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

The Impact of Assigning Weights to Semantic Resources on the Identification of Criminal Suspects on Social Media

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

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.

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.

1. introduction

Social media (e.g., X - formerly Twitter, YouTube, Instagram, Facebook, among others) are part of everyday life for the vast majority of society [1]. Every day, a large amount of data is made available in these networks through various functionalities, such as sharing videos and exchanging messages [2]. In addition, these networks allow for real-time interactions, without geographical space being a limitation [3]. Because of this, *Social Network Analysis*[1] has been of great interest to public and private institutions for a wide variety of purposes [5].

One of the Social Network Analysis tasks that has been of great relevance in recent years is the identification of individuals suspected of crimes on social media (e.g. pedophilia, bullying, terrorism) [6] [7] [8]. This is due to the growing number of people who have used the resources available on networks to carry out acts that can pose risks to other people, both externally and internally to these virtual environments [9] [10]. For example, such acts can have some kind of psychological and/or physical impact on people [4].

[1] Term given to any set of activities that seek to extract knowledge about individuals who use social networks [4].

In the literature, a significant part of the methods that seek to identify individuals who commit crimes on social networks are based on analyzing the textual content provided by people [11] [12] [7]. This analysis is often supported by a controlled vocabulary or a set of terms commonly used by suspicious people in the application domain [11] [13]. These vocabularies or sets of terms can contain expressions with different levels of "dangerousness[2]". There are methods that seek to differentiate these levels [14] [15] [16] and others that do not [12] [11] [17]. Given this scenario, the following research questions arise: *What is the impact of differentiating the dangerousness levels of suspicious terms in a vocabulary or set of terms? Can such differentiation lead to better results in identifying individuals suspected of crimes on social networks*?

In order to find evidence to answer the above questions, the study described in this article carried out experiments with the INSPECTION method[3] [14], considering five sets of data, in two ways. In the first, all the terms were assigned the same weight, i.e., without differentiating their level of dangerousness. In the second, the level of dangerousness of each group of terms was differentiated. Based on the experiments conducted, it was possible to observe that weighting the dangerousness levels of the controlled vocabulary terms led to better results in 82.5% of the experiments, thus positively addressing the research questions raised.

The rest of the paper is organized as follows. Section 2 presents some basic concepts related to social network analysis and unstructured data (texts). Section 3 presents some studies that seek to identify crime suspects through the use of controlled vocabularies or sets of terms according to the application domain. The INSPECTION method and its stages are briefly described in Section 4. Details of the experiments conducted and the results obtained are in Section 5. Finally, Section 6 highlights the contributions of this study and possible future research.

2. Basic concepts

In social network analysis, it is essential to gain a deeper understanding of an individual. To do this, three main pieces of information can be extracted from the network [18]: Topological, Temporal and Contextual.

Topological information, based on the structure of the graph that represents the network, makes it possible to identify how the different nodes (individuals or entities) are interconnected [19]. This reveals interaction patterns and the relative importance of each node in the network [20]. In some social networks, however, this topology is not so obvious and it is necessary to use tools to identify the interconnections between individuals. An example of this is Youtube. The TROY algorithm [21] [16] is a tool that reveals the interactions between YouTube users. By analyzing responses to comments, the algorithm identifies who "received" or to whom a given textual content was "sent", thus generating a multigraph of interactions.

On the other hand, temporal information, which involves analyzing data over time, focuses on the chronology of interactions [22]. Understanding when certain actions take place can provide valuable insights into the behavior of individuals and social dynamics [20].

Contextual information considers the social and cultural environment, as well as intentions in communications [14]. Although it is relevant, it often appears as unstructured data such as messages and comments, which are difficult to analyze [23] [24]. Techniques such as Text Mining and Semantic Annotation help transform these data into actionable

[2] Dangerousness is a term that refers to the quality of something that poses a risk or danger. In today's context, people and/or words that are suspicious according to a domain can be classified according to their level of danger.

[3] A method aimed at identifying people suspected of crimes on social media through textual content. To do this, it uses a controlled vocabulary made up of terms according to the application domain (e.g. pedophilia, etc.). In this method, these terms are weighted and normalized according to their level of dangerousness in the domain in question. Subsequently, people are weighted according to the use of terms belonging to this vocabulary in their texts.

knowledge, enabling a deeper analysis of social networks and human interactions [25] [14].

