Submission date: 20/09/2021
Resubmission date: 07/12/2021
Camera ready submission date: 19/02/2022

1st round notification: 02/11/2021
2nd round notification: 11/02/2022
Available online: 04/03/2022
Publication date: 04/03/2022

Section: extended version
Extended version of selected papers from the XVII Brazilian Symposium on Information Systems

Identificando Suspeitos de Crimes por meio de Interações Implícitas no YouTube

Title: Identifying Criminal Suspects through Implicit YouTube Interactions

Érick S. Florentino[1], Ronaldo R. Goldschmidt[1 2], Maria Claudia Cavalcanti[1 2]

[1]Departamento de Engenharia de Defesa – Instituto Militar de Engenharia (IME)
Rio de Janeiro, RJ – Brasil

[2]Departamento de Engenharia da Computação – Instituto Militar de Engenharia (IME)
Rio de Janeiro, RJ – Brasil

{erick.florentino, ronaldo.rgold, yoko}@ime.eb.br

Abstract. *The identification of criminal suspects on social networks (e.g., pedophilia, terrorism, etc.) has been highlighted in recent years. However, in the literature, interactions derived from the textual content posted on these networks are not always considered. Thus, the present work presents an algorithm, called TROY, capable of making these interactions and their impacts explicit in order to support the identification of suspects. Furthermore, given the difficulties in obtaining datasets in Portuguese for experiments, this work presents a new way to build a dataset for new experiments, using the link prediction task. The results obtained, through the experiments, demonstrate an improvement in the identification of suspects.*

Keywords. *Analysis; Identification; Interactions; Implicit; People; Suspects; Social Networks.*

Resumo. *A identificação de pessoas suspeitas de crimes em redes sociais (e.g. pedofilia, terrorismo etc.) tem tido destaque nos últimos anos. Contudo, na literatura, as interações geradas, a partir de conteúdos textuais postados nessas redes, nem sempre são consideradas. Desse modo, o presente trabalho apresenta o algoritmo, denominado TROY, capaz de explicitar essas interações, bem como seus impactos, a fim de apoiar a identificação de suspeitos. Além disso, perante as dificuldades em se obter dados em português para experimentos, este artigo apresenta uma nova forma de construção de conjuntos de dados para experimentos, utilizando a tarefa de predição de links. Os resultados obtidos, por meio de experimentos, demonstram uma melhora na identificação de suspeitos.*

Palavras-Chave. *Análise; Identificação; Interações; Implícitas; Pessoas; Suspeitos; Redes Sociais.*

Cite as: Florentino, É. S., Goldschmidt, R. R. & Cavalcanti, M. C. (2022). Identifying Criminal Suspects through Implicit YouTube Interactions (Identificando Suspeitos de Crimes por meio de Interações Implícitas no YouTube). iSys: Revista Brasileira de Sistemas de Informação (Brazilian Journal of Information Systems), 15(1), 3:1-3:36. DOI: 10.5753/isys.2022.2227

1. Introdução

Com o passar dos anos, a utilização de redes sociais (e.g. YouTube, Instagram, Twitter, Facebook, entre outras) vem se tornando cada vez mais popular, conectando indivíduos em todo o mundo e que pertencem aos mais diversos padrões culturais, políticos e econômicos [Lévy and Feroldi 1999]. Indo muito além dos meios de comunicação tradicionais, do tipo Um para Um, e Um para Todos, e facilitando, com isso, o intercâmbio cultural.

Existe um grande interesse socioeconômico de instituições públicas e privadas em descobrir características e padrões comportamentais de indivíduos que usam essas redes sociais para diferentes fins [Dorogovtsev and Mendes 2002]. No entanto, para isso é preciso processar um grande volume de dados. Esse é um desafio que tem sido enfrentado por abordagens, propostas na literatura, denominadas análises de redes sociais [Figueiredo 2011]. A análise de redes sociais é composta pelos mais diferentes temas de estudos, dentre os mais importantes a identificação de pessoas suspeitas de crimes tem recebido grande atenção [Pendar 2007]. Tal situação se justifica pelo fato de cada vez mais pessoas estarem usando essas redes a fim de praticarem atos ilícitos, tais como: pedofilia, terrorismo, injúria racial, homofobia, dentre outros [Fernández 2011] [dos Santos and Guedes 2020] [Costa 2019]. Na literatura, a fim de caracterizar e entender os padrões comportamentais de indivíduos em redes sociais, a maioria das pesquisas se baseiam nas comunicações realizadas entre os indivíduos (e.g. conteúdos textuais disponibilizados, interações entre pessoas, entre outros) [Villatoro-Tello et al. 2012].

