VOL.40 Nº2 2023
https://doi.org/10.22491/rmct.v40i2.9271.pt

O Impacto da Ponderação de Recursos Semânticos para Identificação de Suspeitos de Crimes em Redes Sociais

Érick S. Florentino[1], erick.florentino@ime.eb.br, Orcid 0000-0002-0828-4058
Ronaldo R. Goldschmidt[1], ronaldo.rgold@ime.eb.br, Orcid 0000-0003-1688-0586
Maria Cláudia Cavalcanti, yoko@ime.eb.br, Orcid 0000-0003-4965-9941
[1]Instituto Militar de Engenharia – IME

RESUMO: A identificação de pessoas suspeitas de crimes em redes sociais tem sido um tema de grande relevância na análise desse tipo de rede. Na maioria das vezes, os métodos que buscam identificar esses suspeitos utilizam dados textuais disponibilizados pelas pessoas nessas redes (e.g. mensagens, comentários, entre outros). Para analisar os textos, tais métodos costumam utilizar recursos semânticos como vocabulários controlados ou até mesmo simples conjuntos compostos por termos, de acordo com o domínio em questão (e.g. terrorismo, pedofilia, entre outros). A menção de um ou mais desses termos pode levantar suspeitas sobre as pessoas que os utilizaram. No entanto, há termos que levantam mais suspeitas do que outros. Assim sendo, este trabalho busca investigar o impacto da diferenciação do nível de periculosidade dos termos utilizados por um método de identificação de suspeitos de crimes em redes sociais e se isso pode levar a melhores resultados na identificação dos suspeitos. Os resultados obtidos por meio de experimentos no domínio da pedofilia mostraram que a diferenciação do nível de periculosidade dos termos proporcionou melhores resultados em 82.5% dos experimentos realizados.

PALAVRAS-CHAVE: Suspeitos. Redes Sociais, Recursos Semânticos.

ABSTRACT: The identification of criminal suspects on social media has been a topic of great relevance in the analysis of this type of media. Most of the time, the methods that seek to identify these suspects use textual data made available by people on these networks (e.g. messages, comments, among others). To analyze the texts, these methods often use semantic resources such as controlled vocabularies or even simple sets composed of terms, according to the domain in question (e.g. terrorism, pedophilia, among others). The mention of one or more of these terms can raise suspicions about the people who have used them. However, some terms raise more suspicion than others. Therefore, this work seeks to investigate the impact of differentiating the level of dangerousness of the terms used by a method for identifying criminal suspects on social media and whether this can lead to better results in identifying suspects. The results obtained through experiments in the domain of pedophilia showed that differentiating the level of dangerousness of the terms provided better results in 82.5% of the experiments carried out.

KEYWORDS: Suspects. Social Media, Semantic Resources.

1. Introdução

Redes sociais (e.g., X - antigo Twitter, YouTube, Instagram, Facebook, entre outras) fazem parte do cotidiano da grande maioria da sociedade [1]. Diariamente, um grande volume de dados é disponibilizado nessas redes por meio de diversas funcionalidades, tais como compartilhamento de vídeos e troca de mensagens [2]. Além disso, essas redes permitem interações em tempo real, sem ter o espaço geográfico como limitação [3]. Devido a isso, a *Análise de Redes Sociais*[1] tem sido de grande interesse para instituições públicas e privadas para os mais diferentes fins [5].

Uma das tarefas de Análise de Redes Sociais que tem tido grande relevância nos últimos anos é a identificação de pessoas suspeitas de crimes em redes sociais (ex. pedofilia, bullying, terrorismo) [6] [7] [8]. Isso se deve ao crescente número de pessoas que têm utilizado os recursos existentes nas redes a fim de praticar atos que podem trazer riscos a outras pessoas, tanto externamente quanto internamente a esses ambientes virtuais [9] [10]. Por exemplo, tais atos podem gerar

1 Expressão dada a qualquer conjunto de atividades que busque extrair conhecimento sobre os indivíduos que utilizam as redes sociais [4].

algum tipo de impacto psicológico e/ou físico nas pessoas [4].

Na literatura, uma parte significativa dos métodos que buscam identificar pessoas que praticam crimes em redes sociais se baseia em análises do conteúdo textual disponibilizado pelas pessoas [11] [12] [7]. Essa análise, por diversas vezes, conta com o apoio de um vocabulário controlado ou um conjunto composto por termos comumente usados por pessoas suspeitas no domínio da aplicação [11] [13]. Esses vocabulários ou conjunto de termos podem conter expressões com diferentes níveis de "periculosidade[2]". Existem métodos que buscam diferenciar esses níveis [14] [15] [16] e outros que não [12] [11] [17]. Diante desse cenário, surgem as seguintes questões de pesquisa: *Qual o impacto da diferenciação dos níveis de periculosidade dos termos suspeitos de um vocabulário ou conjunto de termos? Tal diferenciação pode levar a melhores resultados na identificação de pessoas suspeitas de crimes em redes sociais?*

A fim de buscar evidências que respondam aos questionamentos acima, o trabalho descrito neste artigo realizou experimentos com o método INSPECTION[3] [14], considerando cinco conjuntos de dados, de duas maneiras. Na primeira, todos os termos foram ponderados com um único peso, ou seja, sem diferenciar o seu nível de periculosidade. Já na segunda, diferenciou-se o nível de periculosidade de cada grupo de termos. A partir dos experimentos realizados, foi possível verificar que a ponderação dos níveis de periculosidade dos termos do vocabulário controlado levou a melhores resultados em 82.5% desses experimentos, respondendo positivamente às questões de pesquisa levantadas.

O restante do trabalho está organizado como se segue. Na Seção 2, são apresentados alguns conceitos básicos relacionados à análise de redes socais e dados não estruturados (textos). Já na Seção 3, são apresentados alguns trabalhos que buscam identificar pessoas suspeitas de crimes por meio do uso de vocabulários controlados ou conjuntos de termos de acordo com o domínio da aplicação. O método INSPECTION, bem como suas etapas, são descritos de forma resumida na Seção 4. Detalhes sobre os experimentos realizados e sobre os resultados obtidos estão na Seção 5. Por fim, na Seção 6, são destacadas as contribuições deste trabalho e possíveis trabalhos futuros.