Text mining is a powerful technique that allows useful knowledge to be extracted from large volumes of textual data, using a variety of resources and approaches [26]. This extraction is carried out using computational tools and techniques that make it possible to identify patterns, trends and relevant information in texts [27]. Among the various methodologies used in text mining are Machine Learning techniques, which allow for the modeling and prediction of behavior from data; statistical analyses, which help to quantify and interpret information; and the Bag-of-Words model, which represents documents as sets of words, facilitating textual analysis in a simplified manner [28].

Semantic annotation of texts uses resources that go beyond simple textual analysis, incorporating semantic elements that enrich understanding of the content [29]. Among the resources available is the Controlled Vocabulary, also known as a Thesaurus, which is made up of a structured list of terms and expressions that describe a particular domain of knowledge and has the function of organizing and standardizing the terminology used, facilitating the search and retrieval of specific information [30].

In comparison, ontologies offer a richer and more complex approach, as they not only define terms, but also establish hierarchies and relationships between concepts, allowing for a dynamic and interconnected representation of knowledge, which makes it easier to understand the relationships between terms and promotes a more in-depth analysis of the data [31]. Both text mining and semantic annotation are essential in extracting and organizing knowledge from textual information, making a significant contribution to data analysis in various areas.

3. Related work

In this section, we will present some studies that use a controlled vocabulary or set of terms to enable contextual analysis of textual data from the network (e.g. messages, comments, among others), in order to identify different types of people suspected of crimes on social networks (e.g. pedophilia, harassment and cyberbullying).

Bretschneider et al. [12] developed a method that seeks to identify online harassment messages. This method uses a pre-defined set of words that can be classified as highly dangerous in the domain in question to help identify messages with suspicious content. From this set, messages that fit into textual patterns are checked. These patterns assess the connection between a person (e.g. personal pronouns, user names, etc.) and a suspicious (or profane) word belonging to the predefined set. In this way, messages that fit these standards are marked as harassing. On the other hand, Bretschneider and Peters [11] use the same method proposed by Bretschneider et al. [12] to find people who practice cyberbullying. However, Bretschneider and Peters [11] enrich the method with topological analysis by considering people's relational behavior. In this way, people are considered suspicious if they have sent a certain number of messages with suspicious content to the same person.

Elzinga et al. [17] developed a non-automated method that uses a temporal relational semantic system to analyze conversations with pedophilic content in chat rooms over time. To do this, Elzinga et al. [17] created 6 (six) categories into which a message can be classified: Where, When, Intimate Parts, Sexual Handlings, Cams and Pictures, Compliments. In this way, each message is manually analyzed and checked to see which category it falls into. Once this is done, given a person with their messages categorized and by means of the temporal relational semantic system, it is possible to visualize the categories in which a person moves through the network over time, as well as their suspicious behavior.

The studies presented consider a set of words in which all are classified as dangerous and do not differentiate the degree of dangerousness. However, in this set, there are terms that are commonly used by people who do not commit any kind of crime on social networks and there are also terms that are more unusual and may raise more suspicion. In Elzinga et al. [17], for example, which addresses the pedophilia domain, the group of words in the "Intimate Parts" category (e.g. penis, breasts, among others) can be considered more dangerous than the group in the "Where" category (e.g. beach, movie theater, among others).

4. The inspection method and its stages

This section presents a brief summary of the INSPECTION method, originally proposed and illustrated in [14]. This method requires as input a data set $N_{v_n} = [M, U]$, where M is a set of messages sent and/or received by people from a set U, and consists of 5 (five) stages: *Terms Preparation*, *Representation of the Network*, *Controlled Vocabulary Weighting*, Contextual *Analysis and Suspect Identification*. The following subsections summarize these steps.

4.1 Terms preparation

In this stage, text mining techniques are used with the aim of normalizing each term t_j on m_x for each message $m_x \in M$. To this end, this stage consists of 3 (three) sub-stages: *Normalization and Extraction of Textual Content, Stop Words Removal* and *Stemming*. In the *Normalization and Extraction of Textual Content* sub-stage, due to textual informality on social networks, the aim is to remove repeated letters, deal with abbreviations, among other things. The *Stop Words Removal* and *Stemming* sub-steps are responsible, respectively, for removing words $t_j \in m_x$ with little meaning and identifying the radical of terms. At the end, the set of treated messages M' is generated.

