocionais, caracterizando-se como um sério problema para a sociedade moderna.

Nesse contexto, o presente artigo realiza um levantamento teórico sobre golpes conduzidos em redes sociais, com foco na análise da literatura existente relacionada a essas atividades fraudulentas. A partir dos objetivos definidos para a busca, seleção e tratamento dos dados, foram identificadas 5 categorias e 19 subcategorias de golpes. Os resultados demonstram que golpes em redes sociais constituem um tema atual na academia, amplamente explorado por pesquisadores ao redor do mundo. Assim, este trabalho contribui ao fornecer uma fundamentação teórica atualizada sobre golpes em redes sociais, apresentando de maneira sistematizada conceitos, referências e discussões em um intervalo de 10 anos, entre 2014 e 2024, e fomentando novas pesquisas na área.

Agradecimentos

Este trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES), Centro de Computação Científica e Software Livre (C3SL) e Fundação de Âmparo à Pesquisa e Inovação de Santa Catarina (FAPESC). Os autores também agradecem o Programa de Pós-Graduação em Informática da Universidade Federal do Paraná.

Referências

[1] Anastasiia Bessarab, Olha Mitchuk, Anna Baranetska, Natalia Kodatska, Olha Kvasnytsia, and Galyna Mykytiv. Social networks as a phenomenon of the information society. *Journal of Optimization in Industrial Engineering*, 14(Special Issue):17–24, 2021.

[2] Claudia Canali, Michele Colajanni, and Riccardo Lancellotti. Characteristics and evolution of content popularity and user relations in social networks. In *The IEEE symposium on Computers and Communications*, pages 750–756. IEEE, 2010.

[3] Tariq Rahim Soomro and Mumtaz Hussain. Social media-related cybercrimes and techniques for their prevention. *Applied Computer Systems*, 24(1):9–17, 2019.

[4] Srishti Gupta, Dhruv Kuchhal, Payas Gupta, Mustaque Ahamad, Manish Gupta, and Ponnurangam Kumaraguru. Under the shadow of sunshine: Characterizing spam campaigns abusing phone numbers across online social networks. In *ACM Conference on Web Science*, pages 67–76. ACM, 2018.

[5] Huy Nghiem, Goran Muric, Fred Morstatter, and Emilio Ferrara. Detecting cryptocurrency pump-and-dump frauds using market and social signals. *Expert Systems with Applications*, 182:115284, 2021.

[6] Dietmar Janetzko, Jonas Krauß, Frederic Haase, and Oliver Rath. On the involvement of bots in promote-hit-and-run scams–the case of rug pulls. In *5th International Conference on Advanced Research Methods and Analytics (CARMA 2023)*, pages 187–194. Editorial Universitat Politècnica de València, 2023.

[7] Vasilka Chergarova, Vinicius Arcanjo, Mel Tomeo, Jeronimo Bezerra, Luis Marin Vera, and Anthony Uloa. Cryptocurrency fraud: A study on the characteristics of criminals who are using fake profiles on a social media platform to persuade individuals to invest into cryptocurrency. *Issues in Information Systems*, 23(3), 2022.

[8] Limeng Cui, Suhang Wang, and Dongwon Lee. Same: sentiment-aware multimodal embedding for detecting fake news. In *International Conference on Advances in Social Networks Analysis and Mining*, pages 41–48. IEEE/ACM, 2019.

[9] Romil Rawat, Vinod Mahor, Sachin Chirgaiya, and Abhishek Singh Rathore. Applications of social network analysis to managing the investigation of suspicious activities in social media platforms. In *Advances in Cybersecurity Management*, pages 315–335. Springer, 2021.

[10] Xigao Li, Amir Rahmati, and Nick Nikiforakis. Like, comment, get scammed: Characterizing comment scams on media platforms. In *Proceedings 2024 Network and Distributed System Security Symposium*, 2024.

[11] Cassandra Cross, Russell G Smith, and Kelly Richards. Challenges of responding to online fraud victimisation in australia. *Trends and issues in crime and criminal justice*, (474):1–6, 2014.

[12] Eddie Bouy B Palad, Marivic S Tangkeko, Lissa Andrea K Magpantay, and Glenn L Sipin. Document classification of filipino online scam incident text using data mining techniques. In *International Symposium on Communications and Information Technologies*, pages 232–237. IEEE, 2019.