2. Conceitos básicos

Na análise de redes sociais, é fundamental compreender um indivíduo de maneira mais profunda. Para isso, podem ser extraídas três principais informações da rede [18]: Topológica, Temporal e Contextual.

A informação topológica, fundamentada na estrutura do grafo que representa a rede, permite identificar como os diferentes nós (indivíduos ou entidades) estão interconectados [19]. Isso revela padrões de interação e a importância relativa de cada nó na rede [20]. Em algumas redes sociais, no entanto, essa topologia não é tão evidente, sendo necessário utilizar ferramentas para identificar as interconexões entre os indivíduos. Um exemplo disso é o Youtube. O algoritmo TROY [21] [16] é uma ferramenta que explicita as interações entre usuários do Youtube. Por meio da análise das respostas a comentários, o algoritmo identifica quem "recebeu" ou para quem foi "enviado" um dado conteúdo textual, gerando, assim, um multigrafo com as interações.

Já a informação temporal, que envolve a análise dos dados ao longo do tempo, foca a cronologia das interações [22]. Entender quando determinadas ações ocorrem pode fornecer insights valiosos sobre o comportamento dos indivíduos e as dinâmicas sociais [20].

A informação contextual considera o ambiente social e cultural, além das intenções nas comunicações [14]. Embora seja relevante, frequentemente

2 Periculosidade é uma expressão que refere-se à qualidade de algo que apresenta risco ou perigo. No contexto atual, pessoas e/ou palavras suspeitas de acordo com um domínio podem ser qualificadas quanto ao seu nível de perigo.

3 Método que busca identificar pessoas suspeitas de crimes em redes sociais por meio de conteúdos textuais. Para isso, é utilizado um vocabulário controlado composto por termos de acordo com domínio da aplicação (e.g. pedofilia, etc.). Nesse método, esses termos são ponderados e normalizados conforme o nível de periculosidade no domínio em questão. E, posteriormente, as pessoas são ponderadas de acordo com a utilização dos termos pertencentes a esse vocabulário em seus textos.

aparece como dados não estruturados como mensagens e comentários, que são difíceis de analisar [23] [24]. Técnicas como Mineração de Texto e Anotação Semântica ajudam a transformar esses dados em conhecimento acionável, permitindo uma análise mais profunda das redes sociais e interações humanas [25] [14].

A mineração de texto é uma técnica poderosa que permite extrair conhecimento útil a partir de grandes volumes de dados textuais, utilizando uma variedade de recursos e abordagens [26]. Essa extração é realizada por meio de ferramentas e técnicas computacionais que tornam possível identificar padrões, tendências e informações relevantes em textos [27]. Entre as diversas metodologias utilizadas na mineração de texto, destacam-se técnicas de *Machine Learning*, que permitem a modelagem e previsão de comportamentos a partir de dados; e análises estatísticas, que ajudam a quantificar e interpretar informações; e o modelo Bag-of-Words, que representa documentos como conjuntos de palavras, facilitando a análise textual de maneira simplificada [28].

Já a anotação semântica de textos utiliza recursos que vão além da simples análise textual, incorporando elementos semânticos que enriquecem a compreensão do conteúdo [29]. Entre os recursos disponíveis, destaca-se o Vocabulário Controlado, também conhecido como Tesauro, que é composto por uma lista estruturada de termos e expressões que descrevem um determinado domínio de conhecimento e tem a função de organizar e padronizar a terminologia utilizada, facilitando a busca e a recuperação de informações específicas [30].

Em comparação, as ontologias oferecem uma abordagem mais rica e complexa, pois não apenas definem termos, mas também estabelecem hierarquias e relacionamentos entre os conceitos, permitindo uma representação dinâmica e interconectada do conhecimento, o que facilita a compreensão das relações entre os termos e promove uma análise mais aprofundada dos dados [31]. Tanto a mineração de texto quanto a anotação semântica são essenciais na extração e organização do conhecimento a partir de informações textuais, contribuindo significativamente para a análise de dados em várias áreas.

3. Trabalhos relacionados

Nesta seção, serão apresentados alguns trabalhos que utilizam um vocabulário controlado ou conjunto de termos para possibilitar análise contextual de dados textuais da rede (e.g. mensagens, comentários, entre outros), a fim de identificar diferentes tipos de pessoas suspeitas de crimes em redes sociais (e.g. pedofilia, assédio e cyberbullying).

Bretschneider et al. [12] desenvolveram um método que busca identificar mensagens de assédio online. Esse método faz uso de um conjunto pré-definido de palavras que podem ser classificadas como de alta periculosidade no domínio em questão, para apoiar a identificação de mensagens com teor suspeito. Sendo assim, a partir desse conjunto, são verificadas as mensagens que se enquadram em padrões textuais. Esses padrões avaliam a conexão entre uma pessoa (e.g. pronomes pessoais, nomes de usuários, entre outros) e uma palavra suspeita (ou profana) pertencente ao conjunto pré-definido. Dessa maneira, mensagens que se enquadram nesses padrões são marcadas como de assédio. Por outro lado, Bretschneider e Peters [11], na busca de pessoas que praticam cyberbullying utilizam o mesmo método proposto por Bretschneider et al. [12]. Mas Bretschneider e Peters [11] enriquecem o método com a análise topológica por considerarem o comportamento relacional das pessoas. Desse modo, pessoas são consideradas suspeitas caso tenham enviado um certo número de mensagens com teor suspeito para uma mesma pessoa.

