UNIVERSIDADE DE SÃO PAULO
Instituto de Ciências Matemáticas e de Computação

Processamento automatizado da comunicação textual de atores de ameaça cibernética para suporte à construção de consciência situacional sobre seus interesses

Giovani Ferreira Silverio da Silva

Monografia - MBA em Inteligência Artificial e Big Data

ICMC USP SÃO CARLOS

Giovani Ferreira Silverio da Silva

Processamento automatizado da comunicação textual de atores de ameaça cibernética para suporte à construção de consciência situacional sobre seus interesses

Monografia apresentada ao Departamento de Ciências de Computação do Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo - ICMC/USP, como parte dos requisitos para obtenção do título de Especialista em Inteligência Artificial e Big Data.

Área de concentração: Inteligência Artificial

Orientador: Prof. Dr. Robson de Oliveira Albuquerque

Versão original

São Carlos

2024

Ficha catalográfica elaborada pela Biblioteca Prof. Achille Bassi
e Seção Técnica de Informática, ICMC/USP,
com os dados inseridos pelo(a) autor(a)

S586p

Silva, Giovani Ferreira Silverio da
Processamento automatizado da comunicação textual de atores de ameaça cibernética para suporte a construção de consciência situacional sobre seus interesses / Giovani Ferreira Silverio da Silva; orientador Robson de Oliveira Albuquerque. -- São Carlos, 2024.
94 p.

Trabalho de conclusão de curso (MBA em Inteligência Artificial e Big Data) -- Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, 2024.

1. inteligência de ameaças. 2. crime cibernético. 3. fraude digital. 4. modelagem de tópicos. I. Albuquerque, Robson de Oliveira, orient. II. Título.

Bibliotecários responsáveis pela estrutura de catalogação da publicação de acordo com a AACR2:
Gláucia Maria Saia Cristianini - CRB - 8/4938
Juliana de Souza Moraes - CRB - 8/6176

Giovani Ferreira Silverio da Silva

Automated processing of textual communication of cyber threat actors to support the development of situational awareness about their interests

Monograph presented to the Departamento de Ciências de Computação do Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo - ICMC/USP, as part of the requirements for obtaining the title of Specialist in Artificial Intelligence and Big Data.

Concentration area: Artificial Intelligence

Advisor: Prof. Dr. Robson de Oliveira Albuquerque

Original version

São Carlos

2024

RESUMO

Silva, G. F. S. **Processamento automatizado da comunicação textual de atores de ameaça cibernética para suporte à construção de consciência situacional sobre seus interesses**. 2024. 89p. Monografia (MBA em Inteligência Artificial e Big Data) - Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos, 2024.

Este trabalho propõe um método para processamento automatizado de dados textuais relacionados às conversas entre atores de ameaça cibernética conduzidas em fóruns de mensagens instantâneas. O método objetiva transformar os dados textuais em informações estruturadas para suportar a criação de consciência situacional sobre os interesses dos grupos e perfis presentes no ecossistema de crimes cibernéticos no Brasil, além de fornecer uma maneira para que analistas de inteligência visualizem esses dados. As informações resultantes consistem da identificação dos grupos e perfis participantes do ecossistema, os tópicos e assuntos em que se interessam, e a similaridade entre eles baseado nos seus respectivos tópicos de interesse. O método proposto é avaliado como capaz de automatizar a percepção dos elementos do ecossistema e suas características (nível 1 de consciência situacional), além de dar suporte para a compreensão de seus significados (nível 2) e projeção de estado futuro (nível 3). Entretanto, os resultados finais obtidos possuem forte dependência na qualidade da etapa de análise de tópicos, requerendo cuidados especiais na implementação a fim de evitar resultados inadequados para o nível de confiança na informação desejado.

Palavras-chave: fraude digital, crime cibernético, análise de tópicos, inteligência de ameaças.

ABSTRACT

Silva, G. F. S. **Automated processing of textual communication of cyber threat actors to support the development of situational awareness about their interests**. 2024. 89p. Monograph (MBA in Artificial Intelligence and Big Data) - Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos, 2024.

This work proposes a method for automated processing of textual data of conversations between cyber threat actors conducted on instant messaging apps. The method aims to transform textual data into structured information to support the creation of situational awareness about the interests of groups and profiles present in the cybercrime ecosystem in Brazil, in addition to providing a way for intelligence analysts to visualize this data. The resulting information consists of the identification of the groups and profiles participating in the ecosystem, the conversational topics in which they are interested, and the similarity between them based on their respective topics of interest. The method is evaluated as capable of automating the perception of ecosystem elements and their characteristics (level 1 of situational awareness), in addition to supporting the understanding of their meanings (level 2) and projection of future state (level 3). However, the final results obtained are strongly dependent on the quality of the topic analysis stage, requiring special care in implementation in order to avoid inadequate results for the desired level of confidence in the information.

Keywords: digital fraud, cybercrime, topic analysis, threat intelligence.

LISTA DE FIGURAS

LISTA DE TABELAS

LISTA DE ABREVIATURAS E SIGLAS

CTI	*Cyber Threat Intelligence*
LSA	*Latent Semantic Analysis*
LDA	*Latent Dirichlet Allocation*
PCA	*Principal Component Analysis*
USP	Universidade de São Paulo
USPSC	Campus USP de São Carlos
BERT	*Bidirectional Encoder Representations from Transformers*
UMAP	*Uniform Manifold Approximation and Projection*
HDBSCAN	*Hierarchical Density-Based Spatial Clustering of Applications with Noise*
TF-IDF	*Term Frequency - Inverse Document Frequency*

SUMÁRIO

1 INTRODUÇÃO

Produzir inteligência sobre ameaças cibernéticas é uma tarefa que exige um bom embasamento metodológico e direcionamento para gerar resultados adequados. Uma das grandes dificuldades para a produção de inteligência cibernética é a evolução temporal rápida das ameaças, que se moldam, atualizam, e formulam novas técnicas de ataque de maneira muito veloz devido a natureza do ambiente cibernético.

Portanto, uma das formas de acompanhar a evolução das ameaças cibernéticas é manter uma constante consciência situacional, que pode ser entendida como a percepção dos elementos dentro de um volume de tempo e espaço, a compreensão do seu significado e a projeção de seu estado em um futuro próximo (ENDSLEY, 1995). O nível de *percepção* da consciência situacional envolve listagem de dados e dos elementos através da identificação e coleta de informações essenciais, procurando respostas para “Quem?”, “Onde?”, “Quando?” e “Quantos?”; enquanto o nível de *compreensão* busca entender a importância dos elementos, agregar significado às informações coletadas, verificar a relevância dessas informações, e procurar respostas para “Porquê?”, “Qual a capacidade?”, “Qual a relevância?” e “Qual ação?” (SILVA, 2023). Manter uma consciência situcional implica em uma constante identificação e monitoramento de ameaças, seus respectivos graus de relevância em relação aos elementos de interesse, suas relações e interdependências, entre outos fatores.

No contexto brasileiro, as ameaças cibernéticas se materializam principalmente no formato de fraudes digitais e ataques cibernéticos, sendo as fraudes digitais mais impactantes no dia-a-dia do cidadão brasileiro (IPESPE - Instituto de Pesquisas Sociais, Políticas e Econômicas, 2023). Os atores de ameaça envolvidos na execução desses crimes digitais possuem a necessidade de se comunicar, compartilhar experências, negociar a venda de dados e informações, contratar serviços criminosos, entre várias outras necessidades, fomentando um fértil ecossistema criminoso. Esse ecosistema fértil para fraudes digitais e ataques cibernéticos atrai muitos criminosos, gerando relações de interdependência onde o resultado de um crime serve como matéria-prima para outros. Exemplos conhecidos de interdependência são as credenciais coletadas por *malware stealers* e as informações sobre cartões de crédito que são amplamente comercializadas para viabilizar outros tipos de crimes (SOCRADAR, 2023).

As necessidades de comunicação e negociação dos atores de ameaças cibernéticas são comumente supridas através do uso de forúns de mensagens que fornecem um grau adequado de privacidade digital, sendo que os fóruns de mensagens selecionados mudam de acordo com a garantia da privacidade oferecida. Essa busca por privacidade se dá pelo interesse dos criminosos em manter suas identidades preservadas, a fim de evitar repressão pelas forças da lei.

O monitoramento e a análise da quantidade massiva de dados textuais produzidas por esses atores de ameaça nos fóruns de mensagens é uma importante fonte de informação sobre seus interesses, instituições alvejadas, métodos utilizados, entre outros fatores. É comum que empresas de inteligência de ameaças cibernética façam investimentos altos na coleta, processamento e análise desses dados. Entretanto, essas empresas normalmente exploram essas informações em um modelo limitado a eventos de interesse, raramente fazendo uso de uma análise estruturada da base de textos coletada. Alguns desafios encontrados para análise de inteligência cibernética sobre esses dados textuais podem ser visualizados na Tabela 1.

Desafio	Razão
Informações de pouca confiabilidade	Qualquer pessoa pode realizar um cadastro e comentar sobre qualquer assunto na maior parte dos fóruns, o que implica em maior dificuldade de rastrear identidades digitais e na possibilidade de tópicos serem discutidos com o intuito de enganar ou induzir erros na análise de inteligência.
Informações irrelevantes	Muitos dos fóruns de mensagens são utilizados por atores de ameaça para discutir tópicos paralelos aos crimes mas de pouca relevância para inteligência cibernética. É necessário filtrar a grande quantidade de tópicos irrelevantes em toda análise de inteligência.
Velocidade das ameaças no ambiente cibernético	O tempo entre a detecção de uma ameaça para uma determinada instituição e a materialização em forma de ataque pode ser bem curto, o que requer maior tempestividade.
Volume de dados	A quantidade de fóruns e mensagens publicadas é massiva. O processamento desses dados exige uma infraestrutura computacional robusta e, possivelmente, incorre em custos operacionais elevados.
Tempo de processamento dos dados	Devido ao volume de dados textuais, o processamento e análise automatizada podem demorar muito tempo, tornando o resultado produzido inoportuno.

Tabela 1 – Desafios na análise de inteligência de fóruns de crimes cibernéticos.

A aplicação de técnicas de mineração de dados e aprendizagem de máquina para análise de dados textuais oriundos de fóruns de mensagens é um tópico com estudos relacionados (SHAH *et al.*, 2020), mas a aplicação dentro do contexto brasileiro, com o objetivo de oferecer suporte adequado a construção de consciência situacional sobre os interesses de atores de ameaças cibernéticas, é um tópico ainda não explorado de acordo com o melhor conhecimento do autor.

O processamento automatizado da estrutura de relacionamentos de atores de ameaça no ecossistema de crimes digitais brasileiro, seus graus de influência, e quais os

conjuntos de atores relacionados a interesses específicos, fornecem um importante contexto sobre o ambiente para analistas de inteligência cibernética. Dessa forma, esta pesquisa busca avaliar técnicas de mineração de dados e processamento de linguagem natural nas características de velocidade de processamento de dados em larga escala, capacidade de identificação acurada dos tópicos discutidos em mensagens de textos curtas e informais, qualidade de agrupamento dos atores de ameaça por similaridade dos seus tópicos de interesse, capacidade de atualização do resultado após novas informações serem coletadas, facilidade de interpretação dos dados por analistas de inteligência, e facilidade de interação com o resultado para definir ou refinar perguntas de interesse; com o propósito de suportar a construção de consciência situacional sobre os atores de ameaça participantes de fóruns de mensagens.

1.1 Motivação e Justificativa

O Brasil constantemente figura entre os países com o maior nível de atividade de ameaças cibernéticas, tanto no aspecto de ataques cibernéticos como também em fraudes digitais. No âmbito de fraudes digitais, a realidade é que elas fazem parte do dia-a-dia do cidadão brasileiro, com aproximadamente 4 em cada 10 entrevistados (38%) sendo vítimas de golpes ou tentativas de golpes digitais (IPESPE - Instituto de Pesquisas Sociais, Políticas e Econômicas, 2023). Entretanto, a investigação e repressão policial desses crimes digitais é limitada (SANTOS; BOTELHO, 2023), requerendo técnicas de investigação alternativas e mais custosas para suprir essas lacunas (PAULINO *et al.*, 2022). Em razão disso, as instituições privadas e governamentais comumente afetadas por ameaças digitais buscam em serviços de inteligência cibernética uma forma de se proteger proativamente, estando um passo a frente das ameaças.

A manutenção da consciência situacional é um requisito que produtos e serviços de inteligência cibernética devem considerar para buscar resultados melhores. Apesar de diversas informações serem necessárias para construção dessa consciência, explorar por completo os dados textuais oriundos de fóruns de mensagens sobre crimes digitais pode possuir muito valor dentro do contexto brasileiro, como por exemplo: revelar relações de interdependência entre atores de ameaça, os seus graus de influência no ecossistema criminoso, os principais tópicos de interesse de cada ator, seus subgrupos preferidos, e seu nível de atividade criminosa. Os dados textuais já são úteis atualmente para detectar ameaças iminentes e novas técnicas de utilizadas por criminosos, mas são pouco explorados para construção de consciência situacional cibernética.

Assim, considerando os aspectos apresentados, é relevante a proposta de um método capaz de processar automaticamente os dados textuais coletados e extrair suas principais características para suportar a construção de consciência situacional.

1.2 Objetivo Geral

Este trabalho tem por objetivo propor um método capaz de processar automaticamente os dados textuais oriundos de fóruns de mensagens instantâneas, identificando os grupos e perfis participantes, seus respectivos tópicos de interesse, e seus graus de similaridade a fim de suportar a construção de consciência situacional sobre o ecossistema de crimes cibernéticos Brasileiro.

1.2.1 Objetivos Específicos

De acordo com o objetivo geral estabelecido, os seguintes objetivos específicos são endereçados:

- Definir as etapas do processo de coleta dos dados textuais dos fóruns de mensagens instantâneas;
- Identificar as etapas de preprocessamento textual necessárias para as características dos textos coletados;
- Identificar técnicas de mineração de dados e processamento de linguagem natural que lidem bem com os desafios impostos pelas características dos dados textuais coletados;
- Avaliar o desempenho de diversas técnicas de mineração de dados e processamento de linguagem natural na análise de tópicos dos textos coletados e no agrupamento dos perfis e grupos por similaridade;
- Avaliar os resultados do método proposto em casos de uso baseados em situações reais;
- Identificar técnicas de visualização de dados que possibilitem um analista de inteligência a compreender as características e relações do grande volume de grupos e perfis.

1.3 Principais Contribuições

As principais contribuições realizadas por este trabalho são:

- Proposta de um método que oferece suporte na criação de consciência situacional sobre os interesses de atores de ameaça baseado em suas comunicações textuais informais;
- Identificação das etapas necessárias para a estruturação de dados textuais em formato que possibilite a investigação iterativa por analista humano na construção de consciência situacional;

- Avaliação da performance de técnicas de análise de tópicos em base de dados contendo textos muito informais e com grande assimetria na quantidade de *tokens* presentes em cada mensagem;
- Avaliação da qualidade de diferentes técnicas de representação de texto em espaço vetorial em bases de dados com textos informais e com grande assimetria de *tokens* presentes em cada documento.

1.4 Organização do Trabalho

Para facilitar a compreensão, este trabalho é organizado nos seguintes capítulos: O capítulo 1 detalha o contexto do problema, a motivação do trabalho e os objetivos almejados; O capítulo 2 elabora os conceitos fundamentais de Inteligência, Mineração de Dados e Processamento de Linguagem Natural para o entendimento completo do método proposto e seus resultados, e discute os trabalhos relacionados, as diferentes abordagens utilizadas por outros autores, e os fatores que diferenciam este trabalho dos outros; O capítulo 3 detalha as etapas que compõem o método proposto e suas respectivas ordens de execução; O capítulo 4 detalha os experimentos realizados para availar o método e algumas etapas específicas, além dos resultados obtidos e as melhores soluções encontradas; Por fim, o capítulo 5 elabora sobre os resultados obitidos, as limitações percebidas do método proposto, os cenários onde a aplicação do método deve ser realizada com cuidados adicionais, e alguns caminhos para continuação da linha de pesquisa.

2 FUNDAMENTAÇÃO TEÓRICA

Este capítulo contém uma revisão teórica dos conceitos fundamentais para o embasamento e entendimento da solução proposta, além de uma análise de trabalhos relacionados e o atual estado da arte. Para facilitar o entendimento, as duas grandes áreas de conhecimento e a análise de trabalhos relacionados estão apresentadas em seções distintas.

Portanto, na seção 2.1, são abordados os conceitos referentes a atividade de inteligência e sua aplicação dentro do ambiente cibernético, além dos tópicos relacionadas a características de atores de ameaça cibernética e seus comportamentos. Na seção 2.2, são tratados os conceitos gerais relacionados a mineração de dados e os conceitos mais específicos relacionados as técnicas testadas ou aplicadas pela solução proposta. Por fim, a seção 2.4 traz uma análise dos principais trabalhos relacionados ao tema, suas soluções propostas e principais resultados.

2.1 Inteligência

Inteligência é um conceito que possui diversas definições dependentes de contexto. De forma mais ampla, pode ser compreendida como produto (conhecimento), as atividades que resultam nesse produto (processo), e as organizações que engajam nessas atividades (KENT, 2015).

Como produto, inteligência pode ser entendida como o conhecimento produzido através da análise e interpretação de dados, informações e outros conhecimentos que foram coletados, processados, integrados e avaliados. Esse produto se materializa através de informes (conhecimento narrativo-descritivo), apreciações (conhecimento interpretativo), e estimativas (conhecimento interpretativo-prospectivo).

Como atividade, inteligência pode ser entendida como o emprego de técnicas e ações especializadas destinadas à: obtenção e análise de dados e informações, e produção e disseminação de conhecimentos; constituindo um instrumento de assessoria para atender as demandas de um tomador de decisão qualquer.

E como organização, inteligência pode ser entendida como uma estrutura de trabalho contínuo responsável pelo exercício profissional da atividade de inteligência. Podem ser classificados como serviços de inteligência, quando têm por finalidade a execução da atividade de inteligência, ou como frações de inteligência, quando integram instituições que possuem outras finalidades (ABIN, 2023).

2.1.1 Dado, Informação e Conhecimento

Dado é a menor unidade de representação de aspecto da realidade. Possui significado descontextualizado, sem nenhuma atribuição extra além do simples registro (ABIN, 2023). Dados podem ser gerados por pessoa ou meio técnico e são normalmente classificados em estruturados, semi-estruturados e não-estruturados.

A análise de um dado exige esforço de identificação de elementos que possam ser atribuídos para composição de estrutura de descrição (definição, aparência, composição, função ou atuação) e de narração (o que, quem, quando, onde, por quê, como).

Figura 1 – Evolução do dado à inteligência

Informação é uma representação de aspecto da realidade com significado contextualizado resultante do processamento (limpeza, correção, seleção, cruzamento, organização, tradução, formatação, sumarização e ordenamento) e interpretação de dados de forma metódica, racional e objetiva (ABIN, 2023). Informações podem ser geradas tanto por pessoas quanto por meios computacionais sem intervenção humana.

Apesar da informação possuir contexto através de estrutura de descrição e narração, possuindo conteúdo e significado compreensíveis, ela não contém corroboração ou demonstração que justique sua veracidade. Portanto, a análise de informação exige busca por coerência interna e externa, e de outras fontes.

Conhecimento é uma representação da realidade com significado contextualizado assumido como verdadeiro e validado. Por conter o significado do objeto observado e ser justificadamente plausível e verossímil, o conhecimento pode ser utilizado para a tomada de decisão.

2.1.2 Disciplinas de Inteligência em Relação a Fonte de Dados

Parte integral do trabalho de inteligência é a coleta de insumos (dados, informações e conhecimentos) para análise. Essa atividade de coleta ocorre através de diferentes disciplinas de coleta de inteligência, como a inteligência de fontes humanas (HUMINT) e a inteligência de fontes abertas (OSINT).

Disciplina	Definição e Contexto
Inteligência de Fontes Abertas (OSINT)	Realizada com base em dados e informações que estão disponíveis de forma pública para qualquer um através de meios legais, incluindo solicitação, observação, ou compra.
Inteligência de Fontes Humanas (HUMINT)	É a inteligência realizada com base em dados obtidos de pessoas. Reúne dados, informações, conhecimentos e percepções originados de relatos feitos ou trazidos por indivíduos externos a organização de inteligência.
Inteligência de Sinais (SIGINT)	Realizada com base em dados obtidos por interpretação e decodificação de comunicações e sinais eletromagnéticos. Vem ganhando subdivisões relevantes, como inteligência produzida com base em dados obtidos no espaço cibernético (CYBINT).
Inteligência de Imagens (IMINT)	Realizada com base em dados obtidos por produção de imagens fotográficas e multiespectrais. Pode incluir desde especialização para imagens geoespaciais obtidas com satélites até a avaliação de fotos digitais ou analógicas.
Inteligência Geoespacial (GEOINT)	Realizada com base em imagens e dados de geolocalização obtidos para descrever, avaliar e representar visualmente características físicas ou atividades geograficamente referenciadas. Considerada parte da IMINT pela comunidade de inteligência dos EUA.
Inteligência de Medidas e Assinaturas (MASINT)	Realizada com base em dados obtidos por aferição de emanações, como a sísmica e a térmica, em geral decorrentes de assinaturas de eventos, como explosões atômicas.
Inteligência de Mídias Sociais (SOCMINT)	Focada em dados e informações publicados em mídias sociais e seus respectivos metadados. Coleta de grande volume de informações para análises de sentimentos, padrões de publicações e avaliação de relevância de temas são exemplos.

Tabela 2 – Principais disciplinas de coleta e algumas de suas subdivisões.