[13] Mudiana Mokhsin, Azhar Abdul Aziz, Amer Shakir Zainol, Norshima Humaidi, and Nur Ain Adnin Zaini. Probability model: Malaysian consumer online shopping behavior towards online shopping scam. *International Journal of Academic Research in Business and Social Sciences*, 9(1), 2019.

[14] Suhaib Al-Rousan, Abdullah Abuhussein, Faisal Alsubaei, Lynn Collen, and Sajjan Shiva. Ads-guard: Detecting scammers in online classified ads. In *Symposium Series on Computational Intelligence*, pages 1492–1498. IEEE, 2020.

[15] Cassandra Cross and Michael Kelly. The problem of "white noise": Examining current prevention approaches to online fraud. *Journal of Financial Crime*, 23(4): 806–818, 2016.

[16] Ross Phillips and Heidi Wilder. Tracing cryptocurrency scams: Clustering replicated advance-fee and phishing websites. In *International Conference on Blockchain and Cryptocurrency*, pages 1–8. IEEE, 2020.

[17] JT Hamrick, Farhang Rouhi, Arghya Mukherjee, Amir Feder, Neil Gandal, Tyler Moore, and Marie Vasek. The economics of cryptocurrency pump and dump schemes. In *Workshop on the Economics of Information Security*. NSF, 2019.

[18] Mehrnoosh Mirtaheri, Sami Abu-El-Haija, Fred Morstatter, Greg Ver Steeg, and Aram Galstyan. Identifying and analyzing cryptocurrency manipulations in social media. *IEEE Transactions on Computational Social Systems*, 8(3):607–617, 2021.

[19] Iman Vakilinia. Cryptocurrency giveaway scam with youtube live stream. In *Annual Ubiquitous Computing, Electronics & Mobile Communication Conference*, pages 0195–0200. IEEE, 2022.

[20] Enze Liu, George Kappos, Eric Mugnier, Luca Invernizzi, Stefan Savage, David Tao, Kurt Thomas, Geoffrey M Voelker, and Sarah Meiklejohn. Give and take: An end-to-end investigation of giveaway scam conversion rates. *arXiv preprint arXiv:2405.09757*, 2024.

[21] Yan Ding, Nurbol Luktarhan, Keqin Li, and Wushour Slamu. A keyword-based combination approach for detecting phishing webpages. *computers & security*, 84:256–275, 2019.

[22] Edwin D Frauenstein and Stephen V Flowerday. Social network phishing: Becoming habituated to clicks and ignorant to threats? In *Information Security for South Africa*, pages 98–105. IEEE, 2016.

[23] Najmeh Miramirkhani, Oleksii Starov, and Nick Nikiforakis. Dial one for scam: Analyzing and detecting technical support scams. In *Annual Network and Distributed System Security Symposium*, volume 16. Internet Society, 2016.

[24] Sampsa Rauti and Ville Leppänen. "you have a potential hacker's infection": A study on technical support scams. In *International Conference on Computer and Information Technology*, pages 197–203. IEEE, 2017.

[25] Patrizia Anesa. Lovextortion: Persuasion strategies in romance cybercrime. *Discourse, Context & Media*, 35:100398, 2020.

[26] Cassandra Cross. Romance baiting, cryptorom and 'pig butchering': an evolutionary step in romance fraud. *Current Issues in Criminal Justice*, 36(3):334–346, 2024.

[27] Jingru Yu, Yi Yu, Xuhong Wang, Yilun Lin, Manzhi Yang, Yu Qiao, and Fei-Yue Wang. The shadow of fraud: The emerging danger of ai-powered social engineering and its possible cure. *arXiv preprint arXiv:2407.15912*, 2024.

[28] Shareen Irshad and Tariq Rahim Soomro. Identity theft and social media. *International Journal of Computer Science and Network Security*, 18(1):43–55, 2018.

[29] Nikita Jain, Pooja Agarwal, and Juhi Pruthi. Hashjacker-detection and analysis of hashtag hijacking on twitter. *International journal of computer applications*, 114(19), 2015.

[30] Emilio Ferrara, Onur Varol, Clayton Davis, Filippo Menczer, and Alessandro Flammini. The rise of social bots. *Communications of the ACM*, 59(7):96–104, 2016.

[31] Zulfikar Alom, Barbara Carminati, and Elena Ferrari. A deep learning model for twitter spam detection. *Online Social Networks and Media*, 18:100079, 2020.

[32] Stefan Erben and Andreas Waldis. Scamspot: Fighting financial fraud in instagram comments. *arXiv preprint arXiv:2402.08869*, 2024.