4.2 Network representation

In the Representation of the Network stage, two multigraphs are constructed from M' in order to represent the network in two ways. In the first, people and their messages are represented. For this, a directed homogeneous multigraph G_{PP} (V_{PP}, E_{PP}) is used, where each $v_{PP} \in V_{PP}$ represents a person and each $e_{PP} \in E_{PP}$ represents a message, which makes it possible to identify the person who sent and/or received that message. In addition, each e_{PP} has an attribute T, responsible for storing the textual content of a message $m_x \in M$. In the second way of representing the network, by means of a directed heterogeneous multigraph, G_{PT} $(V_P \cup V_T E_{PT} \cup E_{TP})$, the People and Terms used in messages are represented. In this case, vertices V_P and V_T, respectively, represent people and the terms used in messages $(t_j \in m_x)$. The directed edges E_{PT} and E_{TP} are responsible for informing, respectively, the person who sent and the person who received a certain term.

4.3 controlled vocabulary Weighting

This sub-stage uses a controlled vocabulary $O = [C, R]$ according to the domain of the application (e.g. Terrorism, Pedophilia, Bullying, among others). This vocabulary is constituted of a set of terms C and a set of relationships between terms R. The set of terms is divided into two disjoint subsets, i.e., $C = [C_r, C_s]$, where C_r contains more generic terms (called root classes) and C_s contains specific terms (called subclasses). Knowing that each class has a w attribute, responsible for storing the weight of the class, the vocabulary is weighted in two phases. In the first phase, a specialist in the domain in question is used to weight the root classes $(c_r \in C_r)$.

Once the root classes have been weighted, the second phase seeks to weight the subclasses $(c_s \in C_s)$. These subclasses go through the *Normalization and Extraction of Textual Content* and *Stemming* sub-steps of the *Data Preparation* stage in the same way as described above for the message terms. At the end of this phase, a set of processed subclasses is generated called. With the subclasses processed $(c_s' \in C_s')$, we check which of them are present in the network to be analyzed. To do this, a filter is applied using the following operation $A = C_s' \cap VT$. The global weight is then calculated (), of each $t_j' \in A$ (Eq. 1).

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

In equation 1, $|E_{PP}|$ and $|\mathrm{n}'_{\mathrm{t_j}}|$ indicate, respectively, the total number of edges (or number of messages) in G_{PP} and the number of edges in E_{PP} that have the term t'_j in messages (i.e., $\mathrm{n}'_{\mathrm{t_j}} = \{e_{ppi} | \mathrm{t}'_{\mathrm{j}} \in e_{ppi}.T\}$). In this way, the global weight $\mathrm{GW}_{\mathrm{t_j}}'$ is responsible for expressing the rarity of each term.

Once the rarity of each term has been identified, it must be normalized according to the root class to which the respective term is linked. In this way, it is possible to differentiate the level of dangerousness of each term. To do this, HGW ($HGW = \log 2\left(\frac{\mathrm{E_{PP}}}{1}\right)$) is calculated, since the normalization is given according to the maximum global weight. Next, the representativeness in percentage terms of the term is checked using $\mathrm{GW}_{\mathrm{t_j}}'$ (Eq. 2).

$$\mathrm{GW}_{\mathrm{t}'_{\mathrm{j}}}^{\%} = \frac{\mathrm{GW}_{\mathrm{t}'_{\mathrm{j}}}^{*1}}{\mathrm{HGW}} \tag{2}$$

The values of this representativeness are used to calculate the weight of the normalized term. Once these values have been calculated, $\mathrm{GW}_{\mathrm{t}'_{\mathrm{j}}}^{\mathrm{N}}$ (Eq. 3) is used to obtain the final Global Weight for each term, within the range of its category.

$$\mathrm{GW}_{\mathrm{t}'_{\mathrm{j}}}^{\mathrm{N}} = \left(\left(\mathrm{Max}\left(\mathrm{C}_{\mathrm{r_k}}\right) - \mathrm{Min}\left(\mathrm{C}_{\mathrm{r_k}}\right)\right) \times \mathrm{GW}_{\mathrm{t}'_{\mathrm{j}}}^{\%}\right) + \mathrm{Min}\left(\mathrm{C}_{\mathrm{r_k}}\right) \tag{3}$$