Elzinga et al. [17] desenvolveram um método não automatizado que utiliza um sistema semântico relacional temporal para analisar conversas com teores pedófilos em salas de bate-papo ao longo do tempo. Para isso, Elzinga et al. [17] criaram 6 (seis) categorias em que uma mensagem pode ser classificada: Onde, Quando, Partes Íntimas, Manipulações Sexuais, Fotos e Câmeras, Elogios. Dessa forma, manualmente, cada mensagem é analisada e verificada em qual categoria

se enquadra. Feito isso, posteriormente, dada uma pessoa com suas mensagens categorizadas e por meio do sistema semântico relacional temporal, é possível visualizar as categorias em que uma pessoa percorre na rede ao longo do tempo, bem como o seu comportamento suspeito.

Os trabalhos apresentados consideram um conjunto de palavras em que todas são classificadas como perigosas e não diferenciam o grau de periculosidade. No entanto, nesse conjunto, há termos que são comumente usados por pessoas que não praticam nenhum tipo de crime em redes sociais e há também termos que são mais incomuns e que podem levantar mais suspeitas. Em Elzinga et al. [17],por exemplo, que abordam o domínio de pedofilia, o grupo de palavras da categoria "Partes Íntimas" (ex. pênis, seios, entre outros) pode ser considerado mais perigoso do que o grupo da categoria "Onde" (ex. praia, cinema, entre outras).

4. O método inspection e suas etapas

Esta seção apresenta um breve resumo do método INSPECTION, originalmente proposto e ilustrado em [14]. Este método requer como entrada um conjunto de dados $N_{v_n} = [M, U]$, em que M é um conjunto de mensagens enviadas e/ou recebidas por pessoas de um conjunto U, e é composto por 5 (cinco) etapas: Preparação de Dados, Representação da Rede, Ponderação do Vocabulário Controlado, Análise Contextual e Identificação de Pessoas Suspeitas. As próximas subseções sumarizam essas etapas.

4.1 Preparação de dados

Nesta etapa, são utilizadas técnicas de mineração de texto com o objetivo de, para cada mensagem $m_x \in M$, normalizar cada termo t_j existente em m_x. Para isso, essa etapa é composta por 3 (três) subetapas: *Normalização e Extração de Conteúdo Textual*, *Remoção de Stop Words* e *Stemming*. Na subetapa de *Normalização e Extração de Conteúdo Textual*, devido à informalidade existente na escrita em redes sociais, busca-se remover letras repetidas, tratar abreviaturas, entre outros. Já as subetapas *Remoção de Stop Words* e *Stemming* são responsáveis, respectivamente, por remover palavras $t_j \in m_x$ com pouco significado e identificar o radical dos termos. Ao final, é gerado o conjunto de mensagens tratadas M'.

4.2 Representação da rede

Na etapa de Representação da Rede, dois multigrafos são construídos a partir de M' a fim de representar a rede de duas formas. Na primeira, são representadas as pessoas e suas mensagens. Para isso, é utilizado um multigrafo homogêneo dirigido $G_{PP}(V_{PP}, E_{PP})$, em que cada $v_{PP} \in V_{PP}$ representa uma pessoa e cada $e_{PP} \in E_{PP}$ representa uma mensagem, a qual permite identificar a pessoa que enviou e/ou recebeu essa mensagem. Além disso, cada e_{PP} tem um atributo T, responsável por armazenar o conteúdo textual de uma mensagem $m_x \in M$. Já na segunda forma de representação da rede, por meio de um multigrafo heterogêneo dirigido, $G_{PT}(V_P \cup V_T, E_{PT} \cup E_{TP})$, são representadas as Pessoas e os Termos utilizados em mensagens. Nesse caso, os vértices V_P e V_T, respectivamente, representam as pessoas e os termos utilizados em mensagens ($t_j \in m_x$). Já as arestas dirigidas E_{PT} e E_{TP} são responsáveis por informar, respectivamente, a pessoa que enviou e a pessoa que recebeu um determinado termo.

4.3 Ponderação do vocabulário controlado

Esta subetapa utiliza um vocabulário controlado $O = [C, R]$ de acordo com o domínio da aplicação (e.g. Terrorismo, Pedofilia, Bullying, entre outros). Esse vocabulário é composto por um conjunto de termos C e um conjunto de relações entre termos R. O conjunto de termos divide-se em dois subconjuntos disjuntos, ie, $C = [C_r, C_s]$, em que C_r tem termos mais genéricos (denominados classes raízes) e C_s contém termos específicos (chamados de subclasses). Sabendo-se que cada classe apresenta um atributo w, responsável por armazenar o peso da classe, o vocabulário é ponderado em duas fases. Na primeira fase, conta-se com o apoio de um especialista, no domínio em questão, para ponderar as classes raízes ($c_r \in C_r$).

Uma vez ponderadas as classes raízes, a segunda fase busca ponderar as subclasses ($c_s \in C_s$). Essas subclasses passam pelas subetapas *Normalização e Extração de Conteúdo Textual* e *Stemming* da etapa de *Preparação de Dados* da mesma forma como descrito anteriormente para os termos das mensagens. Ao final dessa fase, é gerado um conjunto de subclasses processadas denominado C_s'. Com as subclasses tratadas ($c_s' \in C_s'$), são verificadas quais delas estão presentes na rede a ser analisada. Para tanto, é aplicada um filtro por meio da seguinte operação $A = C_s' \cap V_T$. Em seguida, é calculado o peso global (GW_{t_j}'), para cada $t_j' \in A$ (Eq. 1).

$$GW_{t_j}' = \log 2\left(\frac{|E_{pp}|}{|n_{t_j}'|}\right) \quad (1)$$

Na equação 1, $|E_{PP}|$ e $|n_{t_j}'|$ indicam, respectivamente, o número total de arestas (ou número de mensagens) em G_{PP} e o número de arestas em E_{PP} que tenham o termo t'_j em mensagens (ie, $n_{t_j}' = \{e_{ppi} | t_j' \in e_{ppi}.T\}$). Desse modo, o peso global GW_{t_j}', é responsável por expressar a raridade de cada termo.