A comunidade de inteligência dos Estados Unidos da América (EUA) compreende que existem 5 principais disciplinas de coleta de inteligência: inteligência de fontes abertas (OSINT), inteligência de fontes humanas (HUMINT), inteligência de medidas e assinaturas (MASINT), inteligência de sinais (SIGINT), e inteligência de imagens (IMINT), com suas respectivas subdivisões mais especializadas (CLARK, 2013b).

O entendimento como *disciplina* pela comunidade de inteligência dos EUA se refere ao fato de que diferentes tipos de insumos só podem ser coletados e interpretados através de tratamento especializado conforme sua natureza. Por exemplo, a coleta, processamento e análise de imagens obtidas via satélite requer tecnologias especializadas e expertise

profissional específica. Dessa forma, o entendimento como disciplina prevê que cada especialização possui seu próprio ciclo de inteligência com seus respectivos processos e mecanismos especializados (LOWENTHAL; CLARK, 2015).

No Brasil, o entendimento proposto pela Doutrina da Atividade de Inteligência da Agência Brasileira de Inteligência (ABIN, 2023) aborda o conceito de classificação pela origem do dado ao invés de disciplinas de coleta, sendo três categorias de classificação:

- Inteligência de Fontes Humanas (HUMINT): É a inteligência realizada com base em dados obtidos de pessoas;
- Inteligência de Fontes Técnicas (TECHINT): É a inteligência realizada com base em dados obtidos por meios técnicos. Reúne informações e dados originados do emprego de equipamentos, que requerem perícia em seu manuseio;
- Inteligência de Fontes Abertas (OSINT): É a inteligência realizada com base em dados disponíveis, de livre acesso.

A doutrina prevê medidas, necessidades e competências distintas para o tratamento e processamento das diferentes categorias de dados, gerando implicações para o processo de produção, formação de servidores e estruturas organizacionais diferenciadas para coleta e análise de cada categoria de dado (ABIN, 2023).

Entretanto, apesar da previsão de implicações específicas para cada origem de dado, o conceito de disciplina de inteligência é mais amplo, sendo entendido como uma área bem definida de planejamento, coleta, exploração, análise, e comunicação (reporting) de inteligência fazendo uso de uma categoria específica de recursos técnicos ou humanos (JOINT CHIEFS OF STAFF, 2013).

Além das cinco principais disciplinas de inteligência, é compreendido que existem subdivisões mais especializadas delas, como a inteligência de comunicações (COMINT), subdivisão da inteligência de sinais (SIGINT), e a inteligência cibernética (CYBINT), que apesar de não se encaixar perfeitamente dentro de nenhuma disciplina principal, é comumente considerada uma subdivisão da SIGINT (CLARK, 2013a). A definição das principais disciplinas e algumas das suas subdivisões mais conhecidas podem ser observadas na tabela 2.

2.1.2.1 Inteligência de Redes Sociais

A Doutrina da Atividade de Inteligência (ABIN, 2023) define SOCMINT como a inteligência focada em dados e informações publicados em mídias sociais e seus respectivos metadados. Existem diversas maneiras de explorar SOCMINT para os mais diversos contextos. No contexto estatal e governamental, a aplicação de SOCMINT foi inicialmente considerada como uma oportunidade de: consciência situacional em tempo real de violência

e respostas de emergência, entendimento de comportamentos e atividades de grupos de interesse, e identificação de intenção criminal, como observado em (OMAND; MILLER, 2012) e (ŞUŞNEA; IFTENE, 2018).

Entretanto, apesar da crença otimista de que SOCMINT poderia prover feeds de inteligência quase instantâneos e consciência situacional para policiais, no Reino Unido se observou que os algoritmos de análise sentimental e predição de violência e tensão social ainda esbarram em dificuldades como o contexto emocional das mensagens e as relações de causalidade que transformam mensagens nas redes sociais em violência (DOVER, 2020).

Além disso, existem outros fatores limitantes que atrapalham a confiança do uso de SOCMINT, como (DOVER, 2020):

- A maioria dos comentários sobre um evento ocorrerem somente após o acontecimento dele, atrapalhando a capacidade preditiva de curto prazo;
- A necessidade de se apoiar em sistemas de processamento automatizado que podem falhar devido a quantidade massiva de dados que precisam ser coletados e processados;
- A possível necessidade de distinguir entre perfis humanos e não humanos (*bots*);
- As informações coletadas não são avaliadas e confirmadas e, portanto, precisam ser tratadas com mais cautela;
- A possível influência da desinformação intencional (*dezinformatsiya*) ou involuntária;

Apesar dos fatores limitantes, SOCMINT é capaz de prover monitoramento de horizonte (*horizon scanning*), identificação de elementos fora do comum, suporte para definição de novos alvos de coleta, entendimento de sentimento de comunidade, e avisos estratégicos (*strategic warning*), possuindo menos utilidade para alertas de ameaça iminente. No quesito de monitoramento de horizonte, análise de tendências e predição de mudanças de longo prazo são atividades em que SOCMINT pode ser útil (DOVER, 2020).

Portanto, apesar de possuir limitações devido a natureza descontrolada em que os dados de mídias sociais são produzidos e publicados, SOCMINT pode ser utilizada para alcançar resultados de relevância desde que a atenção necessária para com os objetivos e tratamento dos dados seja observada a fim de mitigar as limitações e riscos inerentes.

2.1.3 Ciclo de Inteligência

O conceito de ciclo de inteligência é a representação mais geral e abrangente de como o processo de produção de inteligência é idealizado. O termo *idealizado* é importante pois o conceito têm sido bastante questionado e reavaliado por não representar acuradamente a produção de inteligência atual, como em (CLARK, 2019) e (WARNER, 2013). Apesar

de estar sujeito a questionamentos mais incisivos sobre os aspectos práticos, o ciclo de inteligência ainda é relevante para o entendimento conceitual da produção de inteligência.

Entretanto, como as principais organizações de inteligência (e.g. agências governamentais) se valem do segredo para o sucesso das suas atividades, o detalhamento dos seus métodos nem sempre está disponível. Dessa forma, o conceito de ciclo de inteligência possui diversas interpretações dependentes da visão e características de cada organização.

Porém, apesar das diferentes interpretações, as ações e os objetivos gerais almejados por cada ciclo de inteligência são bem similares, sendo (TANABE, 2023):

- Definir os requisitos de informação de um decisor (Direção);
- Estabelecer um plano para atender a esses requisitos (Planejamento);
- Coletar as informações necessárias para desenvolver um produto final (Coleta);
- Processar as informações coletadas em um formato utilizável (Processamento);
- Analisar as informações para obter significado (Análise);
- Consolidar o produto de Inteligência (Produção);
- Transmitir a Inteligência ao usuário que a demandou (Disseminação);
- Avaliar constantemente todas estas ações (Avaliação).

É importante ressaltar que as etapas podem ser adaptadas ao contexto e não devem ser entendidas como uma sequência encadeada de ações, pois algumas delas permeiam as outras durante toda a execução do ciclo, como é o caso do Planejamento, que organiza todo o desenvolvimento a partir das definições dos requisitos do consumidor, e da Avaliação, que busca garantir que as etapas foram executadas de forma adequada alcançando os objetivos propostos.

A figura 2 traz uma melhor visibilidade sobre o ciclo de inteligência genérico, contendo todas as etapas que são comumente agrupadas ou ocultas em outras interpretações do ciclo de inteligência.

Por fim, existem trabalhos sugerindo novas formas de estruturação alternativa ao ciclo de inteligência, como (GILL; PHYTHIAN, 2013) e (DAVIES; GUSTAFSON; RIGDEN, 2013), propondo modelos mais complexos que procuram evitar uma narrativa linear de produção de inteligência e separação de responsabilidades, como "quem coleta não é responsável sobre o produto final". O entendimento aprofundado dos diferentes modelos propostos não é objetivo deste trabalho.

Figura 2 – Etapas do ciclo de inteligência genérico. Adaptado de (TANABE, 2023).

2.1.4 Consciência Situacional

O campo de estudo da consciência situacional é bastante amplo e possui diversas aplicações de contextos específicos. É possível encontrar aplicações de consciência situacional desde tripulação de navio (MELNYK; BYCHKOVSKY; VOLOSHYN,) e anestesiologistas (SCHULZ *et al.*, 2013), até gestão de riscos de segurança da informação (WEBB *et al.*, 2014), inteligência cibernética (AHMAD *et al.*, 2021; FRANKE; BRYNIELSSON, 2014) e inteligência (GAETA; LOIA; ORCIUOLI, 2021).

De certa forma, consciência situacional pode ser entendida por diversas perspectivas, como a técnica e a cognitiva. Do ponto de vista técnico, a essência da consciência situacional é compilar, processar e fundir dados, sendo que o processamento inclui a necessidade de avaliar os fragmentos de dados, as informações fundidas, e fornecer uma estimativa racional da qualidade da informação (FRANKE; BRYNIELSSON, 2014). Pelo aspecto cognitivo, a consciência situacional diz sobre a capacidade humana de percepção de informações, compreensão de suas implicações, e derivação de conclusões. Nesse sentido, a definição de consciência situacional melhor aceita na literatura é a de (ENDSLEY, 1995), que diz que consciência situacional é a percepção de elementos dentro de um volume de tempo e espaço, a compreensão dos seus significados e a projeção dos seus estados em um futuro próximo.

Essa definição de consciência situacional proposta por (ENDSLEY, 1995) é dividida em níveis progressivos, sendo eles:

1. Percepção de elementos dentro de um volume de tempo e espaço (Nível 1): É quando se percebe, ou toma consciência, do estado atual, dos atributos, e das dinâmicas relevantes sobre os elementos do ambiente observado. A incapacidade de alcançar o nível 1 implica na falha em perceber as informações relevantes sobre o ambiente (ou situação) de acordo com os requisitos de informações e os objetivos almejados;

2. Compreensão do significado (Nível 2): É quando se compara as percepções da situação com o entendimento prévio das informações (e suas associações) que estão sendo recebidas, compreendendo a relação dos elementos entre si e com a situação observada. A incapacidade de alcançar o nível 2 implica na falha de entender o que já foi percebido. Quando o nível 2 é alcançado, a consciência do significado intrínsico das informações recebidas e da sua significância no contexto dos objetivos almejados é alcançada;

3. Projeção de estado em um futuro próximo (Nível 3): É quando se é capaz de extrapolar o entendimento das implicações das percepções dentro do ambiente para predizer o que vai acontecer em um futuro próximo, baseado no entendimento das relações de causa e efeito entre os elementos da situação. O nível 3 habilita a antecipação e o planejamento para futuros alternativos e, para alcançar esse nível, um correto desenvolvimento dos níveis 1 e 2 são necessários.

Essa definição progressiva é um fenômeno que ocorre dentro do contexto de tomada de decisão (Figura 3), dado que (ENDSLEY, 1995) entende consciência situacional como sendo orientada a objetivos e possuindo propósito. Portanto, o resultado de um nível 3 criado a partir de uma avaliação acurada da situação de interesse e de uma interpretação racional dentro do contexto dos objetivos propostos, se torna base para a tomada de decisão informada (ENDSLEY; BOLTÉ; JONES, 2003).

Construir e manter uma consciência situacional fica mais difícil a medida em que a complexidade e a dinâmica do ambiente aumentam. Em ambientes dinâmicos, muitas decisões ocorrem em um curto espaço de tempo e as ações a serem executadas são dependentes de uma análise contínua e atualizada do ambiente e situação. Porém, manter essa análise contínua e atualizada é uma tarefa exigente pois ambientes dinâmicos estão em constante mudança e, muitas vezes, essas mudanças são complexas por natureza (ENDSLEY, 1995).

Além desses aspectos, a consciência situacional deve ser interativa, com a compreensão proporcionada fomentando a busca por novas informações e as novas informações coletadas auxiliando na evolução da compreensão, se tornando um processo orientado a dados e objetivos (ENDSLEY; BOLTÉ; JONES, 2003).

Portanto, consciência situacional se refere a muito mais do que simplesmente perceber as informações sobre um ambiente. Ela inclui a compreensão do significado dessas

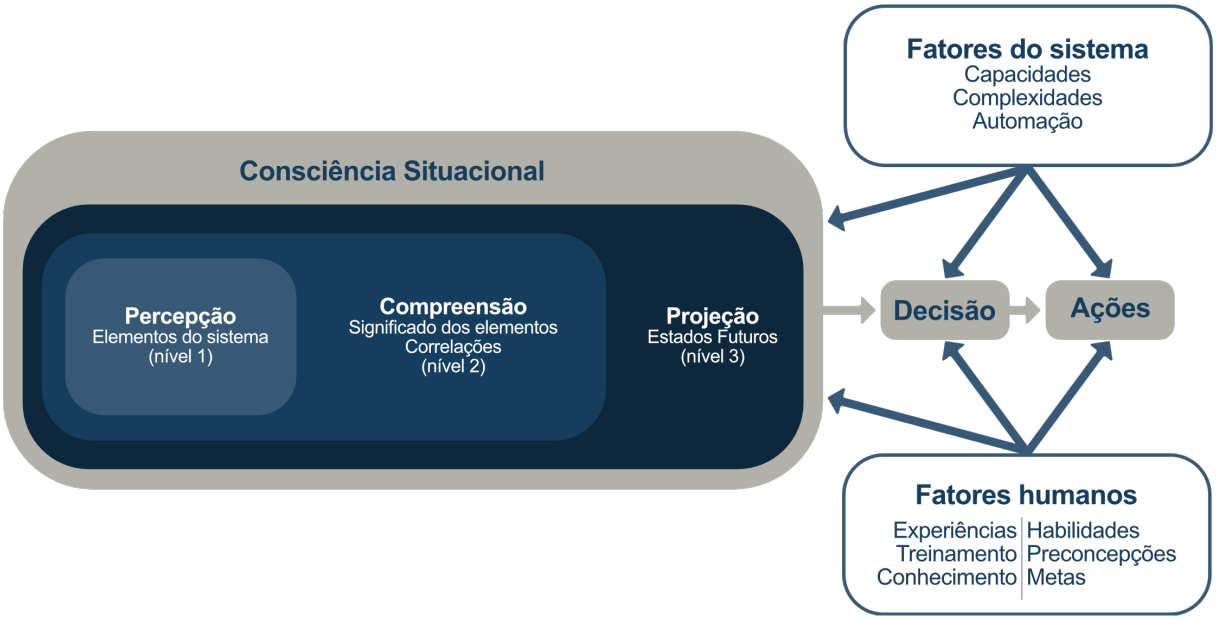

Figura 3 – Modelo de consciência situacional. Adaptado de (SILVA, 2024).

informações de maneira integrada, comparando com os objetivos estabelecidos, e provendo uma projeção futura do estado do ambiente que possua valor para tomada de decisão (ENDSLEY, 1995).

2.1.5 Inteligência de Ameaças

A atividade de inteligência possui a responsabilidade de produção de conhecimento relativo a fatos, eventos, situações ou fenômenos que constituam ou indiquem oportunidades e ameaças aos objetivos de um negócio, organização ou estado. Dessa forma, assinalar e contrapor ameaças a esses objetivos, além de perceber e explorar oportunidades para alcançá-los, é parte integral do trabalho de inteligência (ABIN, 2023).

A nível geral, oportunidades podem ser entendidas como circunstâncias favoráveis a consecução dos objetivos do negócio, organização ou estado, sendo um acontecimento que pode ser aproveitado para impulsionar os seus interesses e garantir vantagens competitivas. Já as ameaças podem ser entendidas como uma oposição antagônica à consecução desses interesses, dificultando, adiando ou impedindo a sua concretização (ABIN, 2023).

Os componentes de uma ameaça são: capacidade e intencionalidade. A capacidade levam em conta, além de outras coisas, as habilidades conhecidas, os recursos logísticos, taxa de sucesso em ataques prévios, sofisticação de ataques prévios, nível de treino, e qualquer outra coisa que é sabida que a ameaça está tentando adquirir, construir ou ter acesso. A capacidade mensurada varia de acordo com a situação e ameaça específica sendo analisada. (STRACHAN-MORRIS, 2012).

Avaliar as capacidades de uma ameaça é fundamental para o trabalho de inteligência,

sendo uma tarefa onde todas as disciplinas de inteligência são normalmente empregadas para coletar informações. Além disso, é comum que uma vasta quantidade de dinheiro seja investida nesse momento, visando entender as capacidades da ameaça e o quão bem elas podem ser utilizadas (STRACHAN-MORRIS, 2012). Porém, quando se trata de avaliar uma ameaça por completo, as capacidades só são uma parte da equação, sendo necessário levar em consideração a intencionalidade da ameaça em usar essas capacidades.

O componente da intencionalidade leva em conta se a ameaça tem o intuito de utilizar suas capacidades, além de como pretende utilizá-las. A intencionalidade pode ser subdivida em vontade e nível de oportunidade. A vontade é uma estimativa do quanto a ameaça quer executar um determinado plano de ação e o nível de oportunidade identifica em quais momentos e locais a ameaça consegue atacar, além de sua capacidade de criar oportunidades para ataque (GILL, 2012).

A análise desses dois componentes provê a avaliação de uma ameaça (*Threat Assessment*). A avaliação de ameaça rigorosa é uma ferramenta importante para tomadores de decisão, pois provê um entendimento da melhor estimativa do plano de ação que uma ameaça pode tomar, auxiliando no planejamento de decisões de budget, além de prover uma base para plano de contingência.

2.1.5.1 Inteligência de Ameaças Cibernéticas

O campo de inteligência de ameaças cibernéticas, do inglês *Cyber Threat Intelligence (CTI)*, é bastante explorado pelo mercado de segurança da informação, muitas vezes pregando que um determinado produto é a solução para estar a frente das ameaças e prevenir ataques cibernéticos. Entretanto, o que se vê na prática é uma dificuldade elevada em construir uma defesa efetiva, com constantes casos de ataques bem-sucedidos causando danos à instituições.

Na literatura, inteligência de ameaças cibernéticas é comumente equiparada aos conceitos de inteligência cibernética e inteligência de ameaças. Entretanto, esses conceitos são distintos:

- Inteligência de ameaças é um conceito bem definido, como organizado na seção 2.1.5. Não sofre tanto com inconsistência de definições;
- Inteligência cibernética possui interpretação ambígua, podendo ser considerada uma disciplina de inteligência em relação ao ambiente cibernético, suportando a tomada de decisão em qualquer domínio, ou inteligência *para* o ambiente cibernético, podendo ser construída a partir de outras disciplinas de inteligência (BONFANTI, 2018);
- Inteligência de ameaças cibernéticas possui diversas definições na literatura, sofrendo com inconsistência no uso do termo (ZIBAK; SAUERWEIN; SIMPSON, 2022).

> Apesar dos diferentes usos e definições dos termos, é compreendido que inteligência de ameaças cibernética é mais específica do que inteligência de ameaças. De forma mais ampla, inteligência de ameaças cibernéticas foca no conhecimento sobre ameaça relativas ao ambiente cibernético e nas ações derivadas desse conhecimento.

A inteligência de ameaças cibernéticas é comumente subdivida em 4 categorias, considerando quem é o consumidor da inteligência e qual é o que se é desejado alcançar (CHISMON; RUKS, 2015):

- Técnica: Se refere a dados e informações consumidos por via técnica, como Indicadores de Comprometimento (IoC) e Indicadores de Ataque (IoA). Em boa parte dos casos pode ser automatizada em infraestruturas de defesa;
- Tática: Se refere a informações sobre como atores de ameaça estão conduzindo os seus ataques. É normalmente consumida por profissionais técnicos responsáveis por defender, alertar, investigar e preparar o ambiente contra essas táticas;
- Operacional: Normalmente se refere a ataques iminentes a instituição e é inicialmente consumida por funcionários de segurança de nível hierárquico mais alto, como gestores de segurança e responsáveis por resposta a incidente;
- Estratégica: Normalmente cobre tópicos como impacto financeiro de atividades cibernéticas, tendências de ataque, e áreas que podem impactar decisões de alto nível da instituição.

As maiores contribuições da área de inteligência de ameaças cibernéticas com relação a defesa de ataques cibernéticos são o entendimento de quais ameaças são relevantes para a instituição, suas capacidades e intencionalidade, e o oferecimento de cursos de ação que aprimorem a tomada de decisão.

2.1.5.2 Atores de Ameaça Cibernética

Em um nível mais amplo, um ator de ameaça é entendido como um indivíduo, grupo ou organização representando uma ameaça (JOHNSON *et al.*, 2016). Portanto, um ator de ameaça cibernética pode ser entendido como um ator de ameaça que representa uma ameaça especializada cibernética.

O entendimento da intencionalidade das ameaças cibernéticas envolve a análise da motivação do seu respectivo ator, pois ela indica a intensidade e persistência dos possíveis ataques. Os tipos de motivações de atores de ameaça cibernética são comumente entendidas como (STANDARD, 2021):

- Acidental: Um ator não hostil cuja intenção benevolente ou inofensiva inadvertidamente causa danos;

- Coerção: Forçado a agir em nome de outra pessoa ou ator;
- Dominância: Desejo de afirmar superioridade sobre alguém ou alguma coisa;
- Ideologia: Paixão por expressar um conjunto de ideias, crenças e valores que podem resultar em atos prejudiciais e ilegais;
- Notoriedade: Buscar prestígio ou se tornar conhecido através de alguma atividade;
- Ganho organizacional: Buscar vantagem sobre uma organização concorrente;
- Ganho pessoal: Desejo de melhorar a própria situação financeira;
- Satisfação pessoal: Desejo de satisfazer um objetivo estritamente pessoal, como: curiosidade, busca de emoção, diversão, ou outros;
- Vingança: Desejo de vingar erros percebidos através de ações prejudiciais, como: sabotagem, violência, roubo, fraude ou constrangimento;
- Imprevisível: Ação sem razão ou propósito identificável, criando eventos imprevisíveis.