This is done by calculating the interval for each c_r ($Max\ (\mathrm{C}_{\mathrm{r_j}}) = \mathrm{C}_{\mathrm{r_j}}.w$ e o $Min\ (\mathrm{C}_{\mathrm{r_j}}) = Max\ (\{\mathrm{C}_{\mathrm{r_i}}.w | \mathrm{C}_{\mathrm{r_i}} \in C_r - \{\mathrm{C}_{\mathrm{r_j}}\} \wedge \mathrm{C}_{\mathrm{r_i}}.w < \mathrm{C}_{\mathrm{r_j}}.w\} \cup \{0\}$. Once we have the value obtained from $\mathrm{GW}_{\mathrm{t}'_{\mathrm{j}}}^{\mathrm{N}}$ and we know that $\mathrm{t}'_{\mathrm{j}} = \mathrm{C}'_{\mathrm{s}}$, this value is assigned to the w attribute of the subclass of the corresponding term (ie., $\mathrm{c}'_{\mathrm{s}}.w = \mathrm{GW}_{\mathrm{t}'_{\mathrm{j}}}^{\mathrm{N}}$)

4.4 Contextual analysis

This stage seeks to identify a score for each person in the network, in order to represent their degree of suspicion in the context of the application domain. To do this, all the terms used by each person are retrieved ($C_{v_T} v_P$), where ($C_{v_T} v_P$) $= \{v_{\mathrm{T}} \mid \exists (v_{\mathrm{P}}, v_{\mathrm{T}}) \in E_{PT}\}$. However, in this retrieval, there may be terms that are not suspicious, i.e., they are not considered important in the domain of the application. To do this, a filter $C^{\cap}_{v_T}(v_P)$ is applied, where $(C^{\cap}_{v_T} v_P) = C_{v_T} \cap (v_P)\ \mathrm{C}'_{\mathrm{s}}$, in order to obtain only the suspicious terms used by the person v_P.

For each personwho has used suspicious terms (i.e. $|(C^{\cap}_{v_T} vP)| > 0$), the metrics M_{GW} and M_{FGW} are used. The metric M_{GW}, using Equation 4, is the sum of all the normalized global weights of the suspicious terms used by a person. Thus, this metric presents the sum of the rarity of each suspect term, normalized by the weight of a root class to which that term is linked.

$$\mathrm{M_{GW}}\left(\upsilon_{\mathrm{P}}\right) = \sum_{\mathrm{C_{s_i}} \in \mathrm{C}^{\cap}_{\upsilon_{\mathrm{T}}}(\upsilon_{\mathrm{P}})} \mathrm{C_{s_i}.w} \tag{4}$$

In the metric M_{GW}, the frequency of each suspect term used by a person is taken into account. This frequency is then multiplied by its respective normalized global weight, indicating the importance of this term for a person in relation to all the messages analyzed. The values obtained are then added together. To do this, Equation 5 is used.

$$\mathrm{M_{FGW}}\left(\upsilon_{\mathrm{P}}\right) = \sum_{\mathrm{C_{s_i}} \in \mathrm{C}^{\cap}_{\upsilon_{\mathrm{T}}}(\upsilon_{\mathrm{P}})} \mathrm{W}\left(\upsilon_{\mathrm{P}}, \mathrm{C_{s_i}}\right) \times \mathrm{C_{s_i}.w} \tag{5}$$

In this way, $W(v_P, \mathrm{C}_{\mathrm{s_i}})$ retrieves the frequency of use of a given suspect term by a person. Formally, $W(v_P, \mathrm{C}_{\mathrm{s_i}}) = |\{(o,t) \in E_{PT} / o = v_P$ e $t = \mathrm{C}_{\mathrm{s_i}}\}|$. Once one of the metrics has been selected, the score obtained with it is applied to the person analyzed ($v_P.st = M_{GW}(v_P)$ ou $v_P.\mathrm{st} = M_{FGW}(v_P)$), expressing their suspicious behavior numerically.

4.5 Suspect Identification

At this last stage, the scores obtained with each of the above metrics are listed in a descending order, so that the most suspicious people are at the top of the list.

