

Identifying Suspects on Social Networks: An Approach based on Non-structured and Non-labeled Data

Érick S. Florentino^a, Ronaldo R. Goldschmidt^b and Maria C. Cavalcanti^c


Defense Engineering and Computer Engineering Departments, Military Institute of Engineering - IME,
Praça Gen. Tibúrcio 80, Rio de Janeiro, Brazil


Keywords: Suspects Identification, Social Network Analysis, Controlled Vocabulary.


Abstract: The identification of suspects of committing virtual crimes (e.g., pedophilia, terrorism, bullying, among others) has become one of the tasks of high relevance when it comes to social network analysis. Most of the time, analysis methods use the supervised machine learning (SML) approach, which requires a previously labeled set of data, i.e., having identified in the network, the users who are and who are not suspects. From such a labeled network data, some SML algorithm generates a model capable of identifying new suspects. However, in practice, when analyzing a social network, one does not know previously who the suspects are (i.e., labeled data are rare and difficult to obtain in this context). Furthermore, social networks have a very dynamic nature, varying significantly, which demands the model to be frequently updated with recent data. Thus, this work presents a method for identifying suspects based on messages and a controlled vocabulary composed of suspicious terms and their categories, according to a given domain. Different from the SML algorithms, the proposed method does not demand labeled data. Instead, it analyzes the messages exchanged on a given social network, and scores people according to the occurrence of the vocabulary terms. It is worth to highlight the endurance aspect of the proposed method since a controlled vocabulary is quite stable and evolves slowly. Moreover, the method was implemented for Portuguese texts and was applied to the “PAN-2012-BR” data set, showing some promising results in the pedophilia domain.

1 INTRODUCTION

The analysis of social or complex networks¹ has attracted great socio-economic interest from the public and private institutions, since through this analysis it is possible to extract characteristics and behavioral patterns of people on these networks (Dorogovtsev and Mendes, 2002; Figueiredo, 2011). For example, the identification of people who use the resources of these networks in order to commit and/or propagate acts that may bring risks inside and outside the network, such as terrorism, pedophilia, virtual bullying (Pendar, 2007; Villatoro-Tello et al., 2012; Santos and Guedes, 2019).

^a  <https://orcid.org/0000-0002-0828-4058>

^b  <https://orcid.org/0000-0003-1688-0586>

^c  <https://orcid.org/0000-0003-4965-9941>

¹A social or complex network, in the context of this work, is a highly interconnected multigraph, where each vertex represents a network item (e.g., person, web page, photo, company, group, etc.) and each edge represents some kind of interaction between the items connected by it (e.g., friendship, collaboration, communication, etc.).

Most authors, when working with this topic, use the supervised machine learning approach, which consists of using a set of previously labeled data, informing the types of people (suspect and not suspect) in the network in order to build classification models (Fire et al., 2012; Villatoro-Tello et al., 2012). However, for the identification of suspects of committing virtual crimes, there is a great difficulty in obtaining a previously labeled set of data (i.e., generally, labeled data in this context are rare), which requires to carry out this labeling manually (Pendar, 2007). Furthermore, social networks have a very dynamic nature, varying significantly, which demands the model to be frequently updated with recent data. Given such a hard situation, the following research question may be posed: *How to identify suspicious people, using messages, on social networks, without depending on a previously labeled data set?* In this direction, the present work raises the hypothesis that the use of a controlled vocabulary over the domain of the application can lead to the identification of suspects of committing a virtual crime without the need of a previously labeled data set.

This work presents a method for identifying suspects of committing virtual crimes using messages from a social network. In the proposed method, these messages are prepared using text mining techniques in order to reduce computational processing and facilitate information extraction. Once prepared, a controlled vocabulary composed of terms related to the field of application is used. This vocabulary is weighted according to a domain expert and to the existence of these terms in the analyzed network. From the weighted vocabulary, it is also possible to score each person according to the use of those terms in their messages. In the end, people are ordered in a descending manner, where the most suspicious of committing virtual crimes will be at the top of the list. It is important to highlight the endurance aspect of the proposed method since a controlled vocabulary is quite stable and evolves slowly if compared to the dynamic nature of the social networks data usually used by most works on this area.

The present text has five other sections, organized as follows: In Section 2 some basic concepts related to the analysis of social networks, text mining, and knowledge representation are presented. In Section 3, some of the main related works are presented, highlighting the contribution of this work. Then, Section 4 describes the proposed method. Details about the experiments performed and the results obtained are in Section 6. Finally, Section 7 reveals the main contributions of this article and points out alternatives for future work.

2 BASIC CONCEPTS

In social network analysis it is common to represent the network using a directed multigraph with attributes. This multigraph may be homogeneous or heterogeneous² (Muniz et al., 2018; Dong et al., 2012). In a homogeneous directed multigraph $G(V, E)$, V is a set of nodes representing individuals or objects (e.g. people), and E is a set of directed edges representing a type of relationship between the vertices (e.g. messages). This type of multigraph makes it possible to represent, for example, messages sent

²A G graph is said to be: (a) homogeneous if, and only if, G has only one type of vertex/node and one type of edge; (b) heterogeneous if, and only if, G has two or more types of vertex/node or edge; (c) a property graph if, and only if, G contains attributes associated to its nodes and/or to its edges; (d) a multigraph if, and only if, G has two or more types of edge connecting the same pair of nodes; (e) a directed graph if, and only if, G has directed edges, making it possible to identify the source and destination nodes of each edge.

and received by nodes of the type person, u and v , where u and $v \in V$, $e = (u, v)$, $e \in E$.

In a heterogeneous directed multigraph $G(V, E)$, the V set is formed by different types of nodes (e.g. person and book), and the E set, by different types of edges (e.g. buy and read). More formally, $V = V_1 \cup V_2 \cup \dots \cup V_n$ and $E = E_1 \cup E_2 \cup \dots \cup E_n$. Thus, V and E are formed by the union, respectively, of different types of nodes and directed edges. For example, consider the heterogeneous directed multigraph $G(V, E)$, where $V = V_P \cup V_B$ and $E = E_{PB} \cup E_{BP}$. In G , each $v_i \in V_P$ and $v_j \in V_B$, are nodes of the type *Person* and *Book*, respectively. Additionally, E_{PB} and E_{BP} , are sets of edges $e_k = (v_i, v_j)$ that represent the buying relationships and $e_m = (v_i, v_j)$ that represents the reading relationships, respectively. Thus, one can express that a person v_1 bought a book v_3 that was read by a person v_2 , as follows: $v_1, v_2 \in V_P$, $v_3 \in V_B$ and $\exists(v_1, v_3) \in E_{PB}$ and $\exists(v_3, v_2) \in E_{BP}$. It is common, in multigraphs, for vertices ($v \in V$) and/or edges ($e \in E$) to be associated with one or more attributes that can represent different types of information about them (e.g. temporal, topological and/or contextual). These attributes could, for example, associate the message exchanged between a pair of persons represented in the multigraph.