Identificada a raridade de cada termo, ela deverá ser normalizada de acordo com a classe raiz, a qual o respectivo termo está vinculado. Desse modo, é possível diferenciar o nível de periculosidade de cada termo. Para isso, é calculado *HGW* ($HGW = \log 2\left(\frac{E_{PP}}{1}\right)$), uma vez que a normalização é dada de acordo com o peso global máximo. Em seguida, é verificada a representatividade em termos percentuais do termo t_j' por meio de GW_{t_j}' (Eq. 2).

$$GW_{t_j'}^{\%} = \frac{GW_{t_j'^{*1}}}{HGW} \quad (2)$$

Os valores dessa representatividade são utilizados para calcular o peso do termo normalizado. Tendo-se esses valores, é utilizada $GW_{t_j'}^{N}$ (Eq. 3) para obter o Peso Global final para cada termo, dentro do intervalo da sua categoria.

$$GW_{t_j'}^{N} = \left(\left(Max\left(C_{r_k}\right) - Min\left(C_{r_k}\right)\right) \times GW_{t_j'}^{\%}\right) + Min\left(C_{r_k}\right) \quad (3)$$

Para isso, é calculado o intervalo para cada c_r, ($Max\ (C_{r_j}) = C_{r_j}.w$ e o $Min\ (C_{r_j}) = Max\ (\{C_{r_i}.w | C_{r_i} \in C_r - \{C_{r_j}\} \wedge C_{r_i}.w < C_{r_j}.w\} \cup \{0\}$. Tendo-se o valor obtido por meio de $GW_{t_j'}^{N}$ e sabendo-se que $t_j' = C_s'$, esse valor é atribuído ao atributo w da subclasse do termo correspondente (ie, $c_s'.w = GW_{t_j'}^{N}$).

4.4 Análise contextual

Esta etapa busca identificar um *score* para cada pessoa na rede, a fim de representar o seu grau de suspeição no contexto do domínio da aplicação. Para isso, para cada pessoa são resgatados todos os termos utilizados por ela $C_{v_T}(v_p)$, onde $C_{v_T}(v_p) = \{v_T | \exists(v_P, v_T) \in E_{PT}\}$. Todavia, nesse resgate, podem existir termos que não são suspeitos, ou seja, não são considerados importantes no domínio da aplicação. Para isso, aplica-se um filtro $C_{v_T}^{\cap}(v_p)$, em que $C_{v_T}^{\cap}(v_p) = C_{v_T}(v_p) \cap C_s'$, com o objetivo de obter apenas os termos suspeitos utilizados pela pessoa

Para cada pessoa v_p que tenha utilizado termos suspeitos (ie, $C_{v_T}^{\cap} | (vP)| > 0$), são utilizadas as métricas M_{GW} e M_{FGW}. A métrica M_{GW}, por meio da Equação 4, realiza a soma de todos os pesos globais normalizados dos termos suspeitos utilizados por uma pessoa. Sendo assim, essa métrica, apresenta a soma da raridade de cada termo suspeito, normalizado pelo peso de uma classe raiz a qual esse termo esteja ligado.

$$M_{GW}(v_P) = \sum_{C_{s_i} \in C_{v_T}^{\cap}(v_P)} C_{s_i}.w \quad (4)$$

Já na métrica M_{GW}, a frequência de cada termo suspeito utilizado por uma pessoa é considerada. Sendo assim, essa frequência é multiplicada pelo o seu respectivo peso global normalizado, indicando a importância desse termo para uma pessoa com relação a todas as mensagens analisadas. Posteriormente, os valores obtidos são somados. Para isso, utiliza-se a Equação 5.

$$M_{FGW}(v_P) = \sum_{C_{s_i} \in C_{v_T}^{\cap}(v_P)} W(v_P, C_{s_i}) \times C_{s_i}.w \quad (5)$$

Dessa maneira, $W(v_P, C_{s_i})$ resgata a frequência do uso de um determinado termo suspeito utilizado por uma pessoa. Formalmente, $W(v_P, C_{s_i}) = |\{(o,t) \in E_{PT}/o = v_P \text{ e } t = C_{s_i}\}|$. Com isso, selecionada uma das métricas, o *score* obtido com ela é aplicado à pessoa analisada ($v_P.st = M_{GW}(v_P)$ ou $v_P.\text{st} = M_{FGW}(v_P)$), expressando numericamente o comportamento suspeito dela.

4.5 Identificação de suspeitos

Nesta última etapa, é feita uma ordenação decrescente pelos *scores* obtidos com cada uma das métricas acima, de tal forma que aquelas pessoas mais suspeitas se encontram no topo da lista.

5. Experimentos e resultados

Esta seção apresenta o conjunto de dados utilizados e reporta os experimentos realizados com o objetivo de responder às questões de pesquisa apresentadas na Seção 1. A escolha do tema pedofilia deveu-se ao fato de as redes sociais terem se tornado um grande atrativo para crianças e adolescentes, fazendo com que esse público tenha participação assídua nesses ambientes [32]. Por isso, é comum que pedófilos utilizem as redes para identificar potenciais vítimas [4].

Em relação ao protótipo do método INSPECTION [14], ele foi desenvolvido em Python 3.0. Na etapa de preparação dos dados, foram utilizadas as bibliotecas NLTK [33], Spacy [34], entre outras. Além disso, foi criado um dicionário de gírias e abreviaturas, comumente usadas em redes sociais em português, para tratar possíveis ruídos no texto. A construção desse dicionário foi feita por meio de pesquisas em sites relacionados [35] [36] [37]. As próximas seções apresentam os detalhes dos experimentos.

Nas próximas subseções, serão apresentados detalhes dos conjuntos de dados (subseção 5.1) e vocabulários controlados (subseção 5.2) utilizados nos experimentos.