Além da motivação, existem outros elementos técnicos importantes de serem considerados na construção de contexto para análise sobre um ator de ameaça cibernético, alguns desses elementos são: caracterização de sua identidade, suspeita de qual é seu resultado pretendido, suas relações com outros atores de ameaça, e suas táticas, técnicas e procedimentos de ataque (STANDARD, 2021).

2.1.6 Deep Web e Dark Web

A internet é uma realidade que possibilitou a formação de uma sociedade interconectada digitalmente, transcendendo limites geográficos. O impacto positivo causado pela internet é muito difícil de ser mensurado em sua completude. Porém, apesar dos impactos positivos, ainda é possível observar que muitos usuários possuem uma visão de que a internet é uma "terra sem lei", onde a impunibilidade é a norma e que suas ações são anônimas, ou seja, que suas ações no ambiente digital não podem ser rastreadas e relacionadas a identidade física do usuário.

Entretanto, os usuários da internet são legalmente vinculados aos seus respectivos países e é possível exercer punições legais por crimes cometidos na internet porque identidades digitais podem ser relacionadas aos indivíduos ou sites que possuem elas. Dessa forma, por meio do rastreamento de ações realizadas no ambiente digital para o real, que as forças da lei conseguem realizar a atribuição de ações conduzidas online. Esses conceitos de anônimato e rastreabilidade são importantes para o entendimento das três categorias da internet em relação a anonimato e rastreabilidade: Web superficial (Surface Web), Web profunda (Deep Web) e Web obscura (Dark Web) (GUPTA; MAYNARD; AHMAD, 2021).

A web superficial (Surface Web) é a parte pública da internet, onde o acesso não é restrito por autenticação ou pagamento, e é indexada por buscadores, como o Google e o Bing. Além disso, todas as partes podem ser identificadas e, portanto, podem ser responsabilizadas legalmente.

A web profunda (Deep Web) é a parte da internet que não é acessível publicamente, ou seja, possui acesso privado, e não é indexada por buscadores. O acesso a essa parte é restrito pela necessidade de autenticação ou por ser parte de uma rede interna. Devido aos requisitos extras para acesso, essa parte costuma fornecer uma capacidade ainda melhor de identificar pessoas e sistemas, aumentando a capacidade de responsabilização legal.

A web obscura (Dark Web) é a parte da internet que não é indexada por buscadores e precisa de softwares especializados para ser acessada. Ela possui tanto elementos públicos como privados desde que o software correto seja utilizado.

A principal diferença entre a Deep Web e a Dark Web é o aspecto de responsabilização, pois os usuários da Dark Web não são identificáveis para a rede ou qualquer pessoa monitorando e, portanto, suas ações são anonimizadas. Além disso, a Dark Web possibilita a hospedagem de serviços web escondidos (*hidden services*) que mantém seus endereços IP reais anônimos e, portanto, sua localização, até mesmo para os usuários que usam esses serviços web. Por conceder anonimidade, o engajamento privado entre pessoas foi institucionalizado na Dark Web. (PELTON; SINGH, 2019)

2.1.7 Fóruns de Mensagens Criminosas

É sabido que a Dark Web é utilizada como uma forma de comercialização digital e ilegal de drogas, armas de fogo, e material relacionado a pedofilia, como demonstrado em (ALDRIDGE; DÉCARY-HÉTU, 2016) e (LIGGETT *et al.*, 2020). Além disso, é comum que a Dark Web seja utilizada para comercialização de crimes cibernéticos ou de seus resultados, como credenciais roubadas, cartões de crédito, malwares, etc. (HOLT, 2013; DUPONT *et al.*, 2017).

Apesar da anonimidade que a Dark Web pode oferecer em ações online, isso tem um custo de expertise para saber entrar e navegas. No Brasil, é comum observarmos que muitas das atividades criminosas acontecem na Deep Web, em fóruns de mensagens em plataformas como Facebook, Telegram e WhatsApp.

A execução de crimes cibernéticos do início ao fim não é simples, envolvendo diversas etapas que requerem conhecimento especializado. A execução de um ataque cibernético em sua totalidade consome bastante tempo e expõe o atacante a um grau de risco muito elevado, pois exige conhecimento específico para executar as diferentes etapas do ataque enquanto protege sua identidade digital.

Considerando uma fraude digital simples, como a compra de um item através de

um cartão de crédito de um terceiro desconhecido, por exemplo, a execução do ataque por completo exige:

1. Identificar as informações de um cartão de crédito válido (número do cartão, data de validade e código de segurança);
2. Burlar (*bypass*) os mecanismos de prevenção a fraude implantados pelo e-commerce alvo;
3. Escapar (*bypass*) dos mecanismos de prevenção a fraude da instituição bancária responsável pela oferta de crédito (avaliação de padrões de consumo, proximidade geográfica, etc);
4. Receber o produto comprado em uma localização que evite conexão à identidade do criminoso, visando dificultar possíveis investigações policiais posteriores;

Todas essas etapas exigem um grau de especialidade diferente. Por exemplo, existem diversas formas de identificar informações de um cartão de crédito válido (etapa 1). Algumas delas são:

- Enumeração por força bruta: comumente denominado "*gerada*"dentro do ecosistema criminoso, é quando fraudadores buscam gerar informações válidas sobre um cartão de crédito a partir de uma informação previamente conhecida (BIN de uma instituição financeira, por exemplo). O programa gerador busca criar diversas combinações diferentes de dados de cartão, eventualmente encontrando uma combinação válida;
- Phishing de cartão de crédito: é quando fraudadores criam páginas falsas de compras com o objetivo de enganar clientes legítimos a entregarem os dados de seus cartões de créditos aos fraudadores;
- Comprometimento de terminais de pagamento: é quando atores de ameaça cibernética comprometem terminais de pagamento de lojas e implantam malwares que coletam as informações dos cartões de crédito.

Cada forma listada de identificar informações válidas de um cartão de crédito exige uma especialização diferente. Um atacante que não domina essa especialização pode se expor a riscos desconhecidos de comprometimento de sua identidade. Além disso, caso sua identidade seja identificada por forças da lei, o atacante pode ser indiciado por mais crimes, visto que executou diversas ações de cunho criminoso diferentes.

Dessa forma, é comum que o ecosistema criminoso seja repleto de atores de ameaça vendendo e comprando serviços e informações diferentes, executando somente os crimes de sua especialidade. Dentro do exemplo de cartões de crédito, a Cybersixgill reporta que a

quantidade de cartões comprometidos está em decadência nos últimos anos mas os serviços de venda de cartões ainda possuem um alto grau de atividade (CYBERSIXGILL, 2024).

Um dos aspectos mais importantes no ecossistema de crimes cibernéticos são as plataformas escolhidas para comunicação e colaboração entre atores de ameaça. Nessas plataformas, os atores de ameaçam trocam ferramentas, dados roubados e serviços ilícitos de forma paralela e equivalente a *dark web* tradicional. Entre as plataformas mais utilizadas atualmente, o Telegram e o Discord figuram em destaque (CYBERSIXGILL, 2024).

2.2 Mineração de Dados

Mineração de dados compreende os principais algoritmos que possibilitam o ganho de percepções e conhecimentos fundamentais em grande quantidade de dados. É um campo interdisciplinar agregando conceitos de áreas aliadas, como sistemas de banco de dados, estatística, aprendizado de máquina, e reconhecimento de padrões (ZAKI; MEIRA, 2014). Como resultado da interdisciplinaridade, mineração de dados possui diferentes definições de acordo com o campo de atuação dos autores. Existem três áreas que são consideradas de maior expressão em relação a contribuição com mineração de dados: Estatística, Banco de Dados, e Aprendizado de Máquina (ZHOU, 2003).

De uma perspectiva estatística, mineração de dados pode ser entendida como a análise de grandes conjuntos de dados observacionais a fim de encontrar relacionamentos inesperados e de resumir os dados de uma forma que eles sejam tanto úteis quanto compreensíveis ao dono dos dados (HAND; MANNILA; SMYTH, 2001).

De uma perspectiva de banco de dados, mineração de dados pode ser entendida como um passo no processo de Descoberta de Conhecimento em Banco de Dados que consiste na realização da análise dos dados e na aplicação de algoritmos de descoberta que, sob limitações computacionais aceitáveis, produzem uma enumeração de padrões dos dados (FAYYAD; PIATETSKY-SHAPIRO; SMYTH, 1996).

Além dessas definições, existem outros entendimentos mais genéricos, como: “Mineração de dados é o estudo de coleta, limpeza, processamento, análise e obtenção de percepções úteis a partir de dados“ (AGGARWAL *et al.*, 2015).

2.2.1 Processos de Mineração de Dados

O processo de mineração de dados é um processo criativo que requer diversas habilidades e conhecimentos distintos. Por muito tempo, o sucesso de uma iniciativa de mineração de dados era muito dependente das habilidades do profissional, sendo incerta a possibilidade de replicar esse sucesso em outras esferas da instituição. Devido a estes fatores, processos para execução estruturada de iniciativas de mineração de dados foram propostas, como o CRISP-DM (WIRTH; HIPP, 2000).

Três processos bastante conhecidos que permeiam este tópico são: Descoberta de conhecimento em banco de dados, do inglês *Knowledge Discovery in Databases* (*KDD*); SEMMA, do inglês *Sample, Explore, Modify, Model, and Assess*; e CRISP-DM, do inglês *Cross Industry Standard Process for Data Mining*. Apesar destes 3 processos permearem o mesmo tópico, existem algumas distinções importantes:

- No processo KDD, mineração de dados é uma das etapas a serem executadas como parte do processo (FAYYAD; PIATETSKY-SHAPIRO; SMYTH, 1996). Por esse ponto de vista, KDD pode ser considerado mais abrangente, limitando mineração de dados a atividade de identificar padrões em dados já pré-processados;

- Tanto SEMMA quanto CRISP-DM podem ser considerados uma implementação do processo de KDD, sendo CRISP-DM mais completo que o SEMMA, apesar do SEMMA poder ser equiparado caso pré-requisitos subentendidos sejam considerados (AZEVEDO; SANTOS, 2008).

Apesar dos diversos processos, KDD costuma ser mais utilizado por pesquisadores e profissionais mais próximos a área de banco de dados, enquanto CRISP-DM é o processo mais considerado por outras áreas de atuação.

2.2.1.1 CRISP-DM

O processo CRISP-DM busca fazer com que projetos grandes de mineração de dados sejam menos custosos, mais confiáveis, mais rápidos, e tenham maior capacidade de reprodução (WIRTH; HIPP, 2000). O processo consiste de 6 etapas maiores e flexíveis quanto a ordem de execução, sendo previsto ciclos de iteração.

Normalmente, a iteração durante o processo é a norma e não a exceção. Percorrer todo o processo sem solucionar o problema não é (geralmente) considerado um falha, dado que todo o processo pode ser usado para guiar a exploração do problema, alcançando entendimento importante e possibilitando uma segunda execução melhor informada (PROVOST; FAWCETT, 2013).

Dessa forma, o processo CRISP-DM pode ser observado na figura [CITAR FIGURA] e suas etapas são detalhadas a seguir (PROVOST; FAWCETT, 2013):

1. Entendimento do negócio: entender o negócio é importante para compreender o problema a ser solucionado e, normalmente, reformular o problema e desenhar uma solução é um processo iterativo de descoberta. A formulação inicial do problema pode não ser completa ou ótima, o que requer múltiplas execuções para uma solução aceitável ser encontrada;

2. Entendimento dos dados: dados são as matérias primas com as quais a solução para o problema de negócio será criada. Portanto, é importante compreender os pontos fortes e as limitações dos dados pois raramente estão adequados a resolução do problema, podendo ser sobre populações diferentes da desejada e possuir níveis de confiabilidade variados. Uma tarefa importante do entendimento dos dados é estimar os custos e benefícios de cada fonte de dados e decidir aonde é válido investir;

3. Preparação dos dados: as tecnologias que serão utilizadas para análise impões restrições e requisitos com relação aos dados que elas conseguem processar, podendo ser um formato de dado específico diferente de como os dados estão disponíveis. Portanto, a etapa de preparação caminha em conjunto com a etapa de entendimento dos dados, manipulando e convertendo dados de forma a alcançar melhores resultados. Alguns exemplos de preparação são: inferir valores em falta, converter para formato tabular, e normalizar valores numéricos;

4. Modelagem: é a primeira etapa onde técnicas de mineração de dados são efetivamente aplicadas aos dados. Nessa etapa, o conhecimento das ideias fundamentais de mineração de dados é fundamental, incluindo técnicas e algoritmos existentes, pois é a etapa onde a maior parte das tecnologias e conhecimento científico podem ser colocados em uso;

5. Avaliação: o objetivo desta etapa é avaliar rigorosamente os resultados da mineração de dados e ganhar confiança de que eles são válidos e confiáveis. É possível realizar a implantação dos modelos diretamente, sem muito escrutínio, mas isso é perigoso pois os resultados alcançados podem não representar padrões de dados relevantes e confiáveis. Além disso, é importante avaliar se o modelo satisfaz os requisitos de negócio previamente estabelecidos;

6. Implantação: os resultados da mineração de dados e as técnicas aplicadas são colocadas em uso para gerar retorno de investimento. Normalmente se busca realizar a implantação das técnicas aplicadas, não somente o modelo gerado, pois o ambiente de atuação do modelo pode evoluir de forma muita rápida ao ponto em que os responsáveis pela mineração de dados podem não ser capazes de adaptar a solução em tempo hábil.

2.2.1.2 Problemas Fundamentais de Mineração de Dados

Apesar da extensiva lista de algoritmos disponíveis atualmente, a quantidade de tarefas (ou problemas) fundamentalmente diferentes que esses algoritmos endereçam é pequena. Alguns autores argumentam que existem 4 problemas fundamentais resolvidos por mineração de dados: agrupamento, classificação, associação de padrões e análise de *outliers* (AGGARWAL *et al.*, 2015). Outros autores, por sua vez, sugerem que são 8 problemas

fundamentais (PROVOST; FAWCETT, 2013). Para fins de completude, discutiremos a abordagem de 8 problemas fundamentais, sendo eles:

- Classificação e probabilidade de classe: Buscam predizer a qual classe (dentro de um conjunto pequeno) um determinado indivíduo de uma população pertence. A tarefa de classificação envolve um modelo que, dado um indivíduo, determina a qual classe ele pertence. A estimativa de probabilidade de classe é uma tarefa muito relacionada a classificação, porém o resultado do modelo aplicado a um indivíduo é a probabilidade de que ele pertença a cada uma das classes. É comum que um modelo que consiga realizar uma dessas tarefas também consiga realizar a outra com alguns ajustes;

- Regressão: Busca estimar ou predizer o valor numérico de uma variável para o indivíduo. Dado o aspecto numérico da regressão, seu resultado é capaz de especificar o quanto algo é predito, diferente da classificação que só é capaz de responder o que é predito;

- Similaridade: busca identificar indivíduos similares baseado em informações conhecidas sobre eles. As medidas de similaridade são fundamentalmente importantes para outras tarefas de mineração de dados, como classificação, regressão e agrupamento;

- Agrupamento: busca agrupar os indivíduos de uma população por sua similaridade sem um direcionamento de quais grupos existem ou quais suas características. Agrupamento costuma ser muito útil na exploracão inicial de um domínio buscando encontrar os grupos naturais existentes, pois podem fornecer informações importantes para outras abordagens de mineração de dados;

- Regras de associação / Agrupamento de coocorrência: buscam encontrar associações entre entidades baseado nas transações envolvendo essas entidades. Enquanto agrupamento busca agrupar objetos baseado na similaridade de suas características, a análise de coocorrência busca agrupar considerando a similaridade da aparição em conjunto em transações. O resultado da análise de coocorrência é a descrição de itens que acontecem juntos, incluindo estatística das frequências e estimativas de quão surpreendente a relação é;

- Criação de perfil e análise de comportamento: Busca caracterizar o comportamento típico indivíduos, grupos ou população. O comportamento analisado é definido de acordo com o contexto, sendo a análise do perfil de clientes de telefonia diferente da análise do comportamento financeiro de clientes bancários para identificar fraudes, por exemplo;

- Predição de conexões: Busca predizer conexões entre dados e uma estimativa da força dessa conexão. Normalmente é usada para sugerindo uma conexão que deveria existir mas ainda não existe, podendo ser interpretada como uma recomendação;

- Redução de dados ou dimensionalidade: Busca transformar um conjunto de dados grande e transformá-lo em um menor, contendo boa parte da informação importante do conjunto maior. Costuma ser usada quando o conjunto de dados é muito grande para ser utilizado eficientemente. Normalmente o conjunto menor sofre com perda de informação;

- Análise de causalidade: Busca auxiliar no entendimento de quais eventos influenciam em outros.

2.2.2 Aprendizado de Máquina

Aprendizado de máquina, do inglês *Machine Learning*, é um campo da ciência da computação que estuda algoritmos e técnicas para automatizar soluções para problemas complexos que são difíceis de serem programados através de métodos convencionais (REBALA *et al.*, 2019). Uma outra definição bastante referenciada na academia, como em (MAHESH, 2020) e (ALZUBI; NAYYAR; KUMAR, 2018), incorpora o conceito de "aprendizado", entendendo aprendizado de máquina como o campo de estudo que busca dar aos computadores a habilidade de aprender a realizar uma tarefa sem ser explicitamente programado para ela. Essa segunda definição é comumente atribuída a (SAMUEL, 1959), mas nem o artigo referenciado nem sua publicação subsequente (SAMUEL, 1967) apresentam essa definição.

Independente da autoria da definição, a incorporação de aprendizado é útil para o entendimento de aprendizado de máquina. Dessa forma, é dito que um programa de computador aprende através de experiência E a realizar uma tarefa T com uma medida de performance P, se a sua performance em T, medida por P, melhora com a experiência E (MITCHELL, 1997).

O conceito de aprendizado é dividido em 3 categorias: aprendizado por instrução, onde o professor ensina o aluno; por analogia, onde um conceito é entendido através da transformação e expansão de conhecimento prévio que apresenta forte semelhança ao novo conceito; e por exemplos, onde o aluno induz uma descrição geral do conceito a partir de um conjunto de exemplos deste conceito. O aprendizado por exemplos é o mais relevante para a área de aprendizado de máquina e considerado essencial (LAMPROPOULOS; TSIHRINTZIS, 2015).

Além disso, o aprendizado por exemplos é discriminado em 3 categorias de acordo com a forma em que o conjunto de exemplos é utilizado (LAMPROPOULOS; TSIHRINTZIS, 2015):

- Aprendizado supervisionado: os exemplos relacionados a um conceito específico são caracterizados por pares de entrada e saída (rótulo). Ou seja, os dados que pertencem ao mesmo conceito já estão associados ao seu respectivo valor alvo;

- Aprendizado não supervisionado: se trata de encontrar uma descrição concisa dos dados através de agrupamento ou mapeamento passivo dos dados de acordo com algum princípio de ordem;

- Aprendizado por reforço: é quando um agente aprende por tentativa e erro a executar uma ação para receber uma recompensa. Normalmente este aprendizado está relacionado a problemas onde o agente de aprendizagem não não sabe a priori o que deve fazer.

Um subtipo de aprendizado supervisionado é o aprendizado semi-supervisionado, onde um pequeno conjunto de rótulos é utilizado em conjunto de uma grande quantidade de dados não rotulados. Essa técnica costuma gerar resultados melhores que o aprendizado não supervisionado mas sem requerer todo o investimento necessário para rotular todos os dados (LEARNING, 2006).

Por fim, o campo de aprendizado de máquina não deve ser confundido com inteligência artificial, sendo aprendizado de máquina uma subárea da inteligência artificial. Inteligência artificial é sobre tornar máquinas inteligentes utilizando diversas abordagens distintas, enquanto aprendizado de máquina é essencialmente uma dessas abordagens possíveis. Essa distinção é importante pois estes termos são observados sendo utilizados de forma equivalente por causa da crença de que aprendizado de máquina é a única forma viável de alcançar os objetivos da inteligência artificial (REBALA *et al.*, 2019).

2.3 Processamento de Linguagem Natural

Processamento de Linguagem Natural (PLN), do inglês *Natural Language Processing* (NLP), é uma área de pesquisa e aplicação que explora como computadores podem ser utilizados para entender e manipular textos ou falas (CHOWDHARY; CHOWDHARY, 2020).

A área de PLN iniciou por volta dos anos de 1950 como uma intersecção da interligência artificial com a linguística, começando como uma área distinta da área de recuperação de informação (que aplica técnicas escaláveis baseadas de estatística para indexar e procurar grandes volumes de texto eficientemente) mas convergindo posteriormente (NADKARNI; OHNO-MACHADO; CHAPMAN, 2011). Atualmente, a PLN é fundamentada por diversas áreas de conhecimento, como: ciência da informação, linguística, matemática, inteligência artificial, psicologia, recuperação de informação, entre outras (CHOWDHARY; CHOWDHARY, 2020).

A aplicação de técnicas de PLN abrangem um grande número de campos de estudo, como: tradução automática, processamento de linguagem natural e sumarização, recuperação de informação em textos em línguas diferetnes (multilíngua), reconhecimento de fala, sistemas especialistas, entre outros (CHOWDHARY; CHOWDHARY, 2020).

O processamento de linguagem natural envolve a análise de diversos níveis interdependentes dos quais as pessoas extraem significado: nível fonológico, morfológico, léxico, síntatico, semântico, discurso, e pragmático (LIDDY, 1998).