Frequently, in a social network, it is required to analyse contextual information, especially messages and descriptions. In this context, the use of text mining can be an interesting approach, because through it, one seeks to extract useful information, using computational techniques (Berry Michael, 2004). Text mining is usually seen as a process composed of several stages, among which we can highlight, the indexation and normalization. In the latter, techniques are employed to reduce and standardize the text. One of them is *stemming*, which reduces a word to its radical, for example, the radical of the words “beautiful” and “beauty” would be “beauti”. Another technique is the removal of *stop words* that remove from the text words with little or no meaning, for example, the words “a”, “and”, among others. Another text mining usual step is to calculate the relevance of the terms, that may be based on measures such as term frequency (TF) and the inverse document frequency (IDF). These measures, respectively, check the frequency and the rarity of a term in a collection of textual data, making it possible to represent numerically the relevance of the term (Morais and Ambrósio, 2007).

Another technique that can be applied to social network analysis is the semantic annotation of texts, which uses semantic resources such as controlled vocabularies (or thesaurus), ontologies, among others

(Moura, 2009). Controlled vocabularies are a set of descriptor terms that are semantically related to a given domain. These terms can be organized in the form of hierarchies, that is, with hierarchical relationships between them (Sales and Café, 2009). On the other hand, an ontology is a more sophisticated semantic resource, because in addition to representing terms from a given field of knowledge, it maintains several types of relationships between them, including hierarchical relationships (Chandrasekaran et al., 1999).

3 RELATED WORK

In the literature, some works focus on the identification of criminal suspects on social networks. Similar to the present work, some of them adopt a contextual analysis approach, i.e., they perform a content analysis of the exchanged messages to identify criminal practices or suspects.

Pendar (2007) sought to identify sexual predators using people's messages in a social network. In this work, the authors count on a set of data previously labeled, informing predators and non-predators. They perform different linguistic analysis of the messages exchanged by these people on the social network, such as *Bag-of-words* and *TF-IDF* statistical measurements. According to the authors, this information is provided to a machine learning algorithm to generate a model that characterizes sexual predator messages, which will enable the identification of sexual predators in an unlabeled network.

Villatoro-Tello et al. (2012) and Santos and Guedes (2019) followed the same line of the method developed by Pendar (2007). The work of Villatoro-Tello et al. (2012) differs in that it attacks both problems at once, i.e., their method is able to identify suspicious conversation and also is able to tell which user is the sexual predator. Santos and Guedes (2019)'s work is very similar to Pendar (2007). Its main contribution is that, to the best of our knowledge, it was one of the first works that developed a method to identify messages with a pedophile content in Portuguese.

Differently, Fire et al. (2012) and Wang (2010) have developed methods to identify *spammers* on social networks that take into account topological information. The method proposed in Fire et al. (2012) assumes that a highly connected user with friends that belong to several non connected communities has a great possibility to be a *spammer*. Yet in Wang (2010), besides the topological information (e.g., the relevance of a user according to the number of messages sent and received), it also takes into account the

content of the messages (e.g., the existence of HTML links, mentions to other people, and reference to trend topics). Similarly to Pendar (2007), both methods use a supervised approach, i.e., they need to use previously labeled data sets, informing *spammers* and *non-spammers*.

Another interesting work was developed by Elzinga et al. (2012). They present a non-automated method, using a time relational semantic system, aiming to analyze messages with a pedophile content in chat rooms for a certain time. The authors identified seven categories of terms used by pedophiles: *sweet greetings*, *compliments*, *intimate parts*, *sexual manipulations*, *cam* and *photos*, *where* and *when*. These categories characterize how pedophiles establish a connection and escalate the conversation on the net, towards a physical meeting. In the proposed analysis, each message is manually framed in one of those categories, and then they analyze the dynamics of the conversation. However, they did not report on the implementation of their method to automate the identification of the conversation dynamics pattern on a social network.

Bretschneider et al. (2014) developed a method to identify harassment messages on social networks. Based on a set of profane words, they select messages that mention them for further checking. If a message contains a profane word directly addressing some people, then they are labeled as harassment messages. In the naive version of the method, the authors labeled messages that mention profane words as harassment messages. Yet in Bretschneider and Peters (2016), a new version of the method was developed to identify Cyberbullying practice. In this version, those who have sent at least two harassment messages to the same person, are labeled as a Cyberbullying offender. Also, they calculate the degree of the offensive based on the use of a property directed multigraph, where nodes represent people, and a directed edge represents people's interaction. In this multigraph, edge attributes represent the interaction harassment degree (the number of received and sent harassment messages). Additionally, node attributes represent one's offensive degree (the number of offended people and the number of harassment messages sent).

Apart from Elzinga et al. (2012), Bretschneider et al. (2014) and Bretschneider and Peters (2016), all other mentioned works need a set of previously labeled data. This network labeling can be very costly and, depending on the network used, sometimes even impossible, as it is mostly a manual task.

More specifically, in Pendar (2007) and Villatoro-Tello et al. (2012), depending on the size of the analyzed dataset, a linguistic analysis of the messages

can have a high computational demand. Yet in Santos and Guedes (2019), the authors used a single analysis of the messages, reducing the computational effort. In Fire et al. (2012) and Wang (2010) the authors follow a topological approach. However, only in Wang (2010), the authors combine topological and contextual approaches, restricted to the Twitter social network. On the other hand, following a vocabulary-based approach, in Elzinga et al. (2012) the authors explore the dynamics of the content of the message, but they did not report on doing so for identifying criminal suspects in a large set of network messages. Yet in Bretschneider et al. (2014) and Bretschneider and Peters (2016), the authors propose an automated method for identifying cyberbullying criminals. However, its content analysis focuses only on this type of crime.

4 PROPOSED METHOD

In order to fill in some of the gaps left by the related work described above, this section presents *INSPECTION*, a new method for identifying suspects from the content of messages exchanged on a social network. Its main differential is that it does not require a labeled base. The idea is to use a categorized and weighted vocabulary. Similar to other related work, the method analyzes the content of messages, identifying suspicious messages based on that vocabulary. Then, according to the results of the analysis of the exchanged messages of each person on the network, it ranks them. The higher the position at the rank, the more suspect, which facilitates the identification of criminal suspects. Figure 1 gives an overview of the proposed method, which is described as follows.

Let a cutout of a network N composed of a set of messages M sent and/or received by people from a set U . Each message m_x , is an ordered set of terms³ ($m_x = t_1, t_2, \dots, t_n$). This method seeks to identify in U , people with suspicious behavior, through the terms used in the M messages ($t_j \in m_x$).

Initially, as Figure 1 shows, in the *Terms Preparation* step, all the terms t_j of every m_x are treated. This treatment allows the standardization of these terms, creating a set of messages with treated terms, called M' . Based on this set (M') and on the set of people (U), the *Representation of the Network* step generates two multigraphs, one to represent the interaction be-

³Formally, each message is represented as a tuple $\langle o, d, m_x \rangle$, where a source person (o), sends to a destination person (d) some message content (m_x). However, at some points of this article, a message is represented as m_x , for the sake of simplification.

tween people, and another that connects people and terms they used in their messages.