5.1 Conjunto de Dados

Nos experimentos aqui realizados, cinco conjuntos de dados foram aplicados ao método INSPECTION. Tais conjuntos foram construídos a partir de comentários e respostas em português extraídos de cinco vídeos do Youtube, de um canal pertencente a uma cantora menor de idade. A escolha dos vídeos, bem como do canal informado, deve-se ao fato de terem uma maior propensão de conter conteúdos textuais suspeitos no tema escolhido para os experimentos. A Tabela 1 apresenta dados estatísticos dos 5 (cinco) vídeos.

Tabela 1 - Dados Estatísticos de cada vídeo *vn* utilizado nos experimentos.

Vídeo	Duração	Visualizações	Comentários e Respostas
v_1	2M:38S	2.551.258	6.897
v_2	3M:14S	57.083	348
v_3	2M:53S	13.041.367	13.080
v_4	2M:56S	18.157.216	18.548
v_5	4M:04S	89.593.403	71.387

Conforme mencionado na Seção 4, o INSPECTION requer como entrada um conjunto de dados $N_{v_n} = [M, U]$. Assim, cada um dos vídeos listados na Tabela 1 foi submetido ao Algoritmo TROY, permitindo a extração das interações entre as pessoas; detalhes desse algoritmo podem ser vistos em [21] e [16]. Esses dados resultantes foram utilizados nos experimentos apresentados neste trabalho e estão resumidos na Tabela 2. O processo de construção desses conjuntos contou ainda com uma etapa de enriquecimento dos dados responsável por incorporar

informações sobre suspeitos de pedofilia existentes no conjunto de dados PAN-2012-BR [39] [7]. O PAN-2012-BR é um conjunto de dados que contém conversas de 39 (trinta e nove) pedófilos, as quais foram disponibilizadas pelo Ministério Público Federal de São Paulo (MPF-SP). A integração desses suspeitos aos conjuntos de dados da Tabela 2 foi feita por meio da tarefa de predição de links, conforme descrito detalhadamente em [16].

Tabela 2 - Dados estatísticos dos conjuntos de dados gerados a partir dos vídeos da Tabela 1.

Vídeo	N_{v_n}		Média de Interações	|U| Suspeitas	|M| Suspeitas
	|U|	|M|			
v_1	1.806	10.647	3	39	1.752
v_2	87	382	2	20	132
v_3	3.688	16.332	3	39	1.484
v_4	5.255	18.488	3	39	1.694
v_5	18.802	75.249	3	39	1.280

5.2 Vocabulário Controlado

O vocabulário controlado foi construído tomando como base seis categorias de palavras, inspiradas em [17]: "onde", "quando", "partes íntimas", "manipulações sexuais", "fotos e câmera" e "elogios". Cada categoria tornou-se uma classe raiz a partir da qual foram adicionadas subclasses extraídas de vocabulários encontrados na literatura (vide Tabela 3). Essas classes e suas subclasses passaram a formar um único vocabulário denominado *O*. Além dessas classes raízes, foi considerada ainda a classe raiz roupa. Assim, para os experimentos com o método INSPECTION foram adotadas 4 (quatro) variações desse vocabulário, como em [16]: O_1^{INT} (sem a classe raiz roupa e ponderado com números inteiros), O_2^{INT} (com a classe raiz roupa e ponderado com números inteiros), O_1^{REAL} (sem a classe raiz roupa e ponderado com números reais) e O_2^{REAL} (com a classe raiz roupa e ponderado com números reais).

É importante ressaltar que, nos vocabulários e , a classe raiz "roupa" é considerada, pois é comum que pessoas suspeitas de pedofilia perguntem sobre as vestimentas de suas possíveis vítimas. Ademais, isso permite a avaliação do desempenho do método com o vocabulário controlado enriquecido por essa classe, tanto com ponderação quanto sem. Além disso, possibilita a comparação do método ponderado de maneira mais rigorosa (*INT*) e mais flexível (*FLOAT*) em relação ao vocabulário não ponderado.

Na Tabela 3, são apresentadas as ponderações das classes mais genéricas dos vocabulários. É válido ressaltar que, para a ponderação do vocabulário controlado, houve o apoio de um policial federal de Aracaju/SE que trabalha com o tema da pedofilia há 11 (onze) anos.

Tabela 3 - Ponderações dos vocabulários controlados e origens das subclasses.

c_r	O_1^{INT}	O_2^{INT}	O_1^{REAL}	O_2^{REAL}	C_s
	Peso (w)	**Peso (w)**	**Peso (w)**	**Peso (w)**	
Quando	1	1	1.5	1.5	[40]
Onde	2	2	2.3	2.3	[41]
Elogios	3	3	4.0	4.0	[42]
Fotos e Câmera	4	5	6.0	6.0	[43]
Partes Íntimas	5	6	5.7	5.7	[44]
Manipulações Sexuais	6	7	5.5	5.5	[45]
Roupa	-	4	-	5.0	[46]

Para os experimentos com o INSPECTION sem a ponderação do vocabulário controlado, todos os termos receberam peso 1 (um) (ie, $C_s'.w = 1$). Com isso, não houve diferenciação no nível de periculosidade dos termos. Dessa maneira, criou-se um referencial para permitir a comparação com os resultados gerados com o INSPECTION a partir da diferenciação do nível de periculosidade dos termos em função da ponderação do vocabulário controlado.

5.3 Resultados

Considerando as variações em relação aos conjuntos de dados, aos vocabulários e suas ponderações, assim como as métricas contextuais, ao todo, o método INSPECTION foi executado 60 (sessenta) vezes. Os resultados dessas execuções encontram-se resumidos na Tabela 4. Para avaliar o desempenho do método INSPECTION, a medida utilizada foi a AUC (Área sob a Curva) [47]. Essa medida calcula a probabilidade de um suspeito ter um $score$ superior a um não suspeito, ambos escolhidos n vezes de maneira aleatória. Neste trabalho, n é igual a 100. É válido ressaltar que o método INSPECTION utiliza o rótulo das pessoas (suspeita e não suspeita) apenas para avaliar o desempenho do método.