O YouTube é uma das principais redes sociais, especializada no compartilhamento de vídeos com os mais diferentes temas e assuntos (ex. música, comédia, entre outros.) [Klausen et al. 2012], e é caracterizada por três níveis de comunicação [Dynel 2014]: O nível I, denominado *nível do orador e seus ouvintes*, refere-se a interações entre pessoas que atuam no próprio vídeo, com seus respectivos papéis (e.g. vídeo-aulas, programas etc.). Já no nível II, denominado *nível do remetente e seus destinatários*, tem-se como remetente o indivíduo responsável pela postagem de um vídeo, e como os destinatários os ouvintes que postam comentários e respostas sobre esse vídeo. Esse nível de comunicação se baseia na interação entre o remetente e o destinatário, a partir dos conteúdos textuais, presentes nos comentários e/ou respostas, disponibilizados no associados ao vídeo. O último, nível III, denominado *nível dos oradores e ouvintes que postam e leem comentários*, inclui as interações, por meio de comentários e respostas, que ocorrem entre os indivíduos, que interagem sobre um dado vídeo. Nesse nível, é válido ressaltar que as interações que ocorrem por meio de comentários e respostas não estão diretamente associadas ao vídeo ou seu remetente.

Geralmente, a grande maioria dos autores se baseia no nível II de comunicação para extração de padrões comportamentais no YouTube, considerando apenas os comentários dos destinatários (ou ouvintes) [Benevenuto et al. 2009], [Benevenuto et al. 2008], [Kwon and Gruzd 2017]. Todavia, cada vez mais pessoas têm realizado interações por meio de comentários e respostas em vídeos, ou seja, utilizam o nível III de comunicação. Esse nível se caracteriza por conter interações potencialmente bem mais ricas, uma vez que não considera apenas o impacto de um comentário e/ou res-

posta em um destinatário, mas também o impacto desse comentário em outros ouvintes. O nível III de comunicação possui o comportamento bem parecido com o que ocorre nas redes sociais de troca de mensagens (*messengers*), possibilitando, dessa maneira, observar e analisar as interações, para identificar comportamentos desses indivíduos no Youtube. Assim sendo, para atender o interesse socioeconômico de empresas de segurança públicas e privadas, surgem as seguintes questões de pesquisa: *a extração das interações, por meio do nível III de comunicação, possibilitaria melhorar a identificação de suspeitos?* Além disso, *como representar as interações oriundas do nível III de comunicação?*

Para responder os questionamentos acima, foi proposto o algoritmo TROY, responsável por formalizar e representar as interações entre pessoas no Youtube, considerando o nível III de comunicação e, portanto, tornando possível identificar a origem e destino do conteúdo textual postado, bem como o impacto desse conteúdo. Esse algoritmo é integrado ao INSPECTION [Florentino et al. 2020b] [Florentino et al. 2021a], um método de identificação de suspeitos em redes sociais baseado em vocabulário controlado que não necessita de conjunto de dados previamente rotulados. Experimentos iniciais realizados sobre conjuntos de dados reais mostraram que o desempenho do INSPECTION melhorou ao considerar o algoritmo TROY [Florentino et al. 2021b].