2.3.1 Representação Textual em Formato Vetorial

Lidar diretamente com dados textuais é algo complexo para computadores, dado que operam exclusivamente com números. Portanto, técnicas que consigam representar textos através de números são imprescindíveis no campo de processamento de linguagem natural.

Uma das técnicas mais simples de realizar essa transformação é a contagem de ocorrência de palavras em um texto, transformando o texto em um vetor de dimensão igual a quantidade de palavras (vocabulário) presente no texto. Essa técnica é denominada saco de palavras (do inglês *bag-of-words*) e possui as características de ser fácil de construir, gerar um vetor sem considerar a ordem das palavras, e não considerar a semântica ou gramática do texto. Portanto, apesar de ser uma técnica bastante utilizada devido a sua baixa necessidade de poder computacional, é considerada um técnica limitada na capacidade de representar os dados textuais.

Uma das limitações mais relevantes deste tipo de representação baseado na contagem de ocorrência de palavras (saco de palavras) é que as relações semânticas entre as palavras são desprezadas. Como essas representações não levam em conta o contexto das palavras na sentença, a técnica de saco de palavras pode falhar em representar o texto de forma acurada (GROOTENDORST, 2022).

Como resposta a esse tipo de problema, técnicas de *embeddings* de texto se tornaram populares no campo de processamento de linguagem natural. Dessa forma, outras técnicas de representação textual foram desenvolvidas na tentativa de capturar informações semânticas, incorporando o conceito de *embeddings* textuais que representam textos e palavras similares de forma próxima no espaço vetorial. Uma dessas técnicas é a word2vec (MIKOLOV, 2013), proposta pela Google.

A técnica de word2vec objetiva representar uma palava em um vetor multidimensional composto de números que capturam as relações entre as palavras. Ou seja, as palavras que aparecem em contextos similares são mapeadas para vetores que estão próximos um dos outros através da medida de similaridade de cosseno (MIKOLOV, 2013).

Posteriormente, outra abordagem competindo com a técnica word2vec foi proposta,

denominada GloVe. O modelo GloVe (*Global Vectors*) utiliza uma abordagem não supervisionada para obter as representações vetoriais das palavras. Essa representação é construída através do mapeamento de palavras em um espaço vetorial onde a distância entre palavras é relacionado as suas similaridades semânticas (PENNINGTON; SOCHER; MANNING, 2014).

Entretanto, as abordagens visando representar palavras em espaço vetorial são consideradas ultrapassadas atualmente devido a evolução dos modelos de linguagem baseados em *transformers*. A arquitetura de redes neurais profundas *transformer* foi proposta por pesquisadores da Google, e visa representar o texto em um espaço vetorial através da transformação de tokens em vetores, seguido da contextualização dentro do escopo da janela de contexto. Esse processo visa ampliar a importância de tokens relevantes enquanto diminui a importância de tokens menos relevantes (VASWANI, 2017).

Dessa forma, o *Bidirectional Encoder Representations from Transformers* (BERT), baseado na arquitetura *transformer*, e suas variantes demonstraram ótimos resultados na geração de representações vetoriais bem contextualizadas de sentenças e palavras. As propriedades semânticas dessas representações vetoriais permitem encapsular o significado dos textos de forma que textos similares fiquem próximos no espaço vetorial (GROOTENDORST, 2022).

2.3.2 Modelagem de Tópicos

Algoritmos de modelagem de tópicos analisam as palavras dos textos originais para descobrir os temas que os permeiam, como esses temas estão conectados entre si, e como eles mudam ao longo do tempo (BLEI, 2012). Esses algoritmos costumavam estar limitados a modelos estatísticos baseados na frequência de palavras em cada texto, mas recentemente outras abordagens de modelagem de tópicos estão ganhando popularidade, como o BERTopic (GROOTENDORST, 2022), que reduzem o problema de modelagem de tópicos a um problema de agrupamento de textos (SIA; DALMIA; MIELKE, 2020).

A modelagem de tópicos é comumente utilizada como uma etapa exploratória da análise de documentos, visando encontrar os principais temas e narrativas discutidas no *corpus*. Essa busca pelos temas discutidos pode nascer de diversas motivações, como: encontrar os principais documentos tratando de um tópico específico de interesse, identificar se um determinado tema foi abordado no *corpus*, identificar as diferentes maneiras em que um tópico é discutido, identificar a proporção de documentos que discutem sobre um tópico específico, entre outros casos de uso (BOYD-GRABER *et al.*, 2017).

Como a modelagem de tópicos é utilizada por seres humanos para explorar e avaliar os tópicos discutidos em um conjunto de documentos, é necessário que essas técnicas sejam capazes de representar os tópicos identificados em um formato interpretável. Isso é normalmente realizado através da identificação das principais palavras de cada tema e

seus respectivos pesos (BOYD-GRABER *et al.*, 2017). Entretanto, não há garantia de que as palavras identificadas são coerentes com as expectativas da pessoa analisando, ou seja, os tópicos encontrados podem ser modelados de forma que são inúteis para a pessoa analisando dado que o conjunto de palavras-chave daquele tópico não são fáceis de serem interpretadas dentro do contexto esperado. Portanto, as técnicas de modelagem de tópicos são normalmente aplicadas de forma iterativa, onde cada iteração realizada pela pessoa analisando busca refinar e definir melhor os tópicos de acordo com a interpretabilidade esperada.

As subseções a seguir descrevem as principais técnicas de modelagem de tópicos utilizadas por este trabalho.

2.3.2.1 Latent Semantic Analysis

Análise semântica latente, do inglês *Latent Semantic Analysis* (LSA), é um método para extrair e representar o significado contextual de palavras através da aplicação de técnicas estatísticas em um grande conjunto de dados textuais. A ideia central é que o agregado de todos os contextos em que uma palavra aparece ou deixa de aparecer fornece um conjunto de restrições mútuas que determina fortemente a similaridade de significado entre palavras e conjunto de palavras se comparados entre si (LANDAUER; FOLTZ; LAHAM, 1998).

Na prática, a técnica LSA é implementada através da criação de uma matriz de contagem de palavras por documento. Essa matriz é submetida a decomposição em valores singulares (*Single Value Decomposition* - SVD), reduzindo a dimensionalidade da contagem de palavras mas preservando sua escala de similaridade. Por fim, os documentos na nova matriz reduzida são comparados através da similaridade de cosseno, que varia de 1 (muito similar) a 0 (muito dissimilar).

2.3.2.2 Latent Dirichlet Allocation

Alocação latente de Dirichlet, do inglês *Latent Dirichlet Allocation* (LDA), é um modelo generativo estatístico que atua sobre um conjunto de documentos e busca capturar a intuição de que cada documento apresenta múltiplos tópicos, e que cada tópico é composto de múltiplas palavras (BLEI, 2012).

Para o modelo LDA, um tópico é formalmente entendido como uma distribuição sobre um vocabulário fixo, assumindo que os tópicos são especificados antes de qualquer dado ser gerado. Dessa forma, as palavras que compõem um determinado tópico são geradas através do seguinte processo (BLEI, 2012):

- Para cada documento no conjunto de documentos:

    1. Uma distribuição sobre os tópicos é escolhida ao acaso;

2. Para cada palavra no documento:
    a) Um tópico da distribuição sobre tópicos do passo 1 é escolhido ao acaso;
    b) Uma palavra da distribuição correspondente sobre o vocabulário é escolhida ao acaso.

Portanto, cada documento apresenta os tópicos em uma proporção diferente (etapa 1) e cada palavra em cada documento é puxada de um dos tópicos (etapa 2b), sendo que esse tópico é escolhido da distribuição de tópicos por documento (etapa 2a). Essa é a característica que diferencia a técnica LDA: todos os documentos compartilham o mesmo conjunto de tópicos, mas cada documento apresenta esses tópicos em diferentes proporções (BLEI, 2012).

2.3.2.3 BERTopic

BERTopic é uma abordagem de modelagem de tópicos que utiliza de *embeddings* textuais gerados por modelos pré-treinados BERT a fim de representar os documentos analisados de forma contextualizada. Essa abordagem é composta das seguintes etapas (GROOTENDORST, 2022):

- Geração de *embeddings*: modelos BERT pré-treinados são utilizados para representar os textos em um espaço vetorial de forma contextualizada, onde documentos similares ficam próximos uns aos outros;
- Redução de dimensionalidade: a técnica UMAP (*Uniform Manifold Approximation and Projection*) é utilizada para reduzir a dimensionalidade dos *embeddings* textuais, buscando melhorar a performance da etapa de agrupamento;
- Agrupamento: o agrupamento por similaridade de documentos é realizado através da técnica de agrupamento HDBSCAN (*Hierarchical Density-Based Spatial Clustering of Applications with Noise*), que realiza o agrupamento baseado na densidade dos grupos;
- Representação dos tópicos: após o agrupamento, os documentos pertencentes ao mesmo grupo são agregados e submetidos a técnica de TF-IDF (*Term Frequency – Inverse Document Frequency*). Dessa forma, ao invés de buscar a representação das palavras mais relevantes para cada documento, esta técnica visa encontrar as palavras mais relevantes para cada grupo, sendo denominada c-TF-IDF (*class TF-IDF*).

Devido a forma como a técnica BERTopic aborda o problema de modelagem de tópicos (agrupamento de documentos similares seguido de representação de palavras mais importantes), existe a limitação que cada documento pertence exclusivamente a um tópico,

diferente de outras técnicas estatísticas que assumem uma distribuição de tópicos em cada documento (como a LDA) (GROOTENDORST, 2022).

Nos experimentos realizados no artigo de publicação da técnica BERTopic, foi identificado que ela fornece resultados melhores do que as técnicas estatísticas tradicionais (como a LDA) utilizando métricas de coerência de tópico (GROOTENDORST, 2022). Entretanto, as soluções de modelagem de tópicos possuem restrições subjetivas com relação a qualidade final, como a necessidade de interpretabilidade dos tópicos por analistas humanos, requerendo uma análise específica para cada caso se esta técnica é viável. Além disso, BERTopic é altamente dependente de *embeddings* textuais com boa representação do texto, o que pode gerar resultados ruins caso os modelos de linguagem utilizados não consigam gerar boas representações dos documentos analisados.

2.4 Trabalhos Relacionados

O trabalho de (SUFI, 2023) realizou coleta massiva de tweets através de busca por palava-chave na API ("cyber"). Posteriormente a coleta, processou os dados identificando a linguagem usada, traduzindo para o inglês, e filtrando mais com termos adicionais (como: "Brazil"para especifica o país). Após esses passos, segmentou os tipos de índice que gostaria de fazer através de mais filtros de palavras, normalizou a frequência desses dados (e.g. count), e aplicou uma abordagem baseada em CNN para computar o nível de ameaça (threat level).

Ano	Autor	Fonte de Dados	Técnicas de Análise	Objetivo do Estudo
2023	(SUFI, 2023)	Twitter	Palavras-chave e redes neurais convolucionais.	Gerar índices de nível de ameaça cibernética (*Threat Level*).
2024	(DEVARAJAN *et al.*, 2024)	Anúncios de venda em mercados na Dark Web	Frequência de termos, regras de associação, SVM e redes neurais.	Agrupamento de produtos e vendedores similares. Identificação dos principais termos usados em conjunto nos anúncios.
2020	(SAMTANI; ZHU; CHEN, 2020)	Fóruns de mensagens na Dark Web	*Diachronic Graph Embedding*	Identificação de ameaças emergentes e tendências baseado em popularidade de termos.
2020	(SHAH *et al.*, 2020)	Grupos do Telegram	Informed Bag of Words	Classificação do grupo de origem de um texto, distinguir entre *bot* e humano, identificar anúncios de venda ilegal, distinguir entre texto de Telegram e de Twitter, e identificar as diferenças entre mensagens lícitas de ilícitas.

Tabela 3 – Estudos sobre extração de características de atores de ameaças através de PLN.

O trabalho de (DEVARAJAN *et al.*, 2024) realiza a coleta de informações de vendas ilícitas em marketplaces da dark web (drogas, armas, tutoriais, ouro, cartões, contrabando, etc), processa o título do post de venda (lowercase, removendo símbolos e stopwords), e utiliza TF-IDF (talvez só TF) em conjunto com regras de associação como entrada para um modelo de classificação utilizando SVM e redes neurais.

O trabalho proposto por (SAMTANI; ZHU; CHEN, 2020) buscou organizar textos coletados de fóruns de mensagens na Dark Web em forma de grafo, para manter o relacionamento de contexto entre as palavras. Geraram embeddings através dos grafos e utilizaram ténicas diachronic (LDA?) para identificar ameaças emergentes em termos de popularidade e funcionalidade. Após isso, o trabalho propõe 3 experimentos diferentes: Analogia semântica (semantic analogy), agrupamento, classificação da plataforma alvo atacada, e classificação do tipo do ataque. Os resultados demonstram que a abordagem de grafos pode ser bem interessante ao invés dos métodos tradicionais, como word2vec. O trabalho dos autores é bastante promissor e pode ser uma base muito interessante para o que estou pretendendo fazer.

O trabalho proposto por (SHAH *et al.*, 2020) buscou classificar quais mensagens dentro do contexto de discussão de grupos do Telegram eram maliciosas. Os atores coletaram manualmente o histórico de diversos grupos cuidadosamente selecionados e aplicaram técnicas de bag of words para alcançar resultados de classificação.

3 CONSCIÊNCIA SITUACIONAL DO ECOSSISTEMA DE CRIMES CIBERNÉTICOS NO BRASIL

A criação e manutenção de uma consciência situacional sobre um determinado tópico de interesse é totalmente dependente do contexto em que é desenvolvida. Ou seja, é requerido um contexto de estratégia e planejamento de inteligência pré-estabelecido para que a produção do conhecimento referente ao estado de um tópico de interesse seja coerente. O contexto referente a solução proposta está representado à esquerda da visão geral da proposta (figura 4) e é abordado com mais detalhes na seção 3.1.

Após a definição das necessidades e dos requisitos para a inteligência, o método proposto segue com a etapa de coleta de dados textuais e seus metadados, o processamento de dados e preparação textual, a aplicação de técnicas de processamento de linguagem natural e aprendizado de máquina, a extração do conhecimento produzido, e a representação desse conhecimento em formato de grafo. As etapas do método proposto e suas respectivas responsabilidades podem ser observadas na figura 4.

O detalhamento de cada etapa de processamento e análise de dados é realizado nas seções a seguir: a seção 3.1 evolui sobre o planejamento de inteligência pré-estabelecido assumido como premissa pelo trabalho, a seção 3.2 aprofunda o entendimento sobre o processo de coleta de mensagens e as características do conjunto de dados utilizado, a seção 3.3 detalha as ações de pré-processamento dos dados textuais e o processamento e normalização dos metadados, a seção 3.4 aprofunda na aplicação de técnicas de mineração de dados para identificar os principais tópicos conversados e caracteriza as técnicas de aprendizado de máquina utilizadas para análise de similaridade, a seção 3.5 detalha como o resultado do processamento e análise e são integrados, e a seção 3.6 evolui sobre como o conhecimento é representado para consulta e interação.

3.1 Estratégia e Planejamento de Inteligência

Um produto de inteligência é construído por uma organização de inteligência e é feito sob medida para o consumidor da inteligência a partir de uma necessidade de conhecer previamente manifestada e da atribuição da responsabilidade de produção para a organização. Portanto, não existe produto sem estratégia e planejamento de inteligência.

Dessa forma, a solução proposta toma como premissa fatores de estratégia e planejamento de inteligência para que o produto gerado seja útil e relevante para os consumidores, essas premissas são:

- Foi atribuído à organização de inteligência a responsabilidade de conhecer sobre o ecossistema de crimes cibernéticos do Brasil, produzindo inteligência sobre as

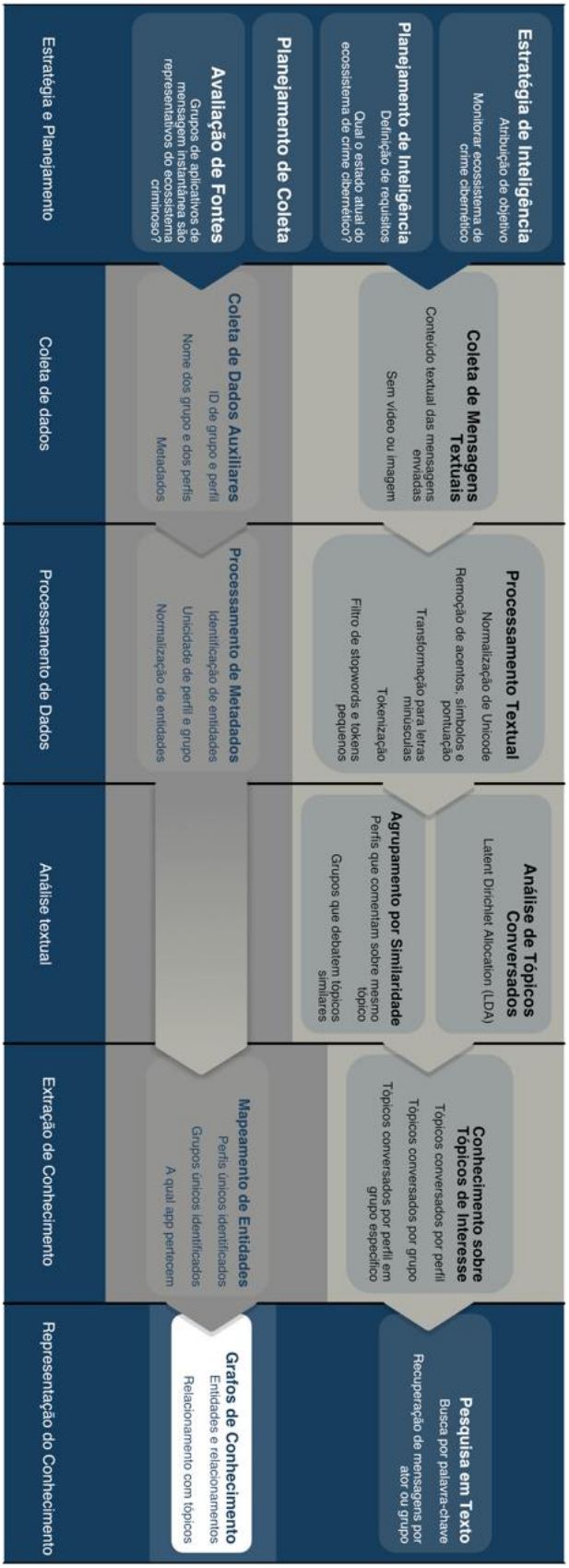

Figura 4 – Visão Geral da Proposta.

ameaças identificadas;

- Se faz necessária a construção de uma consciência situacional sobre os interesses dos atores de ameaça que atuam no ecossistema de crime cibernético brasileiro;
- As mensagens textuais de grupos sobre crimes cibernéticos em aplicativos de mensagens instantâneas são uma fonte de dados relevante para atender os objetivos de coleta.

3.2 Coleta de Dados

O objetivo da etapa de coleta de dados é identificar grupos em aplicativos de mensagens instantâneas que sejam representativos do ecossistema de crime cibernético do Brasil, coletando as mensagens textuais e seus metadados de acordo com os objetivos de coleta estabelecidos no planejamento de inteligência (seção 3.1). Portanto, a identificação e monitoramento de grupos de forma ilimitada não é adequada para atender ao planejamento de inteligência, havendo a necessidade de filtragem prévia.

Dessa forma, o processo de filtragem prévia é realizado através da avaliação de relevância da fonte através da identificação de características associadas ao ecossistema de crimes cibernéticos, sendo elas:

- Terminologia ou gírias características de crime cibernético sendo utilizadas no nome ou descrição do grupo;
- Perfis participantes do grupo com nomes ou imagens referenciando contextos de crime cibernético;
- Mensagens fixadas contendo anúncios de esquemas de fraude ou técnicas de execução de ataques cibernéticos;
- Presença de terminologia ou gírias características de crime nas mensagens do grupo (pode ser verificado através da funcionalidade de pesquisa).

Os grupos que possuem essas características são considerados representativos do ecossistema de crime cibernético, dado que debatem e fomentam atividades criminosas específicas. Deste modo, são grupos selecionados para coleta e monitoramento.

3.3 Processamento de Dados

A etapa de processamento de dados é subdvidida entre o processamento dos metadados e a preparação textual. A etapa de processamento de metadados (subseção 3.3.1) trata dos metadados estruturados referentes aos usuários, grupos e aplicativos de mensagem que foram alvos de coleta, objetivando definir critérios de unicidade e normalizar

os dados de acordo com a entidade representada. Por outro lado, a etapa de preparação textual (subseção 3.3.2) trata as mensagens textuais coletadas e objetiva realizar o pré-processamento dos textos para que técnicas de análise de tópicos possam ser aplicadas com maior grau de sucesso.

3.3.1 Processamento de Metadados

A etapa de processamento de metadados consiste no tratamento dos dados relacionados ao aplicativo, grupo e perfil a fim de definir as entidades analisadas e suas características de unicidade, conforme mostrado na figura 5. O objetivo desta etapa é fornecer uma normalização inicial das entidades processadas para que as técnicas de análise de tópicos de interesse possam ser realizadas em contextos específicos, como principais tópicos mencionados em um determinado grupo ou por um determinado ator de ameaça.

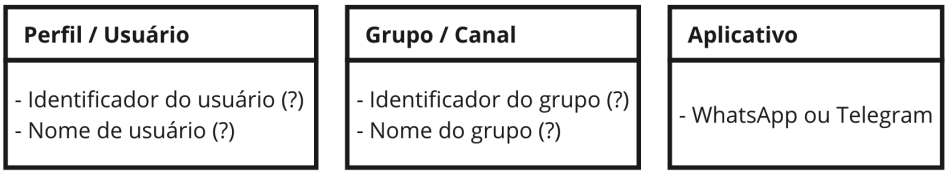

Figura 5 – Entidades processadas através dos metadados.