Em uma análise geral do método INSPECTION [14] com e sem a ponderação do vocabulário, observa-se que todas as execuções do método levaram a resultados superiores ao preditor randômico (AUC >0,5), o que sinaliza para uma boa a capacidade do método em identificar suspeitos de pedofilia, independente do uso ou não de ponderação.

Tabela 4 - Resultados dos experimentos.

Vídeo	**VC**	**INSPECTION C/ PONDERAÇÃO**		**INSPECTION S/ PONDERAÇÃO**	
		M_{FGW}	M_{GW}	M_{FGW}	M_{GW}
v_1	O_1^{INT}	0,905	0,905	0,900	0,865
	O_2^{INT}	0,885	0,890		
	O_1^{REAL}	0,910	0,890	0,910	0,835
	O_2^{REAL}	0,915	0,865		

Vídeo	VC	INSPECTION C/ PONDERAÇÃO		INSPECTION S/ PONDERAÇÃO	
		M_{FGW}	M_{GW}	M_{FGW}	M_{GW}
v_2	O_1^{INT}	0,795	0,845	0,600	0,545
	O_2^{INT}	0,810	0,920		
	O_1^{REAL}	0,835	0,835	0,535	0,525
	O_2^{REAL}	0,840	0,920		
v_3	O_1^{INT}	0,930	0,905	0,900	0,895
	O_2^{INT}	0,930	0,890		
	O_1^{REAL}	0,900	0,915	0,855	0,840
	O_2^{REAL}	0,930	0,925		
v_4	O_1^{INT}	0,940	0,930	0,910	0,885
	O_2^{INT}	0,950	0,965		
	O_1^{REAL}	0,945	0,950	0,945	0,915
	O_2^{REAL}	0,930	0,950		
v_5	O_1^{INT}	0,940	0,885	0,905	0,885
	O_2^{INT}	0,925	0,950		
	O_1^{REAL}	0,930	0,925	0,925	0,900
	O_2^{REAL}	0,920	0,915		

Em uma análise mais detalhada, das 40 (quarenta) comparações de desempenho entre o método INSPECTION com e sem ponderação do vocabulário, o método ponderado obteve vitória em 33 (trinta e três casos) (82,5%), empatou em 3 (três) (7,5%, destacados em negrito e azul na Tabela 4) e perdeu em 4 (quatro) (10%, destacados em vermelho e itálico na Tabela 4). Esses resultados sugerem uma resposta positiva à questão de pesquisa deste artigo, que investiga se a ponderação do vocabulário (relativa ao nível de periculosidade dos termos) pode contribuir para aprimorar a identificação de suspeitos de crimes em redes sociais, especificamente no contexto da pedofilia.

Detalhando ainda mais os resultados do método INSPECTION com a ponderação do vocabulário, observa-se que os melhores desempenhos foram alcançados com o vocabulário O_1^{INT}, o qual obteve sucesso em todos os 5 (cinco) conjuntos de dados, empatando apenas no conjunto de dados v_3 com a métrica M_{GW}. Isso mostra que o enriquecimento do vocabulário com novas classes raízes deve ser feito cuidadosamente. E ainda, que a ponderação das classes raízes de forma mais rígida proporcionou melhores resultados. Já em relação às métricas, a M_{GW} resultou em melhores resultados (dezoito de vinte comparações – 90%). Com isso, a frequência em que a pessoa usou termos suspeitos não é tão relevante.

Em suma, a partir dos resultados apresentados, pode-se perceber que ao ponderar o vocabulário controlado é possível obter melhores resultados no que tange à identificação de pessoas suspeitas de crimes de pedofilia. Em outras palavras, o impacto da ponderação dos termos mostrou-se positivo no processo.

6. Considerações Finais

As redes sociais têm estado presentes cada vez mais no dia a dia da sociedade, atraindo os mais diferentes públicos com as mais diferentes particularidades. Dessa forma, vêm se tornando um meio propício para que pessoas com más intenções pratiquem atos ilícitos na rede. Assim, para evitar riscos à integridade física e psicológica de indivíduos nas redes sociais, a identificação de pessoas suspeitas tem tido grande destaque.

Na literatura, muitos métodos que buscam identificar pessoas suspeitas de crimes em redes sociais utilizam um vocabulário ou conjunto de termos de acordo com o domínio da aplicação. Desse modo, torna-se possível analisar os dados textuais disponibilizados por uma pessoa nas redes sociais a fim de verificar o quanto ela pode ser suspeita. Contudo, dentro de um domínio de aplicação, podem existir termos com diferentes níveis de periculosidade. Diante disso, neste trabalho, foram levantadas as seguintes questões de pesquisa: *Qual o impacto da diferenciação dos níveis de periculosidade dos termos suspeitos de um vocabulário ou conjunto de termos? Se tal diferenciação pode levar a melhores resultados na identificação de pessoas suspeitas de crimes em redes sociais.*

Com a finalidade de responder aos questionamentos acima, foram feitos experimentos com o método INSPECTION , em 5 (cinco) conjuntos de dados, sem e com a ponderação do vocabulário controlado. Os resultados obtidos por meio dos experimentos no domínio da pedofilia mostraram que o impacto da ponderação do vocabulário controlado é positivo e, consequentemente, leva a melhores resultados (82.5% dos experimentos realizados neste trabalho). Dessa maneira, a diferenciação dos níveis de periculosidade dos termos em recursos semânticos (e.g. vocabulários controlados, conjunto de termos, entre outros), para identificação de pessoas suspeitas de crimes em redes sociais, é altamente relevante.

Como trabalhos futuros, destacam-se: a ponderação do vocabulário controlado sem depender do auxílio de um especialista no domínio da aplicação e a realização de experimentos em outras redes sociais e domínios.

Referências

[1] SILVA, C. R. M.; TESSAROLO, F M. 2016. Influenciadores digitais e as redes sociais enquanto plataformas de mídia. *In*: Congresso Brasileiro de Ciências da Comunicação, 39., 2016, São Paulo. *Anais* […]. São Paulo: Intercom, 2016.