5. Experiments and results

This section presents the set of data used and reports on the experiments carried out with the aim of answering the research questions presented in Section 1. The topic of pedophilia was chosen because social networks have become a major attraction for children and teenagers, making this audience participate assiduously in these environments [32]. For this reason, it is common for pedophiles to use social media to identify potential victims [4].

The prototype of the INSPECTION method [14] was developed in Python 3.0. In the terms preparation stage, the NLTK [33] and Spacy [34] libraries, among others, were used. In addition, a dictionary of slang and abbreviations commonly used on social networks in Portuguese was created to deal with possible noise in the text. This dictionary was built by searching related websites [35] [36] [37]. The following sections present the details of the experiments.

In the next subsections, details of the data sets (subsection 5.1) and controlled vocabularies (subsection 5.2) used in the experiments will be presented.

5.1 Data sets

In the experiments carried out here, five data sets were applied to the INSPECTION method. These sets were constructed from comments and responses in Portuguese extracted from five YouTube videos, from a channel belonging to an underage singer. The videos and channel were chosen because they were more likely to contain suspicious textual content on the topic chosen for the experiments. Table 1 shows the statistical data for the five (5) videos.

Table 1 - Statistical data for each video vn used in the experiments.

Video	Duration	Views	Comments and Responses
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

As mentioned in Section 4, INSPECTION requires as input a data set $N_{v_n} = [M, U]$. Thus, each of the videos listed in Table 1 was submitted to the TROY algorithm, allowing the extraction of interactions between people; details of this algorithm can be seen in [21] and [16]. The resulting data was used in the experiments presented in this paper and is summarized in Table 2. The process of building these sets also included a data enrichment stage responsible for incorporating information on suspected pedophiles from the PAN-2012-BR dataset [39] [7]. PAN-2012-BR is a dataset containing the conversations of 39 (thirty-nine) pedophiles, which were made available by the São Paulo Federal Public Prosecutor's Office (MPF-SP). These suspects were integrated into the datasets in Table 2 using the link prediction task, as described in detail in [16].

Table 2 - Statistical data from the datasets generated from the videos in Table 1.

Video	N_{v_n}		Mean of Interactions	|U| Suspects	|M| Suspects
	|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 Controlled vocabulary

The controlled vocabulary was built based on six categories of words, inspired by [17]: “where”, “when”, “intimate parts”, “sexual handlings”, “cams and pictures” and “compliments”. Each category became a root class from which subclasses extracted from vocabularies found in the literature were added (see Table 3). These classes and their subclasses now form a single vocabulary called *O*. In addition to these root classes, the root class clothing was also considered. Thus, for the experiments with the INSPECTION method, 4 (four) variations of this vocabulary were adopted, as in [16]: O_1^{INT} (without the clothing root class and weighted with integers), O_2^{INT} (with the clothing root class and weighted with integers), O_1^{REAL} (without the clothing root class and weighted with real numbers) and O_2^{REAL} (with the clothing root class and weighted with real numbers).

It is important to note that in the vocabularies O_2^{INT} and O_2^{REAL}, the root class “clothing” is taken into account, as it is common for people suspected of pedophilia to ask about the clothing of their potential victims. This also allows to evaluate the performance of the method with the controlled vocabulary enriched by this class, both with and without weighting. It also makes it possible to compare the weighted method more rigorously (*INT*) and more flexibly (*FLOAT*) in relation to the unweighted vocabulary.

Table 3 shows the weightings of the most generic vocabulary classes. It is worth noting that, for the weighting of the controlled vocabulary, there was the support of a federal police officer from Aracaju/SE who has been working on the subject of pedophilia for 11 (eleven) years.

Table 3 - Weightings of the controlled vocabularies and origins of the subclasses.

c_r	O_1^{INT}	O_2^{INT}	O_1^{REAL}	O_2^{REAL}	C_s
	Weight (w)	Weight (w)	Weight (w)	Weight (w)	
When	1	1	1.5	1.5	[40]
Where	2	2	2.3	2.3	[41]
Compliments	3	3	4.0	4.0	[42]
Cams and Pictures	4	5	6.0	6.0	[43]
Intimate parts	5	6	5.7	5.7	[44]
Sexual Handlings	6	7	5.5	5.5	[45]
Clothing	-	4	-	5.0	[46]

For the INSPECTION experiments without weighting the controlled vocabulary, all the terms were given a weight of 1 (one) (ie, $C_s'.w = 1$). As a result, there was no differentiation in the level of dangerousness of the terms. In this way, a benchmark was created to allow comparison with the results generated with INSPECTION, based on differentiating the level of dangerousness of the terms according to the weighting of the controlled vocabulary.