O critério de unicidade para usuários e grupos é a combinação do nome do usuário ou do grupo e seu respectivo ID. Os dados de ID de usuário não estão disponíveis para todos os registros, devido a dificuldade de coleta deste dado. Portanto, é feita a distinção entre nome de usuário associado a um ID específico e o mesmo nome de usuário sem ID disponível. A definição do aplicativo é binária, sendo Telegram ou WhatsApp.

3.3.2 Preparação Textual

A preparação textual consiste de diversas etapas devido às necessidades específicas introduzidas pelo estilo e contexto das mensagens analisadas. As etapas necessárias estão listadas na figura 6.

As mensagens de texto coletadas de grupos de aplicativos de mensagens instantâneas naturalmente fazem uso de linguagem informal. Dentro do contexto de crime cibernético, esse fator é potencializado devido ao uso extensivo de gírias específicas e formatos de escrita alternativos. Uma característica comum é a escrita de mensagens utilizando caracteres unicode de símbolos (como o alfabeto matemático) com o intuito de produzir mensagens mais chamativas para outros atores de ameaça, mas que também causa o efeito colateral de dificultar a análise automatizada de texto. Dessa forma, é necessário realizar a normalização de caracteres unicode para o formato NFKD como uma tentativa de interpretar caracteres especiais como as letras que eles representam (como o símbolo *U+1D402* que representa a

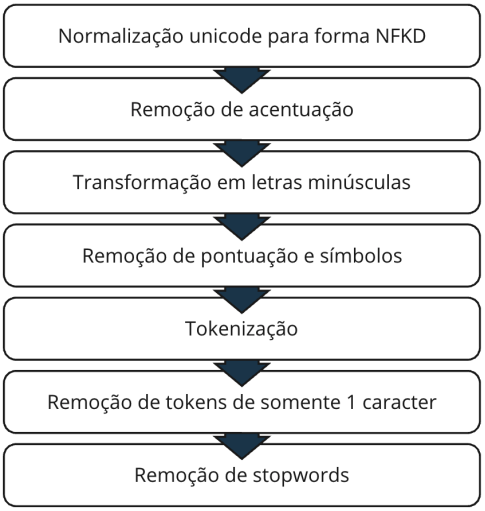

Figura 6 – Etapas da preparação textual.

letra C). Além destas etapas, também é necessário a transformação de letras maiúsculas em minúsculas e a remoção de acentos, pontuação, símbolos, stopwords e tokens com apenas uma letra.

3.4 Análise Textual

A etapa de análise textual é subdividida nas etapas de análise de tópicos e agrupamento por similaridade. A etapa de análise de tópicos (subseção 3.4.1) objetiva identificar os principais assuntos abordados por grupos e usuários através da identificação dos termos mais representativos utilizados e do grau de representatividade de cada termo. Após a análise dos tópicos, a etapa de agrupamento por similaridade (subseção 3.4.2) objetiva identificar *clusters* de usuários e grupos que discutam tópicos similares, com o intuito de revelar especializações dentro do ecossistema de crime cibernético.

3.4.1 Análise de Tópicos Conversados

A etapa de análise de tópicos tem o objetivo de identificar os principais tópicos de interesse dos perfis e grupos através da análise dos textos pré-processados (subseção 3.3.2) referentes as mensagens textuais transmitidas nos grupos pelos perfis.

Entretanto, devido a natureza como esses grupos são organizados e como atores de ameaça se comportam, é necessários que a etapa de análise de tópicos conversados seja capaz de lidar com os seguintes empecilhos:

1. Linguagem informal: As mensagens possuem escrita informal devido ao seu contexto de transmissão em aplicativos de mensagens instantâneas, podendo conter erros gramaticais e contrações de palavras;

2. Termos únicos: As mensagens são geradas com forte uso de gírias e dialetos específicos relacionados ao contexto de crime cibernético no Brasil;

3. Desempenho de processamento: Para que a consciência situacional represente a escala do ecossitema de crimes cibernéticos, é necessário processar uma grande quantidade de dados textuais em um tempo computacionalmente hábil;

4. Desbalanceamento de informações textuais: Poucos atores de ameaça transmitem muitas mensagens, enquanto muitos atores de ameaça transmitem poucas mensagens.

3.4.2 Agrupamento de Perfis e Grupos

O objetivo da etapa de agrupamento de perfis e grupos é fornecer uma noção de similaridade entre os perfis e os grupos, identificando comunidades dentro do ecossistema de crimes cibernéticos que se interessam por tópicos similares. Para alcançar esse objetivo, os seguintes relacionamentos são considerados:

- Perfil $\rightarrow$ Tópico: este relacionamento considera todas as mensagens enviadas por um perfil específico em todos os grupos monitorados, buscando agrupar os atores que possuem os mesmos interesses independente de quais grupos eles fazem parte;
- Grupo $\rightarrow$ Tópico: este relacionamento considera todas as mensagens transmitidas dentro do contexto de um grupo específico, independente de qual perfil realizou o envio, buscando agrupar os grupos que tratem de assuntos similares;
- Perfil em grupo $\rightarrow$ Tópico: este relacionamento considera somente as mensagens enviadas por um perfil específico dentro de um grupo específico, possuindo o objetivo de agrupar os perfis que possuem interesses similares dentro do contexto de um grupo.

3.5 Extração de Conhecimento

A etapa de extração de conhecimento objetiva integrar o resultado produzido pelas etapas de processamento de metadados, análise de tópicos conversados e agrupamento por similaridade em um formato estruturado de entidade e relacionamento, possibilitando a recuperação e correlacionamento de informações. Para alcançar esse objetivo, a extração de conhecimento conta com duas subetapas: mapeamento de entidades, detalhada na subseção 3.5.1, e conhecimento sobre tópicos de interesse (subseção 3.5.2).

3.5.1 Mapeamento de Entidades

A etapa de mapeamento de entidade objetiva utilizar as regras de unicidade e normalização definidas na etapa de processamento de metadados (subseção 3.3.1) para mapear todas as entidades únicas e seus relacionamentos iniciais, possibilitando que outros

conhecimentos possam ser relacionados às entidades. Os relacionamentos usados como base podem ser observados na figura 7.

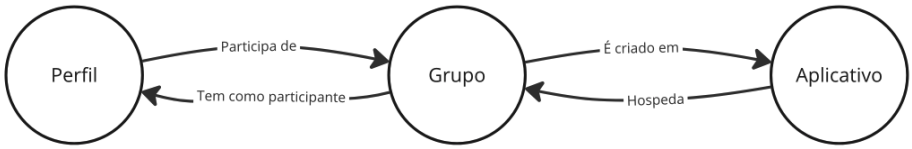

Figura 7 – Entidades mapeadas e seus relacionamentos.

A etapa de mapeamento de entidades fornece a capacidade de compreensão dos relacionamentos fundamentais para a integração futura dos conhecimentos produzidos por outras etapas.

3.5.2 Conhecimento sobre Tópicos de Interesse

A etapa de conhecimento sobre tópicos de interesse objetiva extrair e integrar o conhecimento produzido na análise de tópicos com as entidades e relacionamentos básicos determinados na etapa de mapeamento de entidades (subseção 3.5.1), adicionando novos relacionamentos e entidades referentes ao interesse em tópicos (como observado na figura 8).

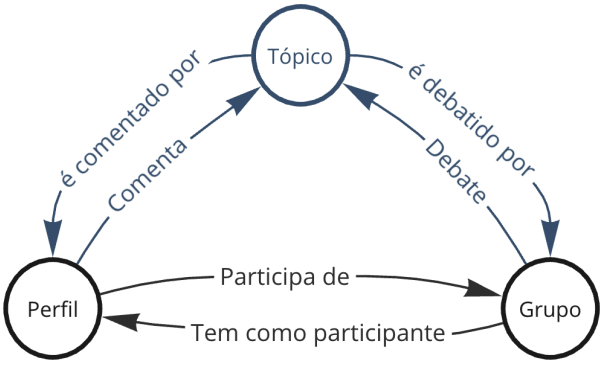

Figura 8 – Integração de tópicos a entidades mapeadas.

Para alcançar esse objetivo, a etapa de conhecimento sobre tópicos de interesse realiza as seguintes ações:

- Adiciona os tópicos como entidades no modelo de dados;
- Cria os relacionamentos entre as entidades base (e.g. perfil e grupo) com os tópicos identificados, disponibilizando o grau de representatividade do tópico como atributo do relacionamento;
- Garante que os tópicos e os relacionamentos criados seguem os padrões de unicidade e normalização a fim de evitar dados inconsistentes ou repetidos.

3.6 Representação do Conhecimento

A etapa de representação do conhecimento objetiva integrar e disponibilizar o conhecimento produzido representando-o através das entidades mapeadas, os relacionamentos criados entre as entidades, e as características dos relacionamentos extraídas através de análise. O formato de disponibilização escolhido foi o de grafos de conhecimento, devido a sua versatilidade para iteração durante análises mais aprofundadas posteriormente. O modelo de dados definido pode ser observado na figura 9.

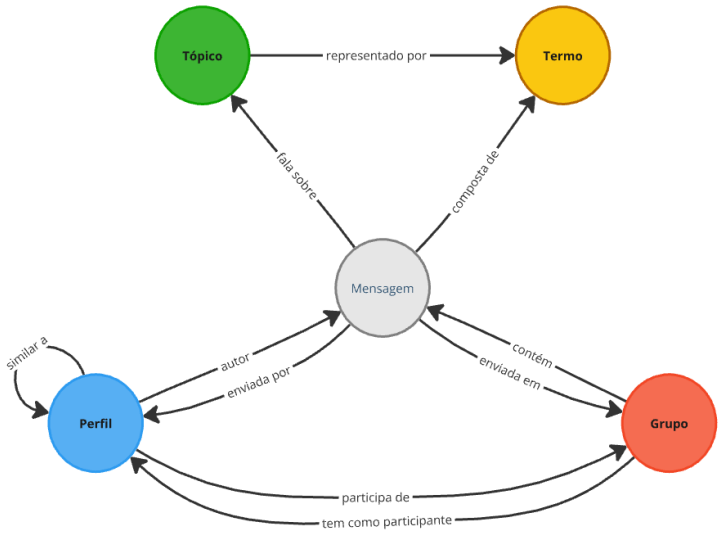

Figura 9 – Conhecimento representado em formato de grafo.

Para facilitar a interação com os grafos de conhecimento através de uma solução visual e com ferramentas robustas de busca em grafos, o sistema de gerenciamento de banco de dados de grafos Neo4j foi escolhido.

4 AVALIAÇÃO EXPERIMENTAL

Este capítulo discute a construção de um conjunto de dados representativo em relação as premissas de planejamento de inteligência estabelecidas previamente (Capítulo 3 seção 3.1), a configuração de experimentos para avaliação da qualidade da análise textual e do suporte à consciência situacional, e os resultados obtidos nos experimentos realizados. Dessa forma, este capítulo está organizado nas seguintes seções: Coleta dos Dados Base 4.1, Estruturação do Conjunto de Dados (seção 4.2), Desafios para Análise de Tópicos no Conjunto de Dados 4.3, Configuração Experimental 4.4, e Resultados e Discussões 4.5.

4.1 Coleta dos Dados Base

A coleta dos dados deve estar alinhada ao planejamento de inteligência, atendendo aos objetivos de inteligência. Dentro do contexto de planejamento discutido na seção 3.1, os seguintes objetivos de inteligência são considerados:

- Identificar os interesses dos perfis e grupos participantes do ecossistema Brasileiro de crimes cibernéticos;
- Identificar perfis e grupos com interesses similares e seus respectivos graus de proximidade;
- Disponibilizar o conhecimento produzido de forma que suporte a criação de consciência situacional sobre esse ecossistema.

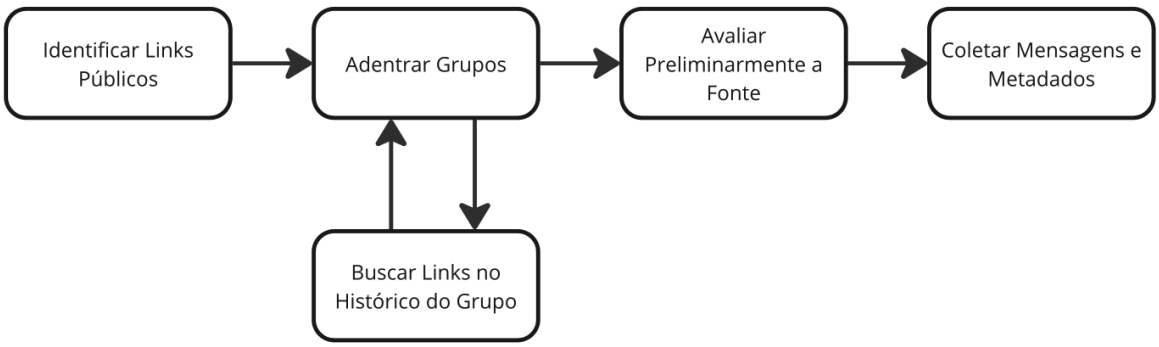

Figura 10 – Processo base de coleta dos dados.

Com os objetivos de inteligência estabelecidos, a etapa de coleta dos dados foi realizada de acordo com o processo demonstrado na figura 10. As etapas que compõem o processo de coleta estão descritas a seguir:

- Identificar Links Públicos: Através de técnicas de OSINT, procurar e selecionar links de convites publicamente compartilhados para grupos que pareçam relacionados a crimes cibernéticos;

- Adentrar Grupos: Utiliza os links identificados na etapa anterior para adentrar (ou pré-visualizar no caso do aplicativo Telegram) os grupos;

- Buscar Links no Histórico do Grupo: Em alguns aplicativos (como o Telegram) é possível consultar o histórico de mensagens do grupo, possibilitando uma busca retroativa por outros links compartilhados restritamente;

- Avaliar Preliminarmente da Fonte: Consiste em analisar o conteúdo do grupo de forma preliminar (análise de histórico ou conteúdo de mensagens recentes) buscando identificar as características que indicam atividade relacionada a crimes cibernéticos. Essas características estão especificadas na Tabela 4;

- Coletar Mensagens e Metadados: Habilita a coleta das mensagens enviadas no grupo através de automação, armazenando os dados e metadados para que possam ser analisados posteriormente.

Critério	Descrição
Linguagem	Presença de terminologia ou gírias características de crime cibernético no nome ou descrição do grupo.
Participantes	Perfis participantes com nomes, fotos ou *nickname* referenciando elementos de crime cibernético.
Conteúdo	Presença de mensagens anunciando esquemas de fraude, técnicas de execução de ataques cibernéticos, ou mencionando outros objetivos criminosos (como estelionato).

Tabela 4 – Critérios de avaliação da representatividade das fontes de dados em relação aos objetivos de inteligência.

Após a realização da coleta durante um período de 31 dias, o conjunto de dados base para a análise foi montado, totalizando mais de 6 milhões de linhas e 7 colunas, com 2 colunas extras sendo adicionadas após a etapa de pré-processamento textual. A Tabela 5 detalha as características das colunas do conjunto de dados base. O período de coleta dos dados e a respectiva quantidade de mensagens coletadas por dia estão visualmente representados na Figura 11.

A coluna *preprocessed* é gerada quando o conteúdo das mensagens (coluna *content*) é submetido a etapa de preprocessamento, que é encarregada de normalizar caracteres unicode, remover emojis e acentos, remover caracteres especiais, remover números, entre outros. Por fim, a transformação do texto preprocessado em tokens (lista de palavras) é

Coluna	Tipo dos Dados	Origem	Valores
messaging_app	String	Metadado	Telegram ou WhatsApp
author_id	String	Coleta	ID do Telegram ou Número do celular
author_name	String	Coleta	Username
channel_id	String	Coleta	ID do grupo no app
channel_name	String	Coleta	Nome do grupo no app
content	String	Coleta	Conteúdo puro da mensagem
event_date	Timestamp	Coleta	12/02/23 - 14/03/2023
preprocessed	String	Processamento	Texto da mensagem pré-processado
tokens	Array	Processamento	Texto da mensagem tokenizado

Tabela 5 – Características do Conjunto de Dados Base.

aplicada no resultado da etapa anterior (coluna *preprocessed*), realizando a remoção de palavras de pouca relevância (comumente denominadas *stopwords*) ao longo da transformação e dando origem a coluna *tokens*.

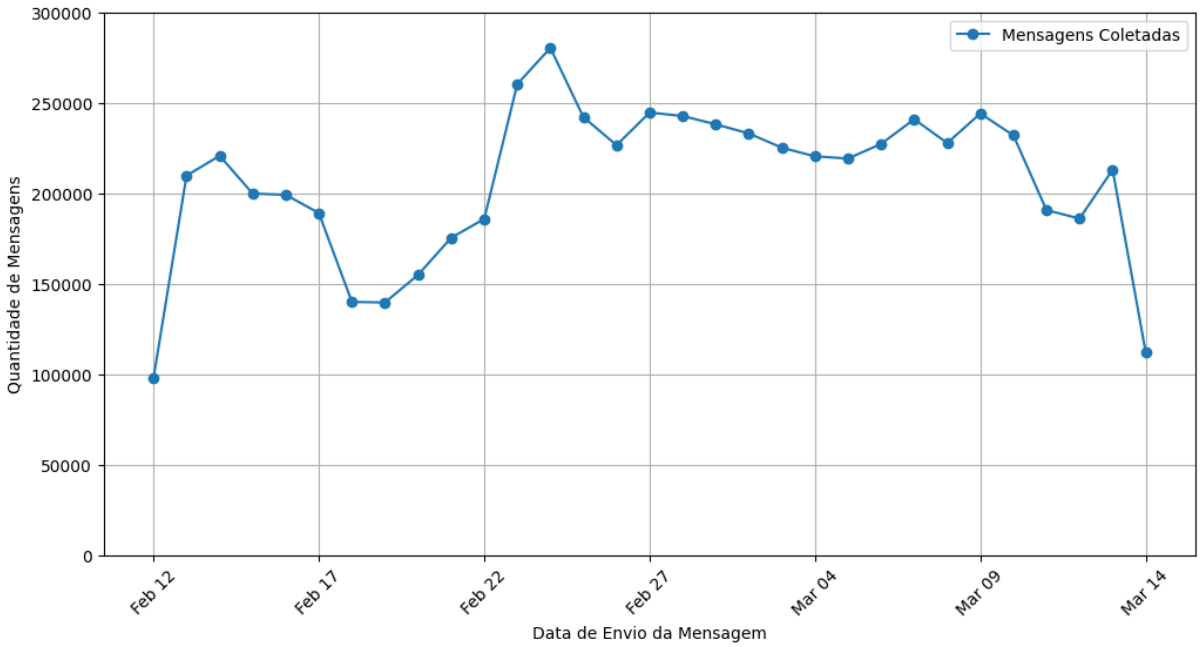

Figura 11 – Quantidade de Mensagens Coletadas por Dia.

A partir do conjunto de dados explicado na tabela 5, a formulação técnica dos principais conceitos utilizados no decorrer do capítulo podem ser explicados:

- Perfil: é identificado pela unicidade dos valores conjuntos das colunas *messaging_app*, *author_id* e *author_name*. Ou seja, todos os valores únicos das 3 colunas ao mesmo

tempo: *UNIQUE(messaging_app, author_id, author_name)*;

- Grupo: é identificado pela unicidade dos valores conjuntos das colunas *messaging_app*, *channel_id* e *channel_name*. Ou seja, todos os valores únicos das 3 colunas ao mesmo tempo: *UNIQUE(messaging_app, channel_id, channel_name)*;

- Mensagens por perfil: é o resultado da contagem de linhas pertencentes ao agrupamento por um determinado perfil, ou seja: *groupby(messaging_app, author_id, author_name).count()*;

- Mensagens por grupo: é o resultado da contagem de linhas pertencentes ao agrupamento por um determinado grupo, ou seja: *groupby(messaging_app, channel_id, channel_name).count()*;

- Tokens por mensagem: é o resultado da contagem de quantos tokens estão presentes no array de tokens (coluna *tokens*). Ou seja: *size(tokens)*;

- Tokens por perfil: é o resultado da contagem de quantos tokens estão presentes no conjunto de arrays de tokens referentes às mensagens enviadas por um perfil. Ou seja: *groupby(perfil).agg(sum(size(tokens)))*.

- Tokens por grupo: é o resultado da contagem de quantos tokens estão presentes no conjunto de arrays de tokens referentes às mensagens enviadas em um grupo. Ou seja: *groupby(grupo).agg(sum(size(tokens)))*;

4.2 Estruturação do Conjunto de Dados Final

Os grupos que possuíam características suficientes para serem aprovados na avaliação preliminar (Tabela 4) se tornaram elegíveis para a segunda avaliação da fonte, onde foi considerado o grau de representatividade da fonte em relação ao ecossistema de crimes cibernéticos brasileiro. Essa avaliação mais aprofundada da fonte representa uma remoção de linhas do conjunto de dados, não afetando colunas. As características analisadas nessa segunda avaliação são:

- Diversidade de mensagens e participantes: grupos dominados por um conjunto limitado de perfis podem não contribuir para a representatividade do ecossistema, especialmente se a maior parte das mensagens forem spam de um mesmo perfil;

- Escopo de mensagens lícitas: alguns grupos são criados focando uma geolocalização específica e, portanto, são dominados por mensagens anunciando atividades lícitas destinadas ao moradores locais (e.g. manicure, pedicure). Por mais que exista tentativa de estelionato, o conteúdo pode não ser suficientemente representativo.