The *Controlled Vocabulary Weighting* step counts on a controlled vocabulary composed of suspicious terms, previously defined by an expert, according to a field of application (e.g. Pedophilia, Terrorism, among others). It is assumed that the vocabulary used is divided into large categories, covering the different aspects of the domain in question. Before weighing the vocabulary, the specialist's view is initially taken to weigh the categories, according to their importance. Each term of the vocabulary is then weighed according to its occurrence in the existing messages (M').

Later, at the *Contextual Analysis* step, each person is analyzed according to the presence of suspicious terms used in their sent messages. At the end, each person is assigned a *score*, which represents numerically the suspicious behavior of a person. Once the *scores* of all users have been calculated at the *Suspicious Person Identification* step, those persons are ranked. Thus the most suspicious of committing virtual crimes, according to the domain of the application, will be at the top of the list.

The following subsections detail each of the steps of the proposed method.

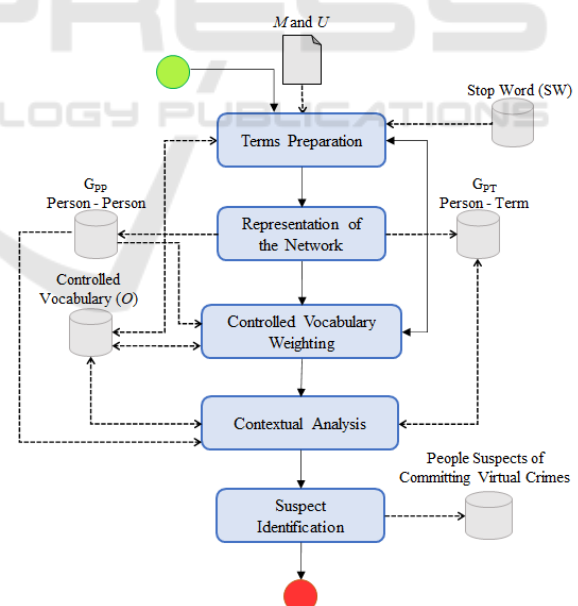


Figure 1: INSPECTION - Overview.

4.1 Terms Preparation

For every $m_x \in M$, this step treats each term $t_j \in m_x$, in order to obtain only the normalized terms that are more relevant to be analyzed. To do so, this step has

the following sub-steps:

- **Normalization and Extraction of Textual Content:** Due to the high textual informality existing in messages on social networks, for each $t_j \in m_x$, it is necessary to remove the use of repeated letters. Besides, to standardize these terms, all of them are placed in lower case and have removed their accents, signs, and punctuation.
- **Stop Words Removal:** This sub-step seeks to remove the *stop words* from m , i.e., remove the t_j with little or no meaning. This removal aims to reduce the computational effort of the next steps, since removing these terms consequently also reduces the number of terms that will be analyzed by the method. As shown in Figure 1, this step counts on the *SW* set, which is the set of terms with little or no meaning. At the end of this step, we have:

$$m_x = m_x - SW \quad (1)$$

- **Stemming:** This sub-step extracts the radical from each $t_j \in m_x$, and generates a new stemmed term (t'_j):

$$t'_j = stem(t_j) \quad (2)$$

In case a given t_j cannot be stemmed, then $t'_j = t_j$. Formally, $\forall m_x \in M \forall t_j \in m_x$ then $t'_j = stem(t_j)$ and $m'_x = m'_x \cup \{t'_j\}$ and $M' = M' \cup \{m'_x\}$. Thus, the new m'_x set is built by means of the stemming of the terms $t_j \in m_x$, and M' is built on each m'_x .

4.2 Representation of the Network

Through M' and U , this step builds two multi-graphs directed. These multigraphs, besides making it possible to identify the people who send and receive one or more messages or terms, also make it possible to carry out the analyses that will be made in the next steps.

- **Representing People and Messages (G_{PP}).** In this representation, $G_{PP}(V_P, E_{PP})$ is a homogeneous directed multigraph representing people and messages from the network. Thus, each $v_P \in V_P$ represents a person of U . As for the $e_{PP} \in E_{PP}$, besides representing a directed connection between two people that exchanged a message, it has a contextual attribute ($e_{PP}.T$) that represents the message text. $G_{PP}(V_P, E_{PP})$ is constructed as follows:

- $V_P = U$, where each $v_P \in V_P$ represents a person who sent and/or received one or more messages;

- $E_{PP} = \{e_{PP} \mid e, r \in V_P, e_{PP} = (o, d) \text{ and } \exists \langle o, d, m'_x \rangle \in M', e_{PP}.T = m'_x\}$, where each $e_{PP} \in E_{PP}$ is a connection that represents a message sent from one person to another.

- **Representing People and Terms (G_{PT}).** In this representation, the heterogeneous directed multi-graph $G_{PT}(V_P \cup V_T, E_{PT} \cup E_{TP})$ is responsible for representing both the people, and the terms (t'_j) used in $m' \in M'$ by those people. So there are two kinds of vertices. Vertices $v_P \in V_P$ represent the users in U , and vertices $v_T \in V_T$ represent terms used in messages exchanged between those users. Every $e_{PT} \in E_{PT}$ and $e_{TP} \in E_{TP}$ are directed edges that represent, respectively, the person who sent and received a certain term. Thus, $G_{PT}(V_P \cup V_T, E_{PT} \cup E_{TP})$ is constructed as follows:

- $V_T = \{v_T \mid \exists m'_x \in M'\}$, where $v_T = t'_j$ and $t'_j \in m'_x$;
- $E_{PT} = \{e = (o, t) \mid o \in V_P \text{ and } t \in V_T \text{ and } \exists e' \in E_{PP}, \text{ where } e' = (o, d) \text{ and } t \in e'.T\}$;
- $E_{TP} = \{e = (t, d) \mid d \in V_P \text{ and } t \in V_T \text{ and } \exists e' \in E_{PP}, \text{ where } e' = (o, d) \text{ and } t \in e'.T\}$.

4.3 Controlled Vocabulary Weighting

As already mentioned, the present method is based on a controlled vocabulary composed of terms related to the domain of the application, or more specifically, to the type of crime under investigation. Vocabulary terms are weighted with the help of a specialist and based on the set of messages exchanged in a social network. Thus, it becomes possible to express numerically how suspicious a term is in the context of a cutout from a social network.