[2] BENEVENUTO, F.; ALMEIDA, J. M.; SILVA, A. S. *Explorando redes sociais online*: da coleta e análise de grandes bases de dados às aplicações. Porto Alegre: Sociedade Brasileira de Computação, 2011.

[3] TAKHTEYEV, Y.; GRUZD, A., WELLMAN, B. Geography of Twitter networks. *Social networks*, Amsterdam, n. 34, v. 1, 73-81, 2012.

[4] DAS, B.; SAHOO, J. S. Social networking sites – a critical analysis of its impact on personal and social life. *International Journal of Business and Social Science*, United States of America, v. 2, n. 14, p. 222-228, 2011.

[5] LÉVY, P.; FEROLDI, D. *Cybercultura*: gli usi sociali delle nuove tecnologie. ITÁLIA: Feltrinelli, 1999.

[6] Andressa Olivetti Costa. 2019. *Ciberterrorismo*. 2019. Trabalho de conclusão de curso (Bacharelado em Direito) – Centro Universitário Antônio Eufrásio de Toledo de Presidente Prudente, Presidente Prudente, 2019.

[7] SANTOS, L. F.; GUEDES, G. Identificação de predadores sexuais brasileiros em conversas textuais na internet por meio de aprendizagem de máquina. *iSys-Brazilian Journal of Information Systems*, Rio de Janeiro, v. 13, n. 4, p. 22-47, 2020.

[8] LEI, Y.; HUANG, B. Prediction of Criminal Suspect Characteristics with Application of Wavelet Neural Networks. *Applied Mathematics and Nonlinear Sciences*, Boston, 2023.

[9] MANN, B. L. Social networking websites–a concatenation of impersonation, denigration, sexual aggressive solicitation, cyber-bullying or happy slapping videos. *International Journal of Law and Information Technology*, Oxford, v. 17, v. 3, p. 252-267, 2009.

[10] SINGH, M.; SINGH, A. How Safe You Are on Social Networks? *Cybernetics and Systems*, Oxford, v. 54, n. 7, p. 1154-1171, 2023.

[11] BRETSCHNEIDER, U.; PETERS, R. Detecting cyberbullying in online communities. *In*: European Conference on Information Systems, 24., Istanbul, 2016. *Anais* […]. Istanbul, 2014.
[12] BRETSCHNEIDER, U.; WÖHNER, T.; PETERS, R. Detecting Online Harassment in Social Networks. *In*: International Conference on Information Systems, 35., Auckland, 2014. *Anais* […]. Auckland, 2014.
[13] FLORENTINO, E. S.; GOLDSCHMIDT, R. R.; CAVALCANTI, M. C. R. Exploring Interactions in YouTube to Support the Identification of Crime Suspects. *In*: Anais do Simpósio Brasileiro de Sistemas de Informação, 17., Uberlândia, 2021. *Anais* […[. Uberlândia: SBC, 2021.
[14] FLORENTINO, E. S.; GOLDSCHMIDT, R. R; CAVALCANTI, M. C. R. Identifying Suspects on Social Networks: An Approach based on Non-structured and Nonlabeled Data. *In*: Proceedings of the International Conference on Enterprise Information Systems, 23., 2021, Setúbal. *Anais* […]. Setúbal: ICEIS, 2021. Disponível em: https://www.scitepress.org/Link.aspx?doi=10.5220/0010440300510062. Acesso em: 22 abr. 2025.
[15] FLORENTINO, E. S.; GOLDSCHMIDT, R. R.; CAAVALCANTI, M. C. Identifying Criminal Suspects on Social Networks: A Vocabulary-Based Method. In: Proceedings of the Brazilian Symposium on Multimedia and the Web, 20., 2020, São Luís. *Anais* […]. São Luís: SIGWEB, 2020.
[16] FLORENTINO, E. S.; GOLDSCHMIDT, R. R.; CAVALCANTI, M. C. Identificando Suspeitos de Crimes por meio de Interações Implícitas no YouTube. *iSys - Revista Brasileira de Sistemas de Informação*, Rio de Janeiro, v. 15, n. 1, p. :36, 2022.
[17] ELZINGA, P.; WOLFF, K. E.; POELMANS, J. Analyzing chat conversations of pedophiles with temporal relational semantic systems. *European Intelligence and Security Informatics Conference*, [*s. l.*], 2012, p. 242-249.
[18] MUNIZ, C. P. M. T. *Investigando a utilização de atributos temporais no problema de predição de links*. 2016. Dissertação (Mestrado em Ciências em Sistemas e Computação) – Instituto Militar de Engenharia, Rio de Janeiro, 2016.
[19] CHEN, Y.; COSKUNUZER, B.; GEL, Y. Topological relational learning on graphs. *Advances in neural information processing systems*, Cambridge, v. 34, p. 27029-27042, 2021.
[20] MUNIZ, C. P.; GOLDSCHMIDT, R.; CHOREN, R. Combining contextual, temporal and topological information for unsupervised link prediction in social networks. *Knowledge-Based Systems*, Amsterdam, v. 156, p. 129-137, 2018.
[21] FLORENTINO, E.S.; GOLDSCHMIDT, R. R.; CAVALCANTI, M. C. R. Exploring Interactions in YouTube to Support the Identification of Crime Suspects. *In*: Anais do Simpósio Brasileiro de Sistemas de Informação, 17., Uberlândia, 2021. Anais […[. Uberlândia: SBC, 2021.
[22] FLORENTINO, E. S.; CAVALCANTE, A. A. B.; GOLDSCHMIDT, R. R. An edge creation history retrieval based method to predict links in social networks. *Knowledge-Based Systems*, Amsterdam, v. 205, p. 106268, 2020.
[23] BAARS, H.; KEMPER, H-G. Management support with structured and unstructured data—an integrated business intelligence framework. *Information Systems Management*, Abingdon, v. 25, n. 2, p. 132-148, 2008.
[24] BERRY, M. W. *Automatic discovery of similar words*. *Survey of Text Mining*: Clustering, Classification and Retrieval. New York: Springer Verlag, 2004.
[25] CARVALHO, R. C. *Aplicação de técnicas de mineração de texto na recuperação de informação clínica em prontuário eletrônico do paciente*. 2017. Dissertação (Mestrado em Ciência da Informação) – Universidade Estadual Paulista, Marília, 2017.
[26] SULOVA, S. Text mining approach for identifying research trends. *In*: International Conference on Computer Systems and Technologies, 21., 2021, Ruse. Anais […]. Ruse: University of Ruse, 2021.
[27] MORAIS, E. A. M.; AMBRÓSIO, A. P. L. *Mineração de textos*. Relatório Técnico–Instituto de Informática. Goiás: Instituto de Informática Universidade Federal de Goiás, 2007.
[28] TAN, A-H. et al. 1999. Text mining: The state of the art and the challenges. *Proceedings of the PAKDD*, 1999.
[29] OLIVEIRA, H. C.; CARVALHO, C. L. *Gestão e representação do conhecimento*. Goiás: UFG, 2008.
[30] SALES, R.; CAFÉ, L. Diferenças entre tesauros e ontologias. *Perspectivas em Ciência da Informação*, Belo Horizonte, v. 14, n. 1, p. 99-116, 2009.
[31] CHANDRASEKARAN, B.; JOSEPHSON, J. R.; BENJAMINS, V. R. What are ontologies, and why do we need them? *IEEE Intelligent Systems and their applications*, Piscataway, v. 14, n. 1, p. 20-26, 1999.
[32] FERNÁNDEZ, A. Clinical Report: The impact of social media on children, adolescents and families. *Archivos de Pediatría del Uruguay*, v. 82, n. 1, p. 31-32, 2011.
[33] BIRD, S.; KLEIN, E.; LOPER, E. Natural language processing with Python: analyzing text with the natural language toolkit. Sebastopol, na Califórnia. Sebastopol: O'Reilly Media, 2009.