Os esforços de avaliação da fonte de dados são imprescindíveis para garantir que o resultado obtido atende às necessidades de inteligência estipuladas no planejamento inicial, filtrando elementos que não contribuem para os objetivos de análise.

Após a avaliação mais aprofundada da fonte, os grupos selecionados como representativos do ecossistema de crime cibernético Brasileiro foram submetidos às etapas de processamento de dados e extração de conhecimento, compondo o conjunto de dados final que foi utlizado para a criação da consciência situacional. As características do conjunto de dados final (após avaliação aprofundada da fonte) em comparação com o conjunto de dados base (Seção 4.1) são detalhadas na tabela 6.

Métrica	Original	Final
Total de Mensagens	6.415.856	4.204.554
Perfis Identificados	67.749	60.095
Grupos Identificados	111	79
Tokens por Mensagem		
Média	45,3	32,7
Desvio Padrão	86,21	78,42
Mínimo	0	1
Mediana	18	9
Máximo	2.099	1.718
Mensagens por Perfil		
Média	94,7	69,96
Desvio Padrão	3.389,25	1.894,45
Mínimo	1	1
Mediana	4	4
Máximo	704.509	281.596
Tokens por Perfil		
Média	4.289,78	2.287,74
Desvio Padrão	158.574,8	100.417,9
Mínimo	0	1
Mediana	13	10
Máximo	27.214.580	16.129.750
Mensagens por Grupo		
Média	57.800,5	53.222,2
Desvio Padrão	81.672,3	54.709,37
Mínimo	1	11.548
Mediana	35.310	38.740
Máximo	388.923	325.333
Tokens por Grupo		
Média	2.618.275	1.740.275
Desvio Padrão	5.076.944	1.605.205
Mínimo	31	211.760
Mediana	1.259.422	1.379.367
Máximo	35.760.600	9.650.417

Tabela 6 – Características do Conjunto de Dados Original e Filtrado.

4.3 Desafios para Análise de Tópicos no Conjunto de Dados

O conjunto de dados analisado, referente a comunicação informal sobre crimes digitais entre perfis de aplicativos de mensagem instantânea, demonstrou ser repleto de desafios para a análise de tópicos conversados. Os principais desafios estão descritos nas subseções a seguir a fim de facilitar a compreensão da dificuldade de conseguir resultados ótimos na análise de tópicos.

4.3.1 Desbalanceamento de Informação Textual

A forma como atores de ameaça se comportam e comunicam dentro dos aplicativos de mensagens instantâneas é bem diversa, consistindo desde perfis que enviaram somente 1 mensagem com um único token (característico de usuários de bots de consulta) até perfis que enviam uma quantidade massiva de mensagens textuais (como os bots de consulta de dados). As características da distribuição da quantidade de tokens extraídos de cada mensagem podem ser observadas na figura 12.

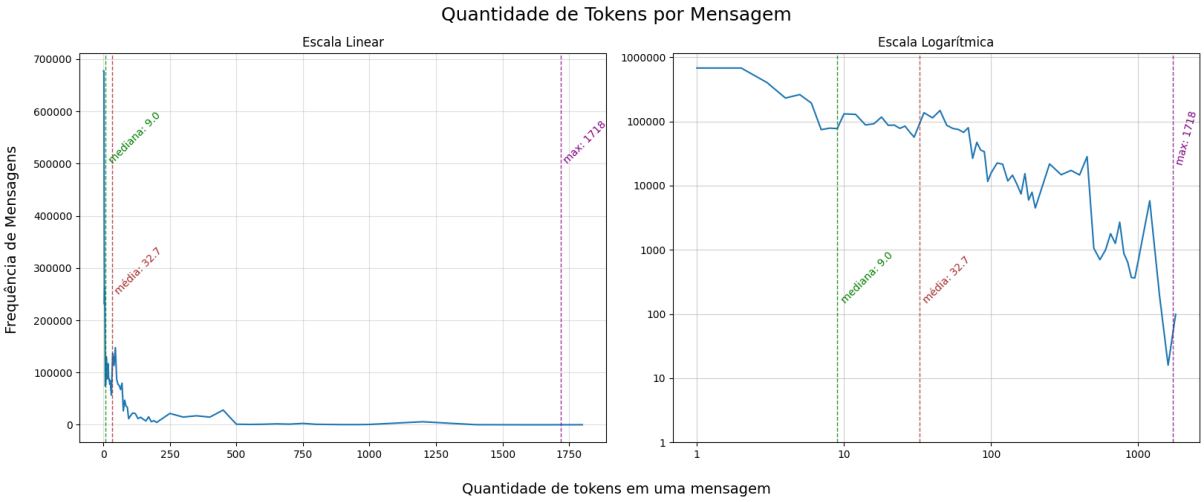

Figura 12 – Características da distribuição de tokens por mensagem.

Portanto, pode-se observar que os dados textuais disponíveis para análise são extremamente desbalanceados, partindo desde mensagens e perfis com pouquíssima informação textual analisável, atrapalhando técnicas baseadas em análise de coocorrência de palavras, até perfis com informação suficiente para análises complexas. Essa dificuldade exige que a solução proposta seja capaz de lidar com alto grau de desbalanceamento de tamanho de textos, modelando tópicos em textos curtos e longos.

4.3.2 Ausência de Estrutura Sintática Coesa

As mensagens mais longas (com maior quantidade de tokens) possuem a característica da ausência de estrutura sintática coesa, estando mais próximas de um conjunto

de frases pequenas aglomeradas do que um texto com parágrafos coesos. Isso pode ser observado em mensagens que anunciam cursos, técnicas e métodos de fraude em diversas instituições, como o exemplo a seguir:

> "canal de metodos e esquemas metodos atualizados metodos e esquemas valor r$ valor unico atualizacao diariamentes sem taxa de mensalidade suporte garantido bins grupo de clientes metodos super facies de fazer canal organizado com conteudos de qualidades metodos estao em pdf texto video aula lista de metodos sensi ff capa netflix ilimitado bins play store esquema ifood na cc metodo disney plus metodo youtube premium (...) metodo deezer dias metodo play pass esquema submarino esquema picpay esquema dimas ff metodo directv go metodo amazom prime chunchyroll link dias tidal link meses globo play dias com vpn prime video link meses como ganhar seguidores no instagram metodo tim meses metodo vivo novo gb plano controle vivo bins atualizadas (...) esquema polishop esquema racao esquemas virando saldo metodo acorntv painel de contas premium metodo globo play renovavel metodo flix ole curso openbullet metodo ancestry metodo ebay (...) tutorial fazer recarga na vivo ativar plano na vivo forma de pagamento pix"

Além desse tipo de mensagem, também é muito comum que as mensagens longas enviadas por perfis bots relacionados a consulta de dados de pessoas sigam um formato semi-estruturado mas com ausência de coesão sintática, como no exemplo a seguir:

> "dados pessoais nome <filtrado> cpf cns nascimento sexo m masculino mae <filtrado> pai municipio de nascimento estado civil nacionalidade brasileira obito nao data do obito nao consta situacao cadastral regular enderecos cep estado sp municipio <filtrado> logradouro <filtrado> tipo r numero cep estado sp municipio <filtrado> bairro <filtrado> logradouro <filtrado> tipo r numero cep estado (...) numero apto telefones proprietario nao informado nao informado nao informado nao informado nao informado vivo nao informado nao informado nao informado nao informado e mails <filtrado> dados economicos renda poder aquisitivo medio alto faixa aquisitiva de r scorecsba scorecsb serasa mosaic descricao profissionais em ascensao social classe adultos urbanos estabelecidos nova descricao amadurecendo confortavelmente no interior nova classe adultos urbanos estabelecidos classe secundaria adultos urbanos estabelecidos observacoes para mais assertividade levar em consideracao (...) pis registro geral numero orgao emissor ssp uf titulo de eleitor parentes nome <filtrado> cpf grau de parentesco mae beneficios beneficio auxilio emergencial total recebido r internet dados de imposto banco banco do brasil (...) cpf nascimento mae

<filtrado> renda endereco <filtrado> sp cep nome <filtrado> cpf nascimento mae <filtrado> renda endereco (...)"

4.3.3 Linguagem Informal, Gírias Específicas e Contrações Textuais

As comunicações informais observadas nos grupos representativos do ecossistema de crimes digitais no Brasil também possuem a característica de utilização de muitas gírias que representam contextos específicos relacionados a crimes. Algumas dessas gírias são: "*lara*", que significa contas laranjas, e "*chk*" que se refere a um software capaz de realizar testes para ver se cartões de créditos são válidos. Exemplos de mensagens que utilizam essas gírias podem ser observados a seguir:

"É caô dms virar consul itau, nunca consegui, vi que vira mais fácil com lara Itaú"

"C tem chk de gg ?"

Além de gírias específicas de contexto, a linguaguem informal observada também é composta por muitas contrações textuais, abreviando palavras em escritas diferentes. Como "C" signficando "você" e "dms" significando "demais" nos exemplos apresentados.

4.3.4 Escrita Incorreta

Por fim, um outro fator de dificuldade imposta pelo conjunto de dados é a necessidade de lidar com escritas diferentes (erradas) para uma mesma palavra. No exemplo a seguir, é possível ver que o intuito da mensagem era perguntar sobre o "curso", mas a escrita ficou como "cruso":

"Como pego o cruso gratis"

Como boa parte dos textos do conjunto de dados são curtos, existe o risco das técnicas de análise de tópicos não serem capazes de compreender a similaridade entre os termos devido a baixa informação de coocorrência de termos.

4.4 Resultados da Análise Textual

A análise textual é subdividida nas etapas de análise de tópicos e agrupamento por similaridade. A etapa de análise de tópicos objetiva identificar os principais assuntos abordados pelos grupos e perfis através da identificação de quais tópicos estão sendo conversados na base de dados e qual a proximidade de cada mensagem com estes tópicos. Após a análise dos tópicos, a etapa de agrupamento por similaridade objetiva identificar

clusters de usuários e grupos que discutam tópicos similares, com o intuito de revelar especializações dentro do ecossistema de crime cibernético.

Como a base de dados não possuia rotulagem originalmente, posteriormente contando apenas com rótulos parciais, a análise de resultados de forma objetiva era inviável, havendo a necessidade de avaliação subjetiva por especialista de domínio baseado na interpretabilidade dos tópicos e sua coerência com as expectativas. Dessa forma, foi estipulada uma escala (apresentada na Tabela 7) para possibilitar uma forma de comparação entre os resultados obtidos.

Nível	Avaliação de Especialista de Domínio
Ruim	• Tópicos identificados são difíceis de interpretar • Tópicos não são coerentes com as expectativas do domínio • Tópicos misturam terminologias que deveriam estar separadas
Médio	• Alguns tópicos possuem boa interpretabilidade • Alguns tópicos são consistentes com a expectativa do domínio • Alguns tópicos estão bem definidos e possuem terminologia segmentada de acordo com as expectativas do domínio
Bom	• Tópicos são majoritariamente fáceis de interpretar • Tópicos majoritariamente coerentes com as expectativas do domínio • A maioria dos tópicos possuem boa segmentação de terminologia de acordo com as expectativas do domínio
Ótimo	• Tópicos fáceis de interpretar e com boa delimitação de termos • Identificação adequada de tópicos relevantes de menor tamanho • Tópicos grandes não são fragmentados em múltiplos tópicos • Os tópicos identificados possuem boa segmentação de terminologia de acordo com as expectativas do domínio

Tabela 7 – Escala de Avaliação de Resultados da Análise de Tópicos.

Após a análise de tópicos de interesse, é realizado um agrupamento de perfis e grupos por tópicos de interesse similares. Esse agrupamento tem a característica de funcionar como uma análise exploratória do relacionamento entre os perfis e grupos com os tópicos de interesse, identificando comunidades dentro do ecossistema interessadas em conjuntos de termos específicos. Dessa forma, os seguintes relacionamentos serão considerados:

- Perfil para tópico: este relacionamento considera todas as mensagens enviadas por um perfil específico em todos os grupos monitorados, buscando agrupar os atores que possuem os mesmos interesses independente de quais grupos eles fazem parte;
- Grupo para tópico: este relacionamento considera todas as mensagens enviadas dentro do contexto de um grupo específico, independente de qual perfil realizou o envio. O objetivo é agrupar os grupos que tratem sobre tópicos similares, fornecendo uma visibilidade de quais deles podem oferecer melhor contexto sobre um determinado tipo de crime;
- Perfil em grupo para tópico: este relacionamento considera somente as mensagens enviadas por um perfil específico dentro de um grupo específico, possuindo o objetivo de agrupar os perfis que possuem interesses similares dentro do contexto de um grupo.

4.4.1 Experimento: Modelagem de Tópicos

A etapa de análise de tópicos é fundamental para um bom resultado na aplicação do método proposto, pois é a parte responsável por possibilitar as análises de interesses e similaridade entre perfis e grupos. Dessa forma, 4 técnicas principais de modelagem de tópicos, não-supervisionadas e semi-supervisionadas, foram avaliadas em busca de melhores resultados.

4.4.1.1 Saco de Palavras

A primeira técnica analisada para modelagem de tópicos foi a *bag-of-words* (saco de palavras), consistindo de: contagem de termos utilizados nas mensagens, redução de dimensionalidade, agrupamento, e identificação de principais termos por grupo através de TF-IDF aplicado as classes. Todas as configurações e algorítmos avaliados estão representados na tabela 8.

De forma geral, a técnica de representação textual através de saco de palavras gerou resultados ruins, especialmente por não representar bem a similaridade entre termos (e.g. tel e telefone). Além disso, a representação textual gerada é um vetor de comprimento igual ao tamanho do vocabulário identificado (aproximadamente 480 mil termos), resultando em vetores grandes e populados com muitos zeros (esparsos). Essas características impõem dificuldades para as técnicas de redução de dimensionalidade avaliadas (PCA e UMAP), resultando em dimensionalidade reduzida de pouca representatividade (e.g. baixa variância explicada). Essas dificuldades implicam em um processo de agrupamento de baixa qualidade, resultando em grupos de mensagens que não são consideradas similares na avaliação de especialistas de domínio. Portanto, os resultados obtidos através dessas técnicas falham em capturar a similaridade entre mensagens e termos, prejudicando a capacidade de extrair tópicos acurados e relevantes.

Saco de Palavras + PCA + K-Means + c-TF-IDF	
Fatores de Complicação	Resultado
Estimar quantidade de *clusters* Equilíbrio entre granularidade e similaridade de tópicos Representação vetorial de texto muito esparsa Qualidade da redução de dimensionalidade (PCA) Captura da similaridade entre termos	Ruim
Saco de Palavras + PCA + UMAP + HDBSCAN + c-TF-IDF	
Fatores de Complicação	Resultado
Tópicos granulares e pouco interpretáveis Representação vetorial de texto muito esparsa Qualidade da redução de dimensionalidade (PCA) Equilíbrio entre tamanho e especificidade de tópicos Captura da similaridade entre termos	Ruim

Tabela 8 – Avaliação da aplicação de bag-of-words para análise de tópicos.

4.4.1.2 Latent Semantic Analysis

A segunda técnica de análise de tópicos realizada foi a LSA (*Latent Semantic Analysis*), que consiste em gerar a matriz de termos presentes em cada mensagem, reduzir a dimensionalidade através da técnica SVD (*Singular Value Decomposition*), agrupar os documentos similares, e identificar os principais termos de cada grupo através da técnica TF-IDF. Os resultados obtidos na aplicação dessa téncnica podem ser observados na tabela 9.

Matriz termo-documento + SVD(100) + K-Means + c-TF-IDF	
Fatores de Complicação	Resultado
Estimar quantidade de *clusters* Disparidade de tamanho entre tópicos/*clusters* Equilíbrio entre tamanho e especificidade de tópicos Captura da similaridade entre termos	Médio
Matriz termo-documento + SVD(100) + UMAP + HDBSCAN + c-TF-IDF	
Fatores de Complicação	Resultado
Tópicos granulares e pouco interpretáveis Captura de similaridade entre mensagens Equilíbrio entre tamanho e especificidade de tópicos	Ruim

Tabela 9 – Avaliação da aplicação de LSA para análise de tópicos.

A técnica LSA é capaz de identificar parcialmente a similaridade entre termos, resultando em representações textuais melhores para análise de similaridade do que a técnica de saco de palavras. A nivel de resultados, o agrupamento via K-Means da representação gerada pela aplicação da LSA gerou alguns grupos e tópicos com boa interpretabilidade, demonstrando potencial de solução do problema de modelagem. Entretanto, a necessidade de especificar previamente a quantidade de grupos e a suposição implícita da técnica K-Means de que os grupos tem tamanhos similares se mostrou um problema para a base de dados. A base de dados analisada possui muitas mensagens que, apesar de possuírem poucos *tokens*, são representativas para alguns tópicos, resultando em grupos com tamanhos extremamente distintos (e.g. 400 mil elementos vs 5 elementos). Essa característica faz com que o agrupamento via K-Means gere resultados ideais somente para os grupos com maior representatividade, falhando em identificar adequadamente os grupos com poucos elementos. Dessa forma, a técnica de agrupamento via K-Means resultou em alguns grupos bem representados, mas impossibilita que todos os tópicos desejados sejam identificados ao mesmo tempo.

A fim de conseguir suprir as dificuldades do K-Means no agrupamento de grupos de tamanhos diferentes, a técnica de agrupamento por densidade HDBSCAN foi aplicada, com uma etapa de redução de dimensionalidade através de *UMAP(init=PCA)*. Porém, o agrupamento resultante demonstrou que diversos documentos não conseguiram ser bem representados pela técnica LSA, com várias grupos de documentos contendo mensagens julgadas não similares por especialista de domínio.

4.4.1.3 Latent Dirichlet Allocation

Devido ao resultado insatisfatório da técnica LSA, a terceira técnica analizada foi a LDA (*Latent Dirichlet allocation*), que consiste de: contagem de termos presentes em cada documento, aplicação do modelo LDA no resultado para identificação dos tópicos presentes no documento e as palavras-chave de cada tópico. Os resultados obtidos com a técnica LDA podem ser observados na tabela 10.

A técnica LDA leva como premissa que um documento (mensagem) é composto por uma distribuição de alguns tópicos, enquanto cada tópico é composto por uma distribuição de termos (palavras). Essa premissa é muito útil para a interpretabilidade dos resultados alcançados dentro da base de dados usada, pois uma boa parte das mensagens é construída a partir da combinação de termos específicos, sem que os termos necessariamente estejam interconectados em uma função sintática (como mostrado na subseção 4.3). Dessa forma, a identificação do grau de proximidade com cada tópico é relevante para análise posterior, mas requer uma identificação de tópicos precisa.

A análise de tópicos através da técnica LDA rendeu os melhores resultados devido a sua capacidade de lidar com textos curtos e longos simultaneamente sem muita penalização

Matriz termo-documento + LDA(topics=<50)	
Fatores de Complicação	Resultado
Tópicos pequenos não são identificados Tópicos grandes são representados repetidamente *Log Likelihood* e *Perplexity* muito altos	Médio
Matriz termo-documento + LDA(topics=<100)	
Tópicos pequenos são agregados em tópicos pouco similares Tópicos grandes são divididos em poucos tópicos muito similares *Log Likelihood* e *Perplexity* muito altos	Ruim
Matriz termo-documento + LDA(topics=<200)	
Tópicos pequenos são representados com pouca confiança Tópicos grandes são divididos em alguns tópicos muito similares *Log Likelihood* e *Perplexity* altos	Médio
Matriz termo-documento + LDA(topics=200)	
Tópicos representados com baixa confiança Tópicos grandes são divididos em tópicos muito similares Interpretar tópicos menores requer que tópicos maiores sejam divididos	Bom
Matriz termo-documento + LDA(topics=>)	
Tópicos granulares e pouco interpretáveis Tópicos grandes são subdivididos em múltiplos tópicos similares	Ruim

Tabela 10 – Avaliação da aplicação de LDA para análise de tópicos.

de performance. Entretanto, a busca (GridSearchCV) pela quantidade ideal de tópicos demonstrou que é inviável alcançar um resultado ótimo no conjunto de dados analisados. Isso se dá por alguns motivos:

- O conjunto de dados possui uma quantidade desproporcional de mensagens pertencentes a tópicos distintos. Ou seja, alguns tópicos possuem centenas de milhares de mensagens com uma quantidade significante de palavras (*tokens*) enquanto alguns tópicos são mencionados somente por algumas unidades de mensagens. Essa desproporcionalidade gerou problemas de especificidade *versus* generalização de tópicos;

- Os tópicos largos, com uma grande quantidade de mensagens, são replicados com baixa distinção de termos quando o hyperparâmetro de número de tópicos é aumentado;

- Os tópicos pouco falados desaparecem quando o hyperparâmetro de número de tópicos é reduzido;

- Valores altos para o hyperparâmetro de número de tópicos resultam em tópicos não generalizados, com termos muito específicos, inviabilizando que tópicos muito similares sejam compreendidos como um único tópico;

- Portanto, não foi possível conciliar uma boa representatividade de tópicos grandes e pequenos ao mesmo tempo, gerando um problema que valores altos trazem bons resultados para tópicos pequenos mas degradam a performance de tópicos maiores, enquanto valores menores resultam em boa generalização de tópicos maiores ao custo de desconsiderar tópicos menores.

Figura 13 – Alguns dos principais tópicos encontrados pelo melhor modelo LDA.