A controlled vocabulary, in the present work, is defined as a tuple $O = [C, R]$, where C is a set of classes and R is a set that represents the generalization/specialization relationships between them. This means that the vocabulary is constituted of a set of taxonomies, one for each category/facet that must be taken into consideration for the crime type in focus. To better explain the *Controlled Vocabulary Weighting* step, C is subdivided into C_r and C_s ($C = C_r \cup C_s$), so that C_r and C_s are, respectively, a set of root classes, one for each taxonomy, and a set of their subclasses. In addition, each $c_{s_i} \in C_s$ is linked directly or indirectly to (descendant of) a single c_{r_j} in C_r . Each c_{r_j} generically represents the terms that correspond to its subclasses. And, each c_{s_i} is previously known as a suspicious term. Each c_r and c_s has a w attribute ($c_{r_j}.w$ and $c_{s_i}.w$) that indicates the relevance of the term in the vocabulary according to the network cutout, i.e., how suspicious the term is in that context.

The first sub-step of the *controlled vocabulary weighting* step begins by arbitrarily assigning the values of the weights to all $c_r.w$, where $c_r \in C_r$, according to an specialist in the crime type under investigation.

Subsequently, the subclasses of the controlled vocabulary are weighted according to their frequency in the available messages. To do this, initially, every $c_{s_i} \in C_s$ goes through the *Terms Preparation* step (and all its sub-steps). Formally, $\forall c_{s_i} \in C_s$ then $c'_{s_i} = stem(c_{s_i})$ and $C'_s = C_s \cup \{c'_{s_i}\}$. This way, it emerges the O' vocabulary, where $O' = [C', R]$ and $C' = C_r \cup C'_s$. Knowing that the subclasses $c'_s \in C'_s$ represent suspicious terms that could be used in messages on the network, the operation described in (3) aims to identify the set (A) of suspicious terms that are present in the cutout of the analyzed social network. In other words, this set will be built with elements $c'_{s_i} \in C'_s$ if exists $t'_j \in V_T$, such that $c'_{s_i} = t'_j$.

$$A = C'_s \cap V_T \quad (3)$$

Then, by means of Equation 4, the Global Weight (Wilbur and Yang, 1996) of each $t'_j \in A$, ($GW_{t'_j}$) is calculated. It is also known as Inverse Document Frequency (or *IDF*) (Robertson, 2004). This equation checks the inverse frequency of the term t'_j , or the rarity of the term, in the set of messages of the network cutout. Therefore, the less the term appears in the message set, the greater is its importance.

$$GW_{t'_j} = \log_2 \left(\frac{|E_{PP}|}{|n_{t'_j}|} \right) \quad (4)$$

Where:

- $|E_{PP}|$ is the total number of edges in G_{PP} ;
- $|n_{t'_j}|$ is the number of edges in E_{PP} that have the term t'_j in the messages they represent ($n_{t'_j} = \{e_{pp_i} | t'_j \in e_{pp_i}.T\}$).

However, the Global Weight ($GW_{t'_j}$) is not sufficient to calculate the relevance of each term in the context of the different categories (root classes) of the vocabulary. Therefore, to adjust the relevance of each $t'_j \in A$ according to its root class (c_{r_k}), a normalization operation is required. The goal is to limit $GW_{t'_j}$ to the weight w of its corresponding c_{r_k} . To do this, initially, it is necessary to find the Highest Global Weight (HGW), which is given by Equation 5. It calculates the maximum rarity of any term in relation to the messages of the network cutout.

$$HGW = \log_2 (|E_{PP}|) \quad (5)$$

Then, Equation 6 may be used to calculate the ratio of the Global Weight of each t'_j with respect to the Highest Global Weight HGW . In this equation, a rule of three is applied, i.e., assuming HGW corresponds to 100%, thus it is possible to obtain each equivalent ratio $GW_{t'_j}^{\%}$.

$$GW_{t'_j}^{\%} = \frac{GW_{t'_j}}{HGW} \quad (6)$$

Now, based on these ratios, each term weight may be normalized according to the weight of their corresponding root class, assigned by the specialist. To do this, for each root class (c_{r_k}) it is necessary to define the corresponding weight interval ($Min()$ and $Max()$). Thus, $Max(c_{r_k}) = c_{r_k}.w$ and $Min(c_{r_k}) = Max(\{c_{r_f}.w | c_{r_f} \in C_r - \{c_{r_k}\} \wedge c_{r_f}.w < c_{r_k}.w\} \cup \{0\})$. Finally, the normalized global weight of each t'_j is given by Equation 7:

$$GW_{t'_j}^N = ((Max(c_{r_k}) - Min(c_{r_k})) \times GW_{t'_j}^{\%}) + Min(c_{r_k}) \quad (7)$$

It is worth noting that this equation normalizes each term global weight ($GW_{t'_j}^N$), placing it within a range limited by two root class weights. The top boundary of such range corresponds to the weight of the root class to which the t'_j is connected ($Max(c_{r_k})$), and the bottom boundary corresponds to the weight of the smaller subsequent root class ($Min(c_{r_k})$), if it exists. If it does not exist, 0 is assumed. Thus, $GW_{t'_j}^N$ is found from $GW_{t'_j}^{\%}$ within the mentioned range.

For each $c'_{s_i} = t'_j$, the operation described in (8) assigns the normalized global weight of a term ($GW_{t'_j}^N$) to the corresponding term (or subclass) in the controlled vocabulary representation.

$$c'_{s_i}.w = GW_{t'_j}^N \quad (8)$$

4.4 Contextual Analysis

At this step, contextual analysis is carried out based on the term weights of each person's messages. Briefly, the idea is to obtain a *score* that expresses how suspicious each person's behavior is. The following sub-steps describe how this *score* is calculated. At the end of this step, this *score* is stored for each person of the network ($v_p.st | v_p \in V_p$ of G_{PT}).

Initially, all terms used by a person are retrieved. For each $v_p \in V_p$ of G_{PT} it is necessary to retrieve all v_T to which it is connected. More formally, for a given v_p , this step retrieves the following set:

$$C_{v_T}(v_p) = \{v_T | \exists (v_p, v_T) \in E_{PT}\} \quad (9)$$

Since not all terms $v_T \in C_{v_T}(v_P)$ are suspect terms, the next operation reduces this set to a new corresponding set, containing only terms that are present in the controlled vocabulary. That way, using the set of subclasses (C'_s) of the vocabulary \mathcal{O}' , and knowing that $v_{T'_j} = c'_{s'_j}$, a reduced set is created for each $v_P \in V_P$ of G_{PT} :

$$C_{v_T}^{\cap}(v_P) = C_{v_T}(v_P) \cap C'_s \quad (10)$$

Next sub-step calculates two metrics that were developed to quantify each person's suspicious behavior: \mathcal{M}_{GW} and \mathcal{M}_{FGW} . These metrics generally quantify the use of all suspicious terms by a person ($v_T \in C_{v_T}^{\cap}(v_P)$). As shown in Equation 11, \mathcal{M}_{GW} metric is the sum of the rarity of each suspect term used by a person in messages, normalized by the weight of a root class, to which this term is connected.

$$\mathcal{M}_{GW}(v_P) = \sum_{c_{s_i} \in C_{v_T}^{\cap}(v_P)} c_{s_i} \cdot w \quad (11)$$