[34] HONNIBAL, M. *et al. spaCy*: Industrial-strength Natural Language Processing in Python. Disponível em: https://zenodo.org/records/10009823. Acesso em: 22 abr. 2025.
[35] Por Da Redação. 2014. SQN, LOL? Entenda as principais expressões e hashtags das redes sociais. *Aconteceu no Vale*. [s. l.], 15 fev. 2014. Disponível em: https://aconteceunovale.com.br/portal/?p=21357#google_vignette. Acesso em: 19 maio 2025.
[36] thecoolcapybara List of internet slangs - gírias da internet. *Reddit*. 20 maio 2020. Disponível em: https://www.reddit.com/r/Portuguese/comments/gn8s1y/list_of_internet_slangs_g%C3%ADrias_da_internet/. Acesso em: 22 abr. 2025.
[37] VEJA lista de abreviações usadas pelos jovens em troca de mensagens na internet. *Gshow*. [*s. l.*], 30 jun. 2021. Disponível em: https://gshow.globo.com/programas/mais-voce/noticia/veja-lista-de-abreviacoes-usadas-pelos-jovens-em-troca-de-mensagens-na-internet.ghtml. Acesso em: 22 abr. 2025.
[38] Ligga Telecom. Gírias nos games e seus significados: quais as mais usadas. https://liggavc.com.br/blog/entretenimento/glossario-conheca-as-girias-mais-usadas-na-internet-e-nos-games/.
[39] ANDRIJAUSKAS, A.; SHIMABUKURO A; MAIA R. F. 2017. DESENVOLVIMENTO DE BASE DE DADOS EM LÍNGUA PORTUGUESA SOBRE CRIMES SEXUAIS. *In*: Simpósio de Iniciação Científica, Didática e Ações Sociais da FEI, 7., 2017. Anais [...]. [s. l.], 2017.
[40] Scheider, Simon, and Peter Kiefer. "(Re-) localization of location-based games." *Geogames and geoplay: game-based approaches to the analysis of geo-information*. Cham: Springer International Publishing, 2017. 131-159. SCHEIDER, S.; KIEFER, P. (Re-) Localization of Location-Based Games. In: AHLQVIST, O.; SCHILIEDER, C. (Eds.). Geogames and Geoplay: Game-based Approaches to the Analysis of Geo-Information. Berlim: Springer. p. 131-159.
[41] Jerry R Hobbs and Feng Pan. Time ontology in OWL. *W3C working draft*, Arlington, 2006.
[42] NEVES, F. Elogios de A a Z. *Dicio Dicionário Online de Português*. [s. l.], [201-?]. Disponível em: https://www.dicio.com.br/elogios-de-a-a-z/. Acesso em: 22 abr. 2025.
[43] MUKHERJEE, S.; JOSHI, S. Sentiment aggregation using ConceptNet ontology. In: MITKOV, R.; PARK, J. C. (Eds.). *Proceedings of the Sixth International Joint Conference on Natural Language Processing*. Nagoya: Nagoya Editora. p. 570-578.
[44] ROSSE C.; MEJINO, J. L. V. The foundational model of anatomy ontology. In: BURGER, A.; DAVIDSON, D.; BALDOCK, R. *Anatomy Ontologies for Bioinformatics*. Heidelberg: Heidelberg Springer. p. 59-117.
[45] KRONK C.; TRAN, G. Q.; WU, D. T. Y. Creating a Queer Ontology: The Gender, Sex, and Sexual Orientation (GSSO) Ontology. *Studies in health technology and informatics*, Amsterdam, v. 264, p. 208-212, 2019.
[46] KUANG Z. et al. Integrating multi-level deep learning and concept ontology for large-scale visual recognition. *Pattern Recognition*, Amsterdam, v. 78, p. 198-214, 2018.
[47] LI S. et al. Similarity-based future common neighbors model for link prediction in complex networks. *Scientific reports*, London, v. 8, n. 1, 1-11, 2018.