O melhor modelo LDA encontrado através de GridSearchCV foi utilizado para identificar os 200 tópicos mais relevantes presentes no conjunto de dados. Alguns dos tópicos mais relevantes são apresentados na figura 13, demonstrando a capacidade da

paraphrase-multilingual-mpnet-base-v2	
Observações	Resultado
Alta dimensionalidade dos *embeddings* dificulta processamento Baixa capacidade de capturar similaridade de gírias Treinado com múltiplas linguagens (inclusive português) Tamanho da sequência de entrada é suficiente para maioria das mensagens	Ruim
paraphrase-multilingual-MiniLM-L12-v2	
Dimensionalidade dos *embeddings* facilita processamento Pouca capacidade de capturar similaridade entre mensagens e gírias Treinado com múltiplas linguagens (inclusive português) Tamanho da sequência de entrada é suficiente para maioria das mensagens	Ruim
distiluse-base-multilingual-cased-v1	
Dimensionalidade dos *embeddings* requer processamento extra Baixa capacidade de capturar similaridade de gírias Treinado com múltiplas linguagens (inclusive português) Tamanho da sequência de entrada é suficiente para maioria das mensagens	Ruim
distiluse-base-multilingual-cased-v2	
Dimensionalidade dos *embeddings* requer processamento extra Pouca capacidade de capturar similaridade entre mensagens e gírias Treinado com múltiplas linguagens (inclusive português) Tamanho da sequência de entrada é suficiente para maioria das mensagens	Ruim

Tabela 11 – Avaliação dos *embeddings* gerados pelos modelos de linguagem pré-treinados.

técnica de identificar tópicos de relevância mesmo com as dificuldades técnicas impostas pelo conjunto de dados (Seção 4.3).

4.4.1.4 BERTopic

A quarta técnica analisada foi a de representação textual através de *embeddings* de modelos de linguagem (BERT e similares) e identificação de tópicos através da ferramenta BERTopic, que realiza a redução de dimensionalidade (UMAP por padrão), agrupamento (HDBSCAN por padrão), e extração de palavras chave através de c-TF-IDF. A avaliação dos *embeddings* obtidos através dos modelos de linguagem testados estão listadas na tabela 11.

A qualidade dos *embeddings* gerados pelos modelos de linguagem utilizados foi

Embeddings + PCA + UMAP + HDBSCAN	
Fatores de Complicação	Resultado
Alta dimensionalidade dos *embeddings* dificulta processamento Baixa capacidade de capturar similaridade de gírias Treinado com múltiplas linguagens (inclusive português) Tamanho da sequência de entrada é suficiente para maioria das mensagens	Ruim
Embeddings + UMAP(init=PCA) + HDBSCAN	
Dimensionalidade dos *embeddings* facilita processamento Pouca capacidade de capturar similaridade entre mensagens e gírias Treinado com múltiplas linguagens (inclusive português) Tamanho da sequência de entrada é suficiente para maioria das mensagens	Ruim
Embeddings + UMAP(init=PCA) + K-Means	
Dimensionalidade dos *embeddings* requer processamento extra Baixa capacidade de capturar similaridade de gírias Treinado com múltiplas linguagens (inclusive português) Tamanho da sequência de entrada é suficiente para maioria das mensagens	Ruim

Tabela 12 – Resultados obtidos com BERTopic de forma não-supervisionada.

baixa dado os desafios impostos pelo conjunto de dados analisados (seção 4.3): capacidade de generalização para textos curto e informais na língua portuguesa, capacidade de generalização para gírias específicas de domínio, e capacidade de compreender escritas erradas de palavras. Dessa forma, o ajuste fino de modelos de linguagem pode ser necessário para uma melhor adequação dos *embeddings* gerados para o cenário encontrado.

Apesar das dificuldades com os *embeddings*, o conjunto de dados foi submetido a análise de tópicos através da ferramenta BERTopic, utilizando os *embeddings* gerados pelos modelos de linguagem. Os resultados obtidos através da ferramenta BERTopic, aplicada de modo não supervisionado, podem ser observados na tabela 12.

Por fim, indepente da abordagem escolhida dentro da aplicação da técnica BERTopic, uma das mais modernas atualmente, o resultado da baixa qualidade dos *embeddings* gerados pelos modelos de linguagem resultam em resultados ruins, incapazes de identificar tópicos adequados conforme as expectativas de especialistas de domínio.

4.4.2 Experimento: Agrupamento por Similaridade

Para a etapa de agrupamento por similaridade, o resultado gerado pelo melhor modelo da etapa de modelagem de tópicos (seção 4.4.1.3) foi utilizado como entrada. Os

dados possuem as seguintes características:

- Matriz de 4.204.554 linhas por 200 colunas: o modelo LDA fornece a probabilidade de um determinado documento (mensagem) estar falando sobre cada um dos 200 tópicos definidos como hiperparâmetro, realizando esse cálculo para cada uma das mensagens;
- A soma dos valores das colunas de uma determinada linha possui o valor máximo de 100, refente a 100% da mensagem alocada entre os tópicos mais prováveis;
- Para avaliar a similaridade entre perfis ou grupos, foi utilizada a estratégia de somar os resultados das 200 colunas de todas as mensagens enviadas pelo perfil ou grupo e utilizar a métrica de distância de cosseno. Em outros testes, os dados somados foram redimensionados através da técnica "StandardScaler".

Devido a característica de dimensionalidade dos dados, a técnica PCA de redução de dimensionalidade foi implementada em algumas configurações para melhorar o desempenho da etapa de agrupamento. Os resultados obtidos nos testes de agrupamento podem ser observadas abaixo:

Técnica	Métrica de Distância	Considerações	Resultado
GroupBy + Sum PCA(10) DBSCAN	Cosseno	• PCA explica pouca variância de 200 dimensões para 10 • Métrica de cosseno não considera grau de interesse • DBSCAN necessita de forte ajuste de hiperparâmetros	Médio
GroupBy + Sum StandardScaler PCA(10) K-Means	Euclideana	• PCA explica pouca variância de 200 dimensões para 10 • Métrica euclideana diferencia graus de interesse em tópicos • K-Means requer conhecimento prévio da quantidade de *clusters*	Ruim
GroupBy + Sum StandardScaler K-Means	Euclideana	• 200 dimensões afetam a velocidade da aplicação do K-Means • Métrica euclideana diferencia graus de interesse em tópicos • K-Means requer conhecimento prévio da quantidade de *clusters*	Ruim
GroupBy + Sum StandardScaler UMAP(5) HDBSCAN	Euclideana	• HDBSCAN não agrupa *outliers* • Métrica euclideana diferencia graus de interesse em tópicos • UMAP requer inicialização por PCA para melhorar velocidade	Ruim

Tabela 13 – Resultados do Agrupamento de Grupos e Perfis por Similaridade.

Por fim, a opção com melhor resultados foi evitar a etapa de agrupamento e lidar apenas com a matriz de distância entre elementos. Dessa forma, foi possível lidar com busca por similaridade e dissimilaridade sem requerer ajuste fino de hiperparâmetros ou avaliação de qualidade de agrupamento.

4.5 Resultados sobre Consciência Situacional

Existem 3 níveis de consciência situacional: percepção, compreensão e projeção, sendo que cada nível requer um conjunto de informações e diferentes níveis de profundidade de entendimento sobre os elementos avaliados (seção 2.1.4). Dessa forma, o nível de consciência situacional que pode ser alcançado através do suporte oferecido pela solução é feito de forma subjetiva por análise de especialista de domínio. A escala de avaliação da completude da consciência situacional e seus respectivos critérios são detalhados na Tabela 14.

Nível	Descrição
Percepção	O quê? Quando? Onde? Quem?
Compreensão	Porquê? Como? Quanto tempo? Quanto?
Projeção	Como estará? O que pode acontecer? O que esperar?

Tabela 14 – Critérios para Avaliação de Completude de Consciência Situacional.

Utilizando a escala de avaliação da completude da consciência situacional como método avaliador do suporte da consciência situacional, são propostos 3 casos de uso relacionados a análises de inteligência de ameaças do contexto de crime cibernético Brasileiro. Esses casos de uso são relacionados a análise sobre o contexto e interesses de um perfil específico (subseção 4.5.1), análise sobre quais perfis demonstram interesse sobre um conjunto de termos e qual o tipo de interesse (subseção 4.5.2), e identificação de grupos e perfis de interesse para monitoramento dedicado (subseção 4.5.3).

4.5.1 Experimento: Análise de Perfil

É comum que organizações de inteligência que possuam a atribuição de produzir inteligência sobre ameaças relacionadas a crimes cibernéticos (desde pequenas equipes até instituições dedicadas) sejam requisitadas a construir conhecimento sobre um determinado perfil de rede social. As razões para este tipo de requisição são diversas, sendo alguns exemplos: um determinado perfil postou uma mensagem ameaçando a instituição ou parceiros que a instituição ajuda a salvaguardar, a instituição recebeu informações que indicam envolvimento de um determinado perfil não mapeado previamente, e um determinado perfil parece possuir alto grau de influência sobre outras ameaças.

Para este caso de uso, é considerado que houve razão para requisição de construção de conhecimento com os seguintes perguntas para a inteligência:

- Quais os principais termos mencionados e tópicos discutidos pelo perfil?
- Quais outros perfis demonstram comportamento e interesse similares a esse perfil?

Este caso de uso voltado a análise de perfil simula uma situação comumente encontrada por organizações responsáveis por produzir inteligência de ameaças, possibilitando uma avaliação de impacto da consciência situacional como suporte a análise. Dessa forma, foi conduzida uma sequência de etapas partindo da pergunta para inteligência até a visualização das informações processadas, essas etapas são:

- Entedimento das perguntas de interesse e estruturação do plano de consultas para organizar as informações necessárias:
  1. Definir consulta pelas informações gerais do perfil;
  2. Definir consulta pelos termos mencionados pelo perfil desejado;
  3. Definir consulta pelos tópicos mais relevantes associados ao perfil desejado,;
  4. Definir consulta pelos perfis similares ao perfil desejado, repetindo as etapas de 1 a 3 para os 5 perfis mais similares.
- Execução progressiva das consultas, agregando os resultados obtidos aos resultados das consultas anteriores;
- Criação de visualização em grafo com as informações consultadas;
- Disponibilização da visualização para interação de analista humano.

O resultado final obtido pela execução dessas etapas para o caso de uso de análise de perfil pode ser observado na figura 14. A nível de qualidade do resultado, o método é capaz de automatizar o nível 1 da construção da consciência situacional, através da disponibilização das informações sobre os elementos observados no ambiente, e é capaz de fornecer suporte ao alcance dos níveis 2 e 3, devido a disponibilização em forma visual das informações coletadas e da possibilidade de consulta e análise de dados históricos já processados.

4.5.2 Experimento: Perfis Interessados em Termos

De forma similar ao caso de uso de análise de perfil apresentado na subseção anterior, é comum que organizações de inteligência sejam solicitadas a identificar perfis interessados em determinados termos e mapear seus interesses. Um exemplo muito comum

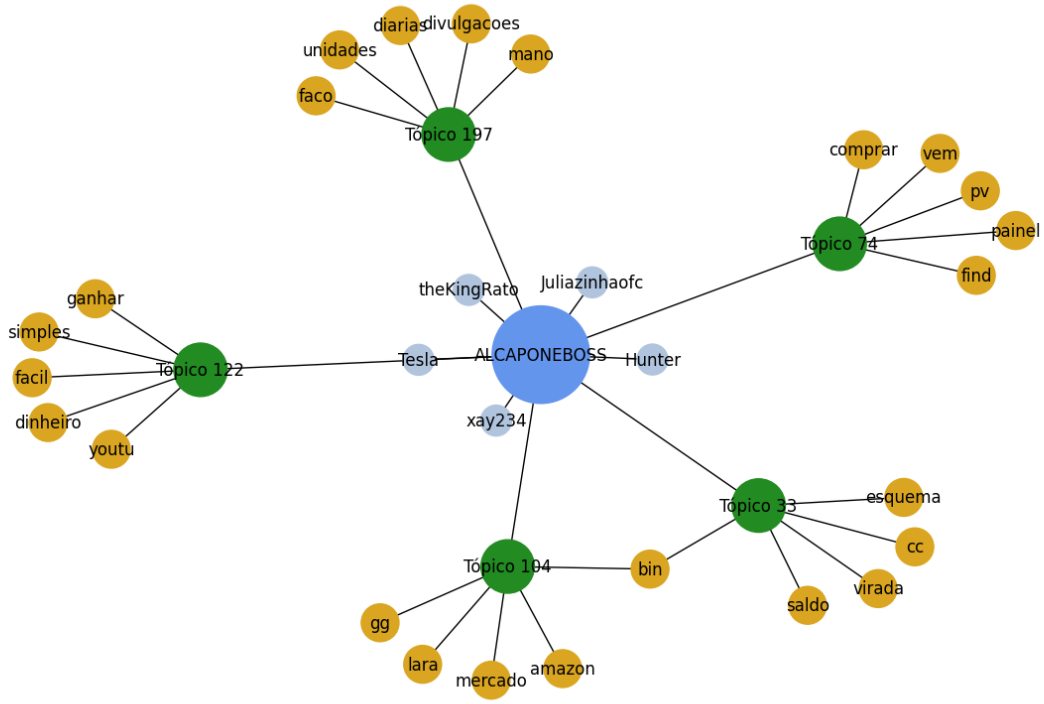

Figura 14 – Resultado do experimento de perfil.

é a busca por termos que se referem a própria organização, objetivando mapear perfis que possuam interesse na instituição e entender se o interesse é maligno (e.g. fraudar, atacar). Esse mapeamento possibilita, entre outras coisas, que a instituição desprenda recursos para monitorar preventivamente esses perfis com o intuito de identificar ataques iminentes ou falhas sendo exploradas.

Para este caso de uso, é considerado que houve solicitação de produção de conhecimento sobre perfis interessados em um conjunto de termos, com os seguintes requisitos para a inteligência:

- Quais perfis possuem interesses nos grandes bancos privados que operam no Brasil (e.g. Itáu, Bradesco e Santander)?
- Qual o contexto disponível desses perfis (e.g. quais grupos participam, tópicos de interesse, e a qual aplicativo de mensagem instantânea pertencem)?

O caso de uso selecionado para o experimento define 3 instituições de interesse a fim de simular um cenário de complexidade moderada, dado que é comum que organizações de inteligência sejam requisitadas a produzir conhecimento desde 1 instituição (normalmente a própria instituição), até várias (e.g. instituições financeiras como um todo). Além de propor

uma dificuldade moderada, esse tipo de requisito é comum na criação de conhecimento sobre possíveis ameaças a uma instituição e possibilitam a avaliação do grau de suporte oferecido pela consciência situacional para a análise.

Portanto, para obter as respostas para as perguntas definidas, foi realizada uma série de etapas relacionadas a consulta de informações e construção de contexto, sendo elas:

- Entedimento das perguntas de interesse e estruturação do plano de consultas para organizar as informações necessárias:

    1. Definir lista de termos associados às instituições de interesse;
    2. Definir consulta para identificar os perfis que utilizaram algum desses termos em suas mensagens;
    3. Definir consulta para identificar os principais tópicos de interesse relacionados a cada perfil;
    4. Definir consulta para identificar em quais grupos monitorados os perfis identificados interagiram;
    5. Definir consulta para atribuir a característica de qual aplicativo de mensagem instantânea esses perfis pertencem;

- Execução progressiva das consultas, agregando os resultados obtidos aos resultados das consultas anteriores;

- Criação de visualização em grafo com as informações consultadas;

- Disponibilização da visualização para interação de analista humano.

O resultado final obtido pela execução dessas etapas para o caso de uso de análise de perfis interessados em termos pode ser visualizado na figura 15. A fim de evitar poluição na figura mostrada, a quantidade de perfis identificados foi limitado a 50, os tópicos de interesse a 5, e os termos relacionados aos tópicos a 10. Na prática, essas limitações podem ser relevadas devido a capacidade de gerar visualizações interativas, que permitem que um analista humano navegue pelo grafo sem restrições quanto a resolução de imagem.

O desempenho do método nesse caso de uso foi avaliado como bom, devido a capacidade de automatizar a percepção dos elementos no ambiente (nível 1 de consciência situacional), e estrutura os dados e informações de forma que possibilita evolução nos níveis 2 (compreensão) e 3 (projeção).

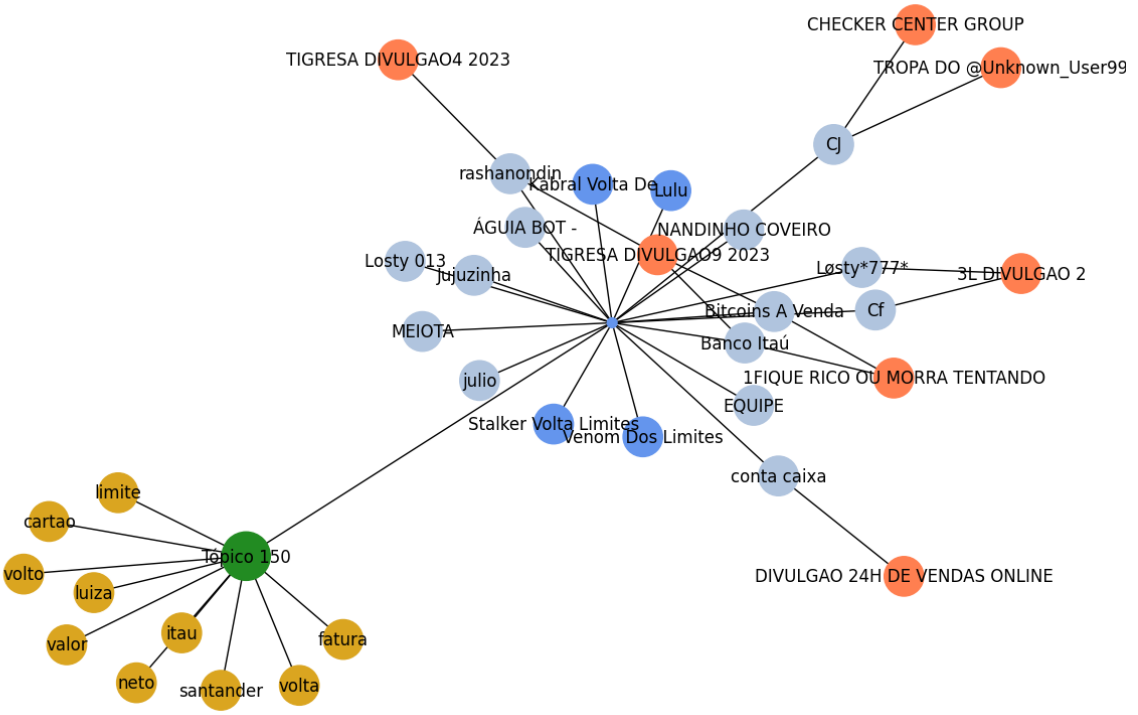

Figura 15 – Resultado do experimento de perfis interessados em temas.

4.5.3 Experimento: Monitoramento de Grupos e Perfis

Um outro caso de uso comumente encontrado por organizações de inteligência que possuem a responsabilidade de produzir inteligência de ameaça é a requisição de identificação e monitoramento de grupos ou perfis relevantes para algum assunto de interesse. Um exemplo comum é quando uma instituição financeira está sofrendo muitas perdas com um determinado tipo de fraude ou ataque cibernético e necessita identificar os atores de ameaça associados ao ataque para tentar reprimir os ataques através das forças da lei. Nesses casos, é comum observar que a necessidade inicial parte do entendimento do ataque ou da fraude, resultando na busca de quais grupos e perfis estão ativamente engajados nessas atividades, e posteriormente implementando monitoramento mais próximo desses grupos e atores com o intuito de embasar solicitação de repressão legal.

Para este caso de uso, considera-se que houve motivação para solicitação a inteligência de produção de conhecimento sobre quais grupos e perfis estão relacionados a um determinado tipo de fraude digital, resultando nos seguintes requisitos para a inteligência:

- Quais são os principais grupos e perfis relacionados ao tipo de fraude (*phishing*) que visa enganar clientes através de sites falsos que imitam empresas?
- Quais são os principais termos mencionados e tópicos discutidos por cada grupo e perfil?

- Qual o grau de similaridade entre os perfis e grupos identificados que tem relação com esse tipo de fraude?

Essas informações solicitadas costumam ser elementos de informação importantes para criar uma base de investigação com o intuito de embasar pedidos de repressão legal aos crimes. É comum que empresas do setor financeiro tenham times internos que atuam com a responsabilidade de realizar investigações preliminares, visando fornecer as forças da lei uma denúncia com maior materialidade, aumentando as chances de sucesso de uma possível repressão. Dessa forma, este caso de uso possibilita a análise do grau de apoio oferecido pelo método nos casos de investigação preliminar para repressão legal.

Portanto, para obter as respostas que embasem essa possível investigação preliminar, foi realizada uma série de etapas relacionadas a consulta de informações e construção de contexto, sendo elas:

- Entedimento das perguntas de interesse e estruturação do plano de consultas para organizar as informações necessárias:
    1. Identificar os tópicos de interesses relacionados ao tipo de fraude requisitado (*phishing*);
    2. Definir consulta para identificar os grupos e perfis que possuem maior grau de relação com os tópicos de interesses selecionados;
    3. Definir consulta para identificar os principais tópicos discutidos por cada grupo e perfil;
    4. Definir consulta para identificar os grupos e perfis que possuem maior similaridade entre si;
- Execução progressiva das consultas, agregando os resultados obtidos aos resultados das consultas anteriores;
- Criação de visualização em grafo com as informações consultadas;
- Disponibilização da visualização para interação de analista humano.

O resultado final obtido pela execução dessas etapas para o caso de uso de monitoramento de grupos e perfis pode ser visualizaod na figura 16. A fim de evitar sobrecarga visual na figura apresentada, a quantidade de perfis e grupos identificados foi limitado a 30, os tópicos de interesse a 5, e os termos relacionados aos tópicos a 10. De forma similar ao caso de uso anterior, essas limitações podem ser relevadas devido a capacidade de gerar visualizações interativas, que permitem que um analista humano navegue pelo grafo sem restrições quanto a resolução de imagem.