Note that Equation 11 does not take into account term frequency, i.e., how frequently a person used each suspect term. Differently, as shown in Equation 12, \mathcal{M}_{FGW} metric multiplies each term normalized global weight by term frequency, indicating the importance of this term for a person in relation to all the messages analyzed.

$$\mathcal{M}_{FGW}(v_P) = \sum_{c_{s_i} \in C_{v_T}^{\cap}(v_P)} W(v_P, c_{s_i}) \times c_{s_i} \cdot w \quad (12)$$

where:

- $v_{T_j} = c_{s_i}$
- $W(v_P, v_{T_j})$ retrieves the frequency of use of a particular suspect term by a person, formally:
 $W(v_P, v_{T_j}) = |\{(o, t) \in E_{PT} \wedge o = v_P \wedge t = v_{T_j}\}|$

Finally, the user may select one of the proposed metrics (Equations 11 or 12), and then the corresponding *score* is assigned to each person, representing his/her suspicious behavior, as follows:

$$v_P.st = \mathcal{M}_{GW}(v_P) \quad (13)$$

$$v_P.st = \mathcal{M}_{FGW}(v_P) \quad (14)$$

It is worth noting that operations 13 and 14 assign the weights, respectively obtained by the Equations 11 and 12, to the analyzed person representation (v_P).

4.5 Suspect Identification

At this step, an ordered list of people is generated according to the selected *score* metric. In a descending order of *scores*, people at the top of the list are the most suspicious ones.

5 EXAMPLE

This section aims to illustrate the usage of the method presented in Section 4. Let the sets of users and messages be, respectively, $U = \{\text{Carlos, Paula, Ana}\}$ and $M = \{\langle \text{Carlos, Paula, How about a beach tomorrow my beautiful?} \rangle, \langle \text{Paula, Ana, Beachhh?} \rangle, \langle \text{Ana, Paula, Not tomorrow!} \rangle, \langle \text{Paula, Ana, Want to?} \rangle, \langle \text{Carlos, Paula, Let's go tomorrow?} \rangle\}$. These sets were processed by the proposed method, and each step is described as follows.

Terms Preparation. In this step some sub-steps such as Normalization and Extraction of Textual Content, Stop Words Removal and Stemming, are executed, and M is converted into $M' = \{\langle \text{Carlos, Paula, \{beach, tomorrow, beauti\}} \rangle, \langle \text{Paula, Ana, \{beach\}} \rangle, \langle \text{Ana, Paula, \{tomorrow\}} \rangle, \langle \text{Paula, Ana, \{want\}} \rangle, \langle \text{Carlos, Paula, \{tomorrow\}} \rangle\}$.

Representation of the Network. From M' and U two multigraphs are built: G_{PP} and G_{PT} . In $G_{PP}(V_{PP}, E_{PP})$, $V_{PP} = \{\text{Carlos, Paula, Ana}\}$ and $E_{PP} = \{(\text{Carlos, Paula}), (\text{Paula, Ana}), (\text{Ana, Paula}), (\text{Carlos, Paula})\}$, where, for instance, $e_{PP_1} = (\text{Carlos, Paula})$, $e_{PP_1}.T = \{\text{beach, tomorrow, beauti}\}$. Figure 2 shows the graphical representation of G_{PP} for the complete example.

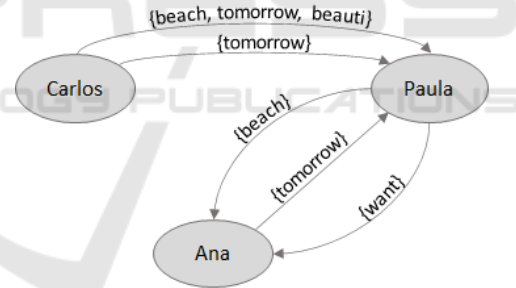


Figure 2: Graphic representation of the multigraph Person-Person (G_{PP}) built from the example.

In $G_{PT}(V_P \cup V_T, E_{PT} \cup E_{TP})$ graph, shown in Fig. 3, V_T is the set of white nodes, while V_P is the set of gray nodes. The edges of the graph correspond to the connections between the nodes from both sets, V_P and

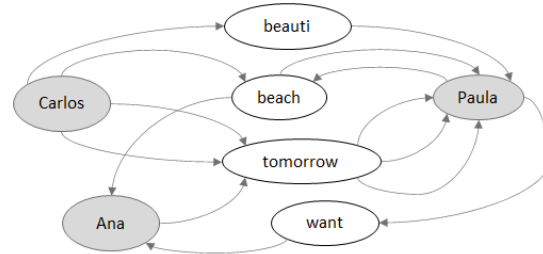


Figure 3: Graphic representation of the Multigraph Person-Term (G_{PT}) built from the example.

V_T . For instance, the edge $(\text{Carlos}, \text{beuti}) \in E_{PT}$, while $(\text{beuti}, \text{Paula}) \in E_{TP}$.

Controlled Vocabulary Weighting. First, it is assumed that there is a controlled vocabulary formed by a set of terms C_s , commonly used in Pedophiles conversation. This vocabulary is organized according to some generic categories (C_r classes). For this example $C_r = \{\text{When}, \text{Compliments}, \text{Clothes}\}$. Fig. 4 shows the terms used in this example, and their organization as subclasses of the categories in C_r . Each of these categories is weighted by a specialist, with values 1, 2 and 3, respectively.

The vocabulary terms (C_s) are weighted according to their usage in the messages under analysis (V_T). Therefore, to avoid the cost of weighting all the terms, first it is necessary to identify which of them need to be weighted. Then, in order to normalize them in the same way of the message terms, each term in the C_s set is submitted to some of the sub-steps of the Terms Preparation step.

The set of selected vocabulary terms (A) is obtained by applying Operation 3. Since $C_s = \{\text{instant}, \text{today}, \text{tomorrow}, \text{care}, \text{beuti}, \text{intim}, \text{panti}, \text{brassier}\}$ and $V_T = \{\text{beach}, \text{tomorrow}, \text{beuti}, \text{want}\}$, thus $A = \{\text{tomorrow}, \text{beuti}\}$. Then, based on the G_{PP} multi-graph of Figure 2, the global weight (GW) of each term $t'_j \in A$ is calculated by applying Equation 4 as follows:

- $GW_{\text{beuti}} = \log_2\left(\frac{5}{1}\right) = 2,32$
- $GW_{\text{tomorrow}} = \log_2\left(\frac{5}{3}\right) = 0,74$.

Subsequently, to normalize these weights according to the Highest Global Weight, first it is calculated by applying Equation 5: $HGW = \log_2(5) = 2,32$. Then, using this value, Equation 6 is applied to obtain the weight rates for each term:

- $GW_{\text{beuti}}^{\%} = \frac{2,32}{2,32} = 1,00$
- $GW_{\text{tomorrow}}^{\%} = \frac{0,74}{2,32} = 0,32$.