5.3 Results

Considering the variations in data sets, vocabularies and their weightings, as well as contextual metrics, the INSPECTION method was run 60 (sixty) times in total. The results of these runs are summarized in Table 4. To assess the performance of the INSPECTION method, the measure used was AUC (Area under the Curve) [47]. This measure calculates the probability of a suspect having a higher score than a non-suspect, both chosen n times at random. In this study, n is equal to 100. It is worth noting that the INSPECTION method uses the label of people (suspect and non-suspect) only to evaluate the performance of the method.

In a general analysis of the INSPECTION method [14] with and without vocabulary weighting, it can be seen that all runs of the method led to results higher than the random predictor (AUC >0.5), which indicates that the method has a good ability to identify suspected pedophiles, regardless of whether or not weighting is used.

Table 4 - Results of the experiments.

Video	VC	INSPECTION WITH WEIGHTING		INSPECTION W/O WEIGHTING	
		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		
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		

Video	VC	INSPECTION WITH WEIGHTING		INSPECTION W/O WEIGHTING	
		M_{FGW}	M_{GW}	M_{FGW}	M_{GW}
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		

In a more detailed analysis, of the 40 (forty) performance comparisons between the INSPECTION method with and without vocabulary weighting, the weighted method won in 33 (thirty-three cases) (82.5%), tied in 3 (three) (7.5%, highlighted in bold and blue in Table 4) and lost in 4 (four) (10%, highlighted in red and italics in Table 4). These results suggest a positive response to the research question of this article, which investigates whether vocabulary weighting (relative to the level of dangerousness of the terms) can help improve the identification of suspected criminals on social networks, specifically in the context of pedophilia.

Further detailing the results of the INSPECTION method with vocabulary weighting, it can be seen that the best performances were achieved with the O_1^{INT} vocabulary, which was successful in all five (5) datasets, only tying in the dataset with the M_{GW} metric. This shows that enriching the vocabulary with new root classes must be done carefully. Furthermore, weighting the root classes more rigidly provided better results. As for the metrics, yielded better results (eighteen out of twenty comparisons - 90%). Therefore, the frequency with which a person used suspicious terms is not so relevant.

In short, from the results presented, it can be seen that weighting the controlled vocabulary gives better results when it comes to identifying people suspected of pedophile crimes. In other words, the impact of weighting the terms proved to be positive in the process.

6. Final considerations

Social media have been increasingly present in society's daily life, attracting the most different audiences with the most different particularities. In this way, they have become a favorable medium for people with bad intentions to commit illegal acts on the web. Therefore, in order to avoid risks to the physical and psychological integrity of individuals on social media, the identification of suspicious individuals has been a major focus.

In the literature, many methods that seek to identify people suspected of crimes on social media use a vocabulary or set of terms according to the application domain. In this way, it becomes possible to analyze the textual data provided by a person on social networks in order to check how suspicious they might be. However, within an application domain, there may be terms with different levels of dangerousness. In view of this, the following research questions were raised in this study: *What is the impact of differentiating the dangerousness levels of suspicious terms in a vocabulary or set of terms? Whether such differentiation can lead to better results in identifying people suspected of crimes on social media.*

In order to answer the above questions, experiments were conducted with the INSPECTION method on 5 (five) sets of data, without and with the weighting of the controlled vocabulary. The results obtained through the experiments in the pedophilia domain showed that the impact of weighting the controlled vocabulary is positive and consequently leads to better results (82.5% of the experiments carried out in this work). In this way, differentiating the levels of dangerousness of terms in semantic resources (e.g. controlled vocabularies, sets of terms, among others) for identifying people suspected of crimes on social networks is highly relevant.

Future work should include: weighting the controlled vocabulary without relying on the help of a specialist in the application domain and conducting experiments on other social networks and domains.

References

[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. SCHEI-

DER, 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.