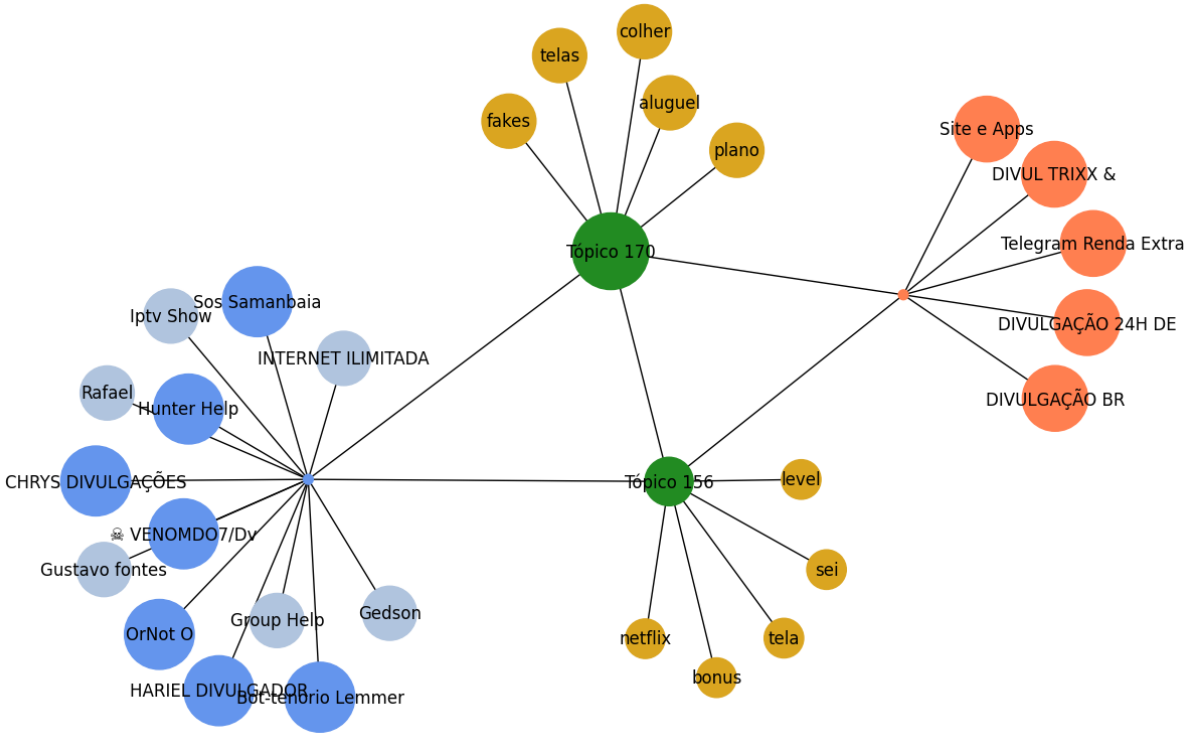

Figura 16 – Resultado do experimento de monitoramento de grupos e perfis.

Dessa forma, o resultado obtido nesse caso de uso foi avaliado como bom, pois consegue fornecer velocidade de construção de conhecimento a partir do nível 1 (percepção) de consicência situacional. Além disso, as informações necessárias para alcançar o nível 2 (compreensão) e nível 3 (projeção) estão estruturadas em formato computacional, facilitando a sua utilização em outras soluções.

5 CONCLUSÕES

O conceito de consciência situacional envolve a compreensão do significado dos elementos percebidos, exigindo a capacidade de responder perguntas sobre as razões pelas quais os elementos estão estabelecidos da maneira como estão ('porquê?'). Dessa forma, consciência situacional presume a existência de uma pessoa capaz de interpretar e compreender os elementos de forma que possa raciocinar e tomar decisões baseadas nesse entendimento.

Portanto, não é possível que um sistema automatizado, mesmo que conte com apoio de inteligência artificial e aprendizado de máquina, consiga alcançar consciência situacional por si só. Entretanto, é viável e recomendado que técnicas de automação e inteligência artificial sejam utilizadas para suportar a criação de consciência situacional, possibilitando que o ser humano utilize de ferramentas que forneçam a capacidade de rapidamente iterar e investigar o ambiente alvo.

Dessa forma, o método proposto neste trabalho demonstrou capacidade de suporte à criação de consciência situacional através da identificação automatizada de elementos do ambiente, disponibilização das suas respectivas característicos de forma estruturada, aplicação de técnicas de aprendizado de máquina para possibilitar a descoberta mais veloz das características do elementos analisados, e organização dos dados para facilitar análises de tendências e previsão de cenários futuros.

Apesar do resultado positivo como suporte viável a criação de consciência situacional, o método proposto sofre de limitações importantes que requerem avaliações cautelosas na implantação:

- Os dados textuais obtidos dos grupos e canais representativos do ecossistema de crimes cibernéticos do Brasil possuem diversas características que degradam a performance de técnicas de análise de tópicos, como: textos curtos, gírias específicas de contexto, linguagem informal, contrações textuais, e escrita incorreta;

- A qualidade do resultado obtido na etapa de análise de tópicos é extremamente influente nos resultados obtidos pelo método. Uma etapa de análise de tópicos com resultados ruins gera um efeito cascata onde os resultados das outras etapas são poluídos com informações imprecisas. Em um cenário que requer dados fidedignos, o método pode ser inviabilizado por causa de baixa qualidade na análise de tópicos;

- Para cenários onde precisão nos resultados da análise de tópicos é crucial, é necessário um trabalho de rotulamento manual dos dados para que técnicas supervisionadas de análise de tópicos possam ser utilizadas.

Por outro lado, mesmo que existam limitações importantes na aplicação deste método em alguns cenários, os resultados obtidos foram positivos, incluindo:

- Capacidade de navegar dados textuais de forma estruturada (grafos)
- Possibilidade de agregar outros dados no modelo de dados (expandir o schema)
- Facilidade de organizar análises de tendência e previsões futuras (dados temporais já organizados)

Além disso, existem algumas outras técnicas que podem ser testadas para avaliar se os resultados da modelagem de tópicos pode ser melhorada:

- Ajuste fino de modelos de linguagem para que generalizem seus resultados para textos com características similares aos observados pelo trabalho, especialmente o entendimento das gírias específicas e a similaridade entre termos;
- Modelagem de tópicos maunal

Dessa forma, as limitações identificadas na aplicação do método devem ser consideradas individualmente por cada necessidade de implantação específica. Em casos onde é requerido que dados sejam fidedignos a realidade (e.g. análise de inteligência), pode ser necessário optar por formas mais conservadoras de análise de tópicos, como modelagem manual ou técnicas supervisionadas. Em outros casos, os resultados obtidos são suficientes para possibilitar a compreensão do ambiente estudado de forma iterativa. Por fim, é possível que o ajuste fino de LLMs (*Large Language Model*) seja capaz de reduzir o impacto das dificuldades encontradas nos textos coletados, possibilitando melhores resultados. Essa abordagem não foi avaliada neste trabalho, mas é recomendada como evolução da tentativa de análise não supervisionada de tópicos.

REFERÊNCIAS

ABIN. **Doutrina da Atividade de Inteligência**. SPO, Brasília - DF, 2023.

AGGARWAL, C. C. *et al.* **Data mining: the textbook**. [*S.l.: s.n.*]: Springer, 2015. v. 1.

AHMAD, A. *et al.* How can organizations develop situation awareness for incident response: A case study of management practice. **Computers & Security**, Elsevier, v. 101, p. 102122, 2021.

ALDRIDGE, J.; DÉCARY-HÉTU, D. Hidden wholesale: The drug diffusing capacity of online drug cryptomarkets. **International Journal of Drug Policy**, Elsevier, v. 35, p. 7–15, 2016.

ALZUBI, J.; NAYYAR, A.; KUMAR, A. Machine learning from theory to algorithms: an overview. *In*: IOP PUBLISHING. **Journal of physics: conference series**. [*S.l.: s.n.*], 2018. v. 1142, p. 012012.

AZEVEDO, A.; SANTOS, M. F. Kdd, semma and crisp-dm: a parallel overview. **IADS-DM**, 2008.

BLEI, D. M. Probabilistic topic models. **Communications of the ACM**, ACM New York, NY, USA, v. 55, n. 4, p. 77–84, 2012.

BONFANTI, M. E. Cyber intelligence: in pursuit of a better understanding for an emerging practice. **Cyber, Intelligence, and Security**, INSS, v. 2, n. 1, p. 105–121, 2018.

BOYD-GRABER, J. *et al.* Applications of topic models. **Foundations and Trends® in Information Retrieval**, Now Publishers, Inc., v. 11, n. 2-3, p. 143–296, 2017.

CHISMON, D.; RUKS, M. Threat intelligence: Collecting, analysing, evaluating. **MWR InfoSecurity Ltd**, v. 3, n. 2, p. 36–42, 2015.

CHOWDHARY, K.; CHOWDHARY, K. Natural language processing. **Fundamentals of artificial intelligence**, Springer, p. 603–649, 2020.

CLARK, R. M. **Intelligence collection**. [*S.l.: s.n.*]: CQ Press, 2013.

CLARK, R. M. Perspectives on intelligence collection. **AFIO's The Intelligencer - Journal of U.S. Intelligence Studies**, v. 20, n. 2, jun. 2013. Available at: <https://www.afio.com/publications/CLARK%20Pages%20from%20AFIO_INTEL_FALLWINTER2013_Vol20_No2.pdf>.

CLARK, R. M. **Intelligence analysis: a target-centric approach**. [*S.l.: s.n.*]: CQ press, 2019.

CYBERSIXGILL. **State of the Cybercrime Underground 2023**. Tel Aviv-Yafo, Israel, 2024. Available at: <https://cybersixgill.com/resources/the-state-of-the-underground-2023>.

DAVIES, P. H.; GUSTAFSON, K.; RIGDEN, I. The intelligence cycle is dead, long live the intelligence cycle: rethinking intelligence fundamentals for a new intelligence doctrine. *In*: **Understanding the intelligence cycle**. [*S.l.: s.n.*]: Routledge, 2013. p. 56–75.

DEVARAJAN, S. *et al.* Enhancing dark web classification: A dynamic crawler and robust classification framework. **International Journal of Intelligent Systems and Applications in Engineering**, v. 12, n. 6s, p. 01–09, 2024.

DOVER, R. Socmint: a shifting balance of opportunity. **Intelligence and National Security**, Routledge, v. 35, n. 2, p. 216–232, 2020. Available at: <https://doi.org/10.1080/02684527.2019.1694132>.

DUPONT, B. *et al.* Darkode: Recruitment patterns and transactional features of "the most dangerous cybercrime forum in the world". **American Behavioral Scientist**, v. 61, n. 11, p. 1219–1243, 2017. Available at: <https://doi.org/10.1177/0002764217734263>.

ENDSLEY, M. R. Toward a theory of situation awareness in dynamic systems. **Human factors**, SAGE Publications Sage CA: Los Angeles, CA, v. 37, n. 1, p. 32–64, 1995.

ENDSLEY, M. R.; BOLTÉ, B.; JONES, D. G. **Designing for situation awareness: An approach to user-centered design**. [*S.l.: s.n.*]: CRC press, 2003.

FAYYAD, U.; PIATETSKY-SHAPIRO, G.; SMYTH, P. From data mining to knowledge discovery in databases. **AI magazine**, v. 17, n. 3, p. 37–37, 1996.

FRANKE, U.; BRYNIELSSON, J. Cyber situational awareness–a systematic review of the literature. **Computers & security**, Elsevier, v. 46, p. 18–31, 2014.

GAETA, A.; LOIA, V.; ORCIUOLI, F. A comprehensive model and computational methods to improve situation awareness in intelligence scenarios. **Applied Intelligence**, Springer, v. 51, n. 9, p. 6585–6608, 2021.

GILL, P. Intelligence, threat, risk and the challenge of oversight. **Intelligence and National Security**, Taylor & Francis, v. 27, n. 2, p. 206–222, 2012.

GILL, P.; PHYTHIAN, M. From intelligence cycle to web of intelligence: Complexity and the conceptualisation of intelligence 1. *In*: **Understanding the intelligence cycle**. [*S.l.: s.n.*]: Routledge, 2013. p. 21–42.

GROOTENDORST, M. Bertopic: Neural topic modeling with a class-based tf-idf procedure. **arXiv preprint arXiv:2203.05794**, 2022.

GUPTA, A.; MAYNARD, S. B.; AHMAD, A. The dark web phenomenon: A review and research agenda. **arXiv preprint arXiv:2104.07138**, 2021.

HAND, D. J.; MANNILA, H.; SMYTH, P. **Principles of Data Mining**. The MIT Press, 2001. (Adaptive Computation and Machine Learning). ISBN 9780262082907. Available at: <https://mitpress.mit.edu/9780262082907/principles-of-data-mining/>.

HOLT, T. J. Examining the forces shaping cybercrime markets online. **Social Science Computer Review**, v. 31, n. 2, p. 165–177, 2013. Available at: <https://doi.org/10.1177/0894439312452998>.

IPESPE - Instituto de Pesquisas Sociais, Políticas e Econômicas. **RADAR FEBRABAN (Dezembro 2023)**. Praça Doutor Fernando Figueira, 30, Ilha do Leite, Recife - PE, Brasil, 2023. Available at: <https://portal.febraban.org.br/noticia/4039/pt-br/>.

JOHNSON, C. *et al.* Guide to cyber threat information sharing. **NIST special publication**, v. 800, n. 150, p. 35, 2016.

JOINT CHIEFS OF STAFF. **JP 2-0, Joint Intelligence**. Washington, DC, USA, 2013. Available at: <https://www.jcs.mil/Doctrine/Joint-Doctrine-Pubs/2-0-Intelligence-Series/>.

KENT, S. **Strategic intelligence for American world policy**. [*S.l.: s.n.*]: Princeton University Press, 2015. v. 2377.

LAMPROPOULOS, A. S.; TSIHRINTZIS, G. A. **Machine learning paradigms**. [*S.l.: s.n.*]: Springer, 2015.

LANDAUER, T. K.; FOLTZ, P. W.; LAHAM, D. An introduction to latent semantic analysis. **Discourse processes**, Taylor & Francis, v. 25, n. 2-3, p. 259–284, 1998.

LEARNING, S.-S. Semi-supervised learning. **CSZ2006. html**, v. 5, p. 2, 2006.

LIDDY, E. D. Enhanced text retrieval using natural language processing. **Bulletin of the American Society for Information Science and Technology**, New Jersey, v. 24, n. 4, p. 14–16, 1998.

LIGGETT, R. *et al.* The dark web as a platform for crime: An exploration of illicit drug, firearm, csam, and cybercrime markets. **The Palgrave handbook of international cybercrime and cyberdeviance**, Springer, p. 91–116, 2020.

LOWENTHAL, M. M.; CLARK, R. M. **The five disciplines of intelligence collection**. [*S.l.: s.n.*]: Sage, 2015.

MAHESH, B. Machine learning algorithms-a review. **International Journal of Science and Research (IJSR).[Internet]**, v. 9, n. 1, p. 381–386, 2020.

MELNYK, O.; BYCHKOVSKY, Y.; VOLOSHYN, A. Maritime situational awareness as a key measure for safe ship operation. **Zeszyty Naukowe. Transport/Politechnika Ślaska**.

MIKOLOV, T. Efficient estimation of word representations in vector space. **arXiv preprint arXiv:1301.3781**, 2013.

MITCHELL, T. M. **Machine Learning**. [*S.l.: s.n.*]: McGraw-Hill Education, 1997.

NADKARNI, P. M.; OHNO-MACHADO, L.; CHAPMAN, W. W. Natural language processing: an introduction. **Journal of the American Medical Informatics Association**, BMJ Group BMA House, Tavistock Square, London, WC1H 9JR, v. 18, n. 5, p. 544–551, 2011.

OMAND, J. B. D.; MILLER, C. Introducing social media intelligence (socmint). **Intelligence and National Security**, Routledge, v. 27, n. 6, p. 801–823, 2012. Available at: <https://doi.org/10.1080/02684527.2012.716965>.

PAULINO, G. d. C. *et al.* **TEĆNICAS AVANÇADAS DE INVESTIGAÇÃO: Perspectivas prat́ica e jurisprudencial**. [*S.l.: s.n.*]: Escola Superior do Ministério Público da União, 2022. ISBN ISBN 978-65-88299-89-0.

PELTON, J. N.; SINGH, I. B. Coping with the dark web, cyber-criminals and techno-terrorists in a smart city. *In*: _____. **Smart Cities of Today and Tomorrow: Better Technology, Infrastructure and Security**. Cham: Springer International Publishing, 2019. p. 171–183. ISBN 978-3-319-95822-4. Available at: <https://doi.org/10.1007/978-3-319-95822-4_11>.

PENNINGTON, J.; SOCHER, R.; MANNING, C. D. Glove: Global vectors for word representation. *In*: **Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)**. [*S.l.: s.n.*], 2014. p. 1532–1543.

PROVOST, F.; FAWCETT, T. **Data Science for Business: What you need to know about data mining and data-analytic thinking**. [*S.l.: s.n.*]: "O'Reilly Media, Inc.", 2013.

REBALA, G. *et al.* Machine learning definition and basics. **An introduction to machine learning**, Springer, p. 1–17, 2019.

SAMTANI, S.; ZHU, H.; CHEN, H. Proactively identifying emerging hacker threats from the dark web: A diachronic graph embedding framework (d-gef). Association for Computing Machinery, New York, NY, USA, v. 23, n. 4, aug 2020. ISSN 2471-2566. Available at: <https://doi.org/10.1145/3409289>.

SAMUEL, A. L. Some studies in machine learning using the game of checkers. **IBM Journal of research and development**, IBM, v. 3, n. 3, p. 210–229, 1959.

SAMUEL, A. L. Some studies in machine learning using the game of checkers. ii—recent progress. **IBM Journal of research and development**, IBM, v. 11, n. 6, p. 601–617, 1967.

SANTOS, Y. C. M. d. M. d.; BOTELHO, D. G. A influência das novas tecnologias no direito penal–desafios e perspectivas. **Revista Ibero-Americana de Humanidades, Ciências e Educação**, v. 9, n. 11, p. 2713–2726, 2023.

SCHULZ, C. M. *et al.* Situation awareness in anesthesia: concept and research. **The Journal of the American Society of Anesthesiologists**, The American Society of Anesthesiologists, v. 118, n. 3, p. 729–742, 2013.

SHAH, D. *et al.* Illicit activity detection in large-scale dark and opaque web social networks. **2020 IEEE International Conference on Big Data (Big Data)**, p. 4341–4350, 2020.

SIA, S.; DALMIA, A.; MIELKE, S. J. Tired of topic models? clusters of pretrained word embeddings make for fast and good topics too! **arXiv preprint arXiv:2004.14914**, 2020.

SILVA, R. M. d. **PROPOSTA DE UM FRAMEWORK PARA MELHORIA DA QUALIDADE NA PRODUÇÃO DE INTELIGÊNCIA DE AMEAÇA CIBERNÉTICA**. 2023.

SILVA, R. M. d. **Proposta de um framework para melhoria da qualidade na produção de inteligência de ameaça cibernética**. 2024.

SOCRADAR. **Brazil Threat Landscape Report**. 254 Chapman Rd, Ste 208 Newark, Delaware 19702 USA, 2023. Available at: <https://socradar.io/resource/brazil-threat-landscape-report/>.

STANDARD, O. Stix version 2.1. jun. 2021. Available at: <https://docs.oasis-open.org/cti/stix/v2.1/os/stix-v2.1-os.html>.

STRACHAN-MORRIS, D. Threat and risk: what is the difference and why does it matter? **Intelligence and National Security**, Taylor & Francis, v. 27, n. 2, p. 172–186, 2012.

SUFI, F. A new social media-driven cyber threat intelligence. **Electronics**, MDPI, v. 12, n. 5, p. 1242, 2023.

ŞUŞNEA, E.; IFTENE, A. The significance of online monitoring activities for the social media intelligence (socmint). *In*: **Conference on mathematical foundations of informatics**. [*S.l.: s.n.*], 2018. p. 230–240.

TANABE, R. **Proposta de um método para inteligência de fontes abertas: valores e princípios para uma atividade ética e profissional**. 2023.

VASWANI, A. Attention is all you need. **Advances in Neural Information Processing Systems**, 2017.

WARNER, M. The past and future of the intelligence cycle. *In*: **Understanding the intelligence cycle**. [*S.l.: s.n.*]: Routledge, 2013. p. 9–20.

WEBB, J. *et al.* A situation awareness model for information security risk management. **Computers & security**, Elsevier, v. 44, p. 1–15, 2014.

WIRTH, R.; HIPP, J. Crisp-dm: Towards a standard process model for data mining. *In*: MANCHESTER. **Proceedings of the 4th international conference on the practical applications of knowledge discovery and data mining**. [*S.l.: s.n.*], 2000. v. 1, p. 29–39.

ZAKI, M. J.; MEIRA, W. **Data mining and analysis: fundamental concepts and algorithms**. [*S.l.: s.n.*]: Cambridge University Press, 2014.

ZHOU, Z.-H. **Three perspectives of data mining**. [*S.l.: s.n.*]: Elsevier, 2003.

ZIBAK, A.; SAUERWEIN, C.; SIMPSON, A. C. Threat intelligence quality dimensions for research and practice. **Digital Threats: Research and Practice**, ACM New York, NY, v. 3, n. 4, p. 1–22, 2022.