Next, these rates are used to recalculate each term weight and normalize them according to the weights of their corresponding categories. Considering the fact that *beuti* and *tomorrow* are related to *Compliments* ($w = 2$) and *When* ($w = 1$) categories, respectively, then only these two categories are used as references. Therefore, the reference values for each category are calculated as follows:

- $\text{Min}(\text{Compliments}) = \text{Max}(\{1, 0\} \cup \{0\}) = 1,0$
- $\text{Min}(\text{When}) = \text{Max}(\{0\} \cup \{0\}) = 0$.

Finally, Equation 8 is used to obtain the final Global Weight for each term, within the interval of its corresponding category:

- $GW_{\text{beuti}}^N = ((2,0 - 1,0) \times 1,0) + 1,0 = 2,0$

- $GW_{\text{tomorrow}}^N = ((1,0 - 0) \times 0,32) + 0 = 0,32$.

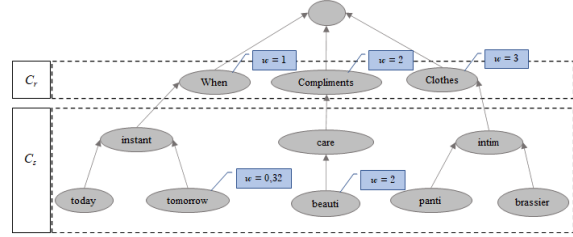


Figure 4: Controlled vocabulary: weighted and normalized.

Contextual Analysis. Now that the weights were calculated for the selected vocabulary terms, it is possible to calculate the *score* of each member of the V_P set. Table 1 summarizes the results of applying Operations 9 and 10. Note that Carlos is the one that mentions more terms that belong to the vocabulary. Subsequently, based on the weights of the terms they mention, a *score* is calculated for each person. Table 2 shows the *scores* for both metrics, $\mathcal{M}_{FGW}(v_P)$ and $\mathcal{M}_{GW}(v_P)$ (Eq. 12 and 11), for all users in V_P . These *scores* indicate how suspicious a person is.

Table 1: Contextual analysis of the example.

v_P	$C_{v_P}(v_P)$ (Eq. 9)	C_s	$C_{v_P}^N(v_P)$ (Eq. 10)
Carlos	{beach, tomorrow, beuti}	{tomorrow, beuti}	{tomorrow, beuti}
Paula	{beach, want}		{}
Ana	{tomorrow}		{tomorrow}

Table 2: Scores according to \mathcal{M}_{GW} and \mathcal{M}_{FGW} metrics.

v_P	$C_{v_P}(v_P)$	$c_{v_P, w}$	$\mathcal{M}_{GW}(v_P)$	$W(v_P, C_{v_P})$	$\mathcal{M}_{FGW}(v_P)$
Carlos	tomorrow	0,32	2,32	2	2,64
	beuti	2		1	
Paula	-	0	0	0	0
Ana	tomorrow	0,32	0,32	1	0,32

textbfSuspect Identification. In both metrics it was verified that Carlos got the highest *scores* ($\mathcal{M}_{GW}(\text{Carlos}) = 2,32$ and $\mathcal{M}_{FGW}(\text{Carlos}) = 2,64$). This indicates that he is the most suspicious among the people of the U set. On the other hand, Paula did not use any term considered suspicious in her vocabulary. Thus, in both metrics her *score* was 0.

6 EXPERIMENTS AND RESULTS

In order to evaluate the proposed method, a prototype was implemented in Python v. 2.7, using several APIs⁴. The experiments ran on this prototype. They focused on the identification of pedophilia suspects, and were carried out using the ‘‘PAN-2012-

⁴APIs used: *NLTK* for portuguese text manipulation, *NetworkX* to deal with graph structures, and *AnyTree* to deal with the controlled vocabulary structure.

BR” data set (Santos and Guedes, 2019; Andrijauskas et al., 2017). This data set was developed in partnership with the Federal Prosecutor’s Office of São Paulo (MPF-SP), the University Center of the Fundação Educacional Inaciana (FEI), and the Federal University of Minas Gerais (UFMG). The “PAN-2012-BR” data set is composed of conversations in Portuguese, i.e., people and the messages exchanged between them. These conversations and people are labeled as predators or non-predators. It is worth noticing that the notion of predator is different from the notion of suspect. However, due to the difficulty of obtaining a data set composed of common and suspicious people for the evaluation of the proposed method, we decided to proceed with the experiment using the predator label, instead of suspect label.

Table 3: “PAN-2012-BR” social network statistical information (Santos and Guedes, 2019; Andrijauskas et al., 2017).

Label	People	Messages	People	Messages
Non Predators	330	21.909	368	22.255
Predators	77	436	39	90
TOTAL	407	22.345	407	22.345

Table 3 presents statistical information about the “PAN-2012-BR” social network, organized by label, users and messages. It should be noted that the predator/non-predator labels will only be used at the end of the experiment, to verify the method performance. Thus, only information about people (sender and receiver) and their exchanged messages will be used as input for this experiment.

While planning for the experiment, it was noticed that the number of messages exchanged by each user in the “PAN-2012-BR” network cutout varied considerably. Taking into account that the proposed method *scores* people according to their messages, active people (who sends a high number of messages) could be heavily scored in contrast to less active ones. To avoid this problem, the number of exchanged messages to be analyzed per person was limited to a threshold. Thus, to investigate the impact of this message limitation, two experiment configurations were defined. The first one executes the method on all messages of the dataset, i.e., with no limitation (E_1). The second one (E_2) analyzes a limited number of messages per person. In this work, the median (M_d) was used as the maximum number of messages per person for the whole network cutout.

A controlled vocabulary for Pedophilia crimes, named O_1 , was created and organized according to 6 (six) root classes (or categories) of terms, based on the categories proposed by Elzinga et al. (2012): “where”, “when”, “intimate parts”, “sexual manipu-

lations”, “cam and photos” and “compliments”. The *sweet greetings* category was discarded due to its similarity to the *compliments* category. It is worth to point out that the subclasses for each root class or category, were built from real ontology cutouts. Table 5 shows each $c_r \in C_r$ of the O_1 controlled vocabulary, and their respective sources, i.e., the semantic resources from which the set of corresponding subclasses (c_s) were extracted.

To execute the experiment according to the E_2 configuration, the M_d was calculated and only the first 13 (thirteen) messages in M from each person in U were selected. The experiments for both configurations began with the *Terms Preparation* step, when the *PAN-2012-BR* dataset selected messages went through the *normalization*, *Stemming* and *Stop Word Removal* sub-steps. Then, the *Representation of the network* step was executed. Table 4 shows the data for both multigraphs (homogeneous or heterogeneous) that were created to represent the “PAN-2012-BR” data. Note that different multigraphs were created for each experiment configuration, i.e., with (E_2) or without (E_1) message number limitation.

Table 4: Multigraph data for the “PAN-2012-BR” dataset.

# of Messages	Multigraph	Nodes		Edges
		People	Terms	
All msgs	G_{PP}	407	-	22.345
	G_{PT}		7.680	119.939
First 13 msgs(M_d)	G_{PP}	-	-	4.014
	G_{PT}		2.915	18.666

At the beginning of the *Controlled Vocabulary Weighting* step, the root classes ($c_r \in C_r$) were weighted according to the experience of a Brazilian Federal Policeman, who is a specialist on Pedophilia crimes. After a brief presentation of the method, the specialist assigned values according to the importance of each category, within a scale of 1 (one - less important) to 6 (six - most important), as shown in Table 5. In addition, it was recommended to assign a distinct value to each category. The reason for this is to enhance the distinction between the term weights within the range of values of each category.

Table 5: Source of the subclasses (terms) linked to each root class of the controlled vocabulary O_1 .

C_r	w	Source
Where	2.3	(Scheider and Kiefer, 2018)
When	1.5	(Hobbs and Pan, 2006)
Intimate Parts	5.7	(Rosse and Mejino, 2008)
Sexual Manipulations	5.5	(Kronk et al., 2019)
Cam and Photos	6.0	(Mukherjee and Joshi, 2013)
Compliments	4.0	(Neves, nd)

According to the pedophilia specialist, the categories used reflect the usual interaction of a typical pedophile. First of all, the criminal usually demands

photos, films, etc, which is why the highest importance was given to this category (*Cam and photos*). *Intimate parts* and *Sexual Manipulations* are also categories of obvious importance. *Compliments* are often used to gain the victim's trust. Regarding *When* and *Where* categories, these were not so important, not only because they are very commonly used in any conversation, but especially because the main goal of a pedophile is not to schedule meetings, but to obtain images, photos and films.

Once the root classes were weighted, the experiments proceeded to the weighting of the O_1 remaining subclasses. Each subclass was weighted based on their corresponding root class weight, and on the heterogeneous multigraph extracted from the messages in the *PAN-2012-BR* data set, as explained in Section 4.3, operations/equations (3) to (8).

The subsequent step of the method, *Context Analysis*, was also performed. It calculated both metrics (Equations (11) and (12)) for each person in each heterogeneous multigraph. Thus, for each experiment configuration, it was generated two predator rankings. Since the experiments counted on a labeled data set, it was possible to evaluate the method performance with respect to those rankings. Table 6 shows the performance results compared to the Naive approach of the method proposed by Bretschneider et al. (2014) and Bretschneider and Peters (2016), applied to the same vocabulary. Considering that it is the evaluation of a score ranking, an adaptation of the measure *Area Under the Curve (AUC)* (Li et al., 2018) was used. This measure expresses the probability that a suspect always has a *score* higher than a non-suspect, both chosen n times randomly. In this work, $n = 100$.

Note that Table 6 shows the results for both experiment configurations (E_1 and E_2) with two vocabularies: O_1 and O_2 . The O_2 vocabulary is an evolution of O_1 . It was created to include a new category of terms to represent clothing terms (*Clothes* category). For this category, the specialist assigned a weight value of 5.0 (five) for its importance. The subclasses for this new category (root class) were created based on the ontology proposed in Kuang et al. (2018).

The Naive method assumes that messages are suspicious if they have at least one term present in the vocabulary. Once identified suspicious messages, a person is suspicious if he/she had sent more than two suspicious messages to the same person. To compare the method proposed in the present work to the Naive method, for the latter one, people were ranked according to the number of suspicious messages sent.

Analyzing the results in Table 6, for the experiments performed using the O_1 vocabulary, the method proposed here obtained superior results for both met-

rics, if compared to the Naive method ($AUC = 0.255$). Note that the E_1 experiment configuration obtained the best performance for the \mathcal{M}_{GW} metric ($AUC = 0.478$), showing a difference of 0.023, when compared to the other \mathcal{M}_{FGW} metric ($AUC=0.455$). However, none of the metrics obtained an AUC value greater than 0.5, which may lead to the conclusion that the E_1 experiment setup could have been inappropriate.

On the other hand, for the E_2 experiment setup, both metrics (\mathcal{M}_{GW} and \mathcal{M}_{FGW}) obtained results higher than 0.5, showing a better performance of the method when it is configured to analyze a limited number of messages per person. The best AUC value (0.660) was obtained with the \mathcal{M}_{GW} metric. It shows a difference of 0.04 when compared to the AUC value obtained with the \mathcal{M}_{FGW} metric ($AUC=0.620$).

With respect to the use of the evolved vocabulary (O_2), the experiments' performance is very similar to the ones using the O_1 vocabulary. Again, the method proposed in this work obtained better results than the Naive method Bretschneider and Peters (2014), for both experiment setups and metrics.

Note that the E_2 experiment configuration obtained AUC values greater than 0.5 for both metrics, showing a good performance of the method in executions with both vocabularies. With the use of O_2 vocabulary, the best performance was obtained with \mathcal{M}_{GW} metric ($AUC = 0.655$), which shows an improvement of 0.060 when compared to the AUC value obtained with the \mathcal{M}_{FGW} metric ($AUC = 0.595$). For this setup, the \mathcal{M}_{GW} metric obtained the best AUC values in all cases, and the best one was with the O_1 vocabulary ($AUC = 0.660$). When compared to the AUC value obtained using the O_2 vocabulary ($AUC = 0.650$), it shows a downward tendency, which leads to the idea that the evolution of the vocabulary was not well conducted. Therefore, taking into account that the best results were obtained with a limited number of messages per user within the network cutout (no more than M_d messages), it is possible to conclude that this may be the best setup. Moreover, it also reduces the computational effort, and consequently, shortens the execution time, which benefits the method users. Another point to highlight is that the method showed worse results when using the frequency of suspicious terms in each person's messages. Then, the proposed method should recommend preferably the \mathcal{M}_{GW} metric rather than the \mathcal{M}_{FGW} metric that uses the term frequency. Finally, the enrichment of the controlled vocabulary should be carefully conducted to avoid a method performance loss.

In short, the hypothesis raised at the introduction of this work seems to have been confirmed, i.e., the

Table 6: Results obtained with the metrics \mathcal{M}_{FGW} and \mathcal{M}_{GW} , considering the experiment configurations $E1$ and $E2$.

Controlled Vocabulary	Bretcheneider (Naive)	Proposed Method			
		$E1$		$E2$	
		\mathcal{M}_{GW}	\mathcal{M}_{FGW}	\mathcal{M}_{GW}	\mathcal{M}_{FGW}
O_1	0,255	0,478	0,455	0,660	0,620
O_2	0,275	0,430	0,430	0,655	0,595

use of a controlled vocabulary over a crime type domain may be a way towards the identification of suspicious persons. It is important to highlight that despite the AUC values presented by INSPECTION are not comparable to the ones produced by machine learning methods, the proposed method does not demand labeled data as these methods do.

7 CONCLUSIONS

One of the main concerns in the analysis of social networks is to identify suspicious people. It is a hard task to find out who makes use of these networks to practice crimes or spread risk to other people. There are already several machine learning based methods to identify suspicious people in social networks. Although most of them have shown promising results, they demand previously labeled data (indicating who are the suspects) to build their classification models. Such demand hampers their use in real applications because labeled data in the context of virtual crimes are usually rare and difficult to obtain. Given this scenario, the present work proposed INSPECTION, a method that uses a controlled vocabulary, specifically built according to the crime type in focus, to identify suspicious people in the social network, without the need of previously labeled data sets.

To evaluate the proposed method, this work reports on experiments for the pedophilia criminal scenario. To perform these experiments, a prototype was implemented. Also, a specific controlled vocabulary was built (in Portuguese), based on other existing vocabularies. The results show that the INSPECTION method is a promising approach to identify suspicious people without depending on a previously labeled data.

Future works include evaluating the performance of the proposed method applied to other social networks, and other crime types. In addition, an ongoing work is the extension of the proposed method to include social network topological analysis. Such analysis may lead to improvements on the INSPECTION's performance. Moreover, the inclusion of a new stage in the INSPECTION process to enrich the controlled vocabulary is foreseen, as well as, to consider semantic information while weighting the controlled vocabulary.

ACKNOWLEDGEMENTS

This study was partially funded by the Cybernetic Research Subproject of the Brazilian Army Strategic Project. In addition, the authors would like to thank Dr. Paulo Renato da Costa Pereira, a specialist in pedophilia crimes of the Brazilian Federal Police, for his valuable support throughout the experiments.

REFERENCES

- Andrijauskas, A., Shimabukuro, A., and Maia, R. F. (2017). Desenvolvimento de base de dados em língua portuguesa sobre crimes sexuais (*in Portuguese*). VII Simpósio de Iniciação Científica, Didática e Ações Sociais da FEI.
- Berry Michael, W. (2004). Automatic discovery of similar words. *Survey of Text Mining: Clustering, Classification and Retrieval*, Springer Verlag, New York, LLC, pages 24–43.
- Bretschneider, U. and Peters, R. (2016). Detecting cyberbullying in online communities. *European Conference on Information Systems*.
- Bretschneider, U., Wöhner, T., and Peters, R. (2014). Detecting online harassment in social networks. *International Conference on Information Systems*.
- Chandrasekaran, B., Josephson, J. R., and Benjamins, V. R. (1999). What are ontologies, and why do we need them? *IEEE Intelligent Systems and their applications*, 14(1):20–26.
- Dong, Y., Tang, J., Wu, S., Tian, J., Chawla, N. V., Rao, J., and Cao, H. (2012). Link prediction and recommendation across heterogeneous social networks. In *2012 IEEE 12th International conference on data mining*, pages 181–190. IEEE.
- Dorogovtsev, S. N. and Mendes, J. F. (2002). Evolution of networks. *Advances in physics*, 51(4):1079–1187.
- Elzinga, P., Wolff, K. E., and Poelmans, J. (2012). Analyzing chat conversations of pedophiles with temporal relational semantic systems. In *Europ. Intel. and Security Informatics Conf.*, 2012, pages 242–249. IEEE.
- Figueiredo, D. R. (2011). *Introdução a redes complexas (in Portuguese)*, pages 303–358. Sociedade Brasileira de Computação, Rio de Janeiro.
- Fire, M., Katz, G., and Elovici, Y. (2012). Strangers intrusion detection-detecting spammers and fake profiles in social networks based on topology anomalies. *Human Journal*, pages 26–39.
- Hobbs, J. R. and Pan, F. (2006). Time ontology in owl. *W3C working draft*, 27:133.
- Kronk, C., Tran, G. Q., and Wu, D. T. (2019). Creating a queer ontology: The gender, sex, and sexual orientation (gss) ontology. *Studies in health technology and informatics*, 264:208–212.
- Kuang, Z., Yu, J., Li, Z., Zhang, B., and Fan, J. (2018). Integrating multi-level deep learning and concept ontology for large-scale visual recognition. *Pattern Recognition*, 78:198–214.

- Li, S., Huang, J., Zhang, Z., Liu, J., Huang, T., and Chen, H. (2018). Similarity-based future common neighbors model for link prediction in complex networks. *Scientific reports*, 8(1):1–11.
- Morais, E. A. M. and Ambrósio, A. P. L. (2007). Mineração de textos (*in Portuguese*). Relatório Técnico–Instituto de Informática (UFG).
- Moura, M. A. (2009). Informação, ferramentas ontológicas e redes sociais ad hoc: a interoperabilidade na construção de tesouros e ontologias (*in Portuguese*). *Informação & Sociedade: Estudos*, 19:59–73.
- Mukherjee, S. and Joshi, S. (2013). Sentiment aggregation using conceptnet ontology. In *Proceedings of the Sixth International Joint Conference on Natural Language Processing*, pages 570–578.
- Muniz, C. P., Goldschmidt, R., and Choren, R. (2018). Combining contextual, temporal and topological information for unsupervised link prediction in social networks. *Knowledge-Based Systems*.
- Neves, F. (n.d.). Elogios de A a Z (*in Portuguese*). <https://www.dicio.com.br/elogios-de-a-a-z>. Accessed: 2020-07-18.
- Pendar, N. (2007). Toward spotting the pedophile telling victim from predator in text chats. *ICSC*, pages 235–241.
- Robertson, S. (2004). Understanding inverse document frequency: on theoretical arguments for idf. *Journal of documentation*.
- Rosse, C. and Mejino, J. L. (2008). The foundational model of anatomy ontology. In *Anatomy Ontologies for Bioinformatics*, pages 59–117. Springer.
- Sales, R. d. and Café, L. (2009). Diferenças entre tesouros e ontologias (*in Portuguese*). *Perspectivas em Ciência da Informação*, 14(1):99–116.
- Santos, L. and Guedes, G. P. (2019). Identificação de predadores sexuais brasileiros por meio de análise de conversas realizadas na internet (*in Portuguese*). XXXIX Congresso da Sociedade Brasileira de Computação.
- Scheider, S. and Kiefer, P. (2018). (re-) localization of location-based games. In *Geogames and Geoplay*, pages 131–159. Springer.
- Villatoro-Tello, E., Juárez-González, A., Escalante, H. J., Montes-y Gómez, M., and Pineda, L. V. (2012). A two-step approach for effective detection of misbehaving users in chats. *CLEF*.
- Wang, A. H. (2010). Don't follow me: Spam detection in twitter. In *International Conference On Security and Cryptography (SECRYPT)*, pages 1–10. IEEE.
- Wilbur, W. J. and Yang, Y. (1996). An analysis of statistical term strength and its use in the indexing and retrieval of molecular biology texts. *Computers in Biology and Medicine*, 26(3):209–222.