No sentido de reforçar esses resultados, novos experimentos foram realizados, considerando novos vídeos a partir de um canal com alta propensão a interação de suspeitos de pedofilia. Para marcação dos suspeitos, os conteúdos dos comentários e respostas postados por cada pessoa foram analisados minuciosamente de maneira manual. O cenário construído a partir dessa análise é denominado *CD01*. Além disso, a fim de se ter uma melhor precisão na marcação de suspeitos, considerando a dificuldade em se obter um conjunto de dados com pessoas suspeitas de pedofilia em redes sociais, é apresentado o cenário *CD02*. Esse cenário realiza a integração dos pedófilos existentes no conjunto de dados PAN-2012-BR [Andrijauskas et al. 2017] [dos Santos and Guedes 2020], a pessoas em cada um dos quatro vídeos extraídos para os novos experimentos. O PAN-2012-BR possui conversas de pedófilos liberadas por meio do Ministério Público Federal de São Paulo (MPF-SP).

Em resumo, as principais contribuições do presente trabalho são: o algoritmo TROY, bem como a integração desse algoritmo ao Método INSPECTION; um novo cenário (*CD02*) para experimentos; uma nova forma de construção de cenários de experimentos, utilizando a tarefa de predição de *links*; e os resultados dos novos experimentos que validam os resultados preliminares obtidos em [Florentino et al. 2021b], considerando tanto o cenário *CD02*, quanto o cenário *CD01*.

O presente trabalho encontra-se organizado em mais 6 seções. Na Seção 2 são apresentados conceitos das diferentes áreas da computação com que este trabalho interage. Na Seção 3 são apresentados os principais trabalhos relacionados no que tange a representação das interações existentes na rede social YouTube, considerando os níveis de comunicação, e suas diferenças em relação ao proposto neste artigo. Além disso, ainda nessa seção, são apresentados diferentes métodos/trabalhos para identificação de suspeitos e suas limitações. Tais limitações levaram a escolha do INSPECTION [Florentino et al. 2020b] [Florentino et al. 2021a] para realizar os experimen-

tos. A descrição do algoritmo TROY e o exemplo da aplicação desse algoritmo utilizando o método INSPECTION para identificação de pessoas suspeitas, encontram-se, respectivamente, nas seções 4 e 5. Já na Seção 6, além dos resultados obtidos com experimentos, é realizado um detalhamento da construção dos conjuntos de dados de acordo com dois cenários (*CD01* e *CD02*). Ainda, nessa seção, são apresentados os vocabulários e suas ponderações. Por último, na seção 7 são apresentadas as contribuições do presente artigo e trabalhos futuros.

2. Conceitos Básicos

O presente trabalho interage com diferentes áreas da computação com objetivo de realizar experimentos de análise em redes sociais. Dentre elas está a área de análise de redes complexas (seção 2.1), cujos principais conceitos são apresentados em duas subseções. Na primeira, os tipos de representações de uma rede (seção 2.1.1), as quais permite a extração de informações que demonstrem o comportamento de indivíduos. Na segunda descreve-se a tarefa de predição de link (seção 2.1.2), que sugere possíveis conexões entre esses indivíduos. Já na área de descoberta de conhecimento (seção 2.2), apresentamos os conceitos de mineração de texto fundamentais para se trabalhar com dados textuais em redes sociais, permitindo identificar termos e comportamentos por meio de dados não estruturados. Por último, a área de representação do conhecimento (seção 2.3), que oferece técnicas para representar e hierarquizar termos, bem como suas relações, dado um domínio, possibilitando identificar perfis específicos e até mesmo correlações em redes sociais.

2.1. Análise de Redes Complexas

2.1.1. Representação da Rede

O multigrafo[1] dirigido[2] com atributos[3], seja ele homogêneo[4] e/ou heterogêneo[5], é comumente utilizado, na literatura, a fim de representar redes sociais [3][17]. A seguir é apresentado um exemplo de um multigrafo homogêneo com atributos contextuais nas arestas.

Um multigrafo homogêneo dirigido $G(V, E)$, V é um conjunto de vértices que representam indivíduos ou objetos (e.g. pessoas, livros, entre outros) e E é um conjunto de arestas dirigidas que representam um tipo de relacionamento entre os vértices (e.g. mensagens, laços de amizades, entre outros). Esse tipo de multigrafo possibilita representar, por exemplo, uma mensagem m enviada e recebida, respectivamente, pelos vértices do tipo pessoa u e v, onde u e $v \in V$, $e = (u, v)$, $e \in E$. Em um multigrafo heterogêneo dirigido $G(V, E)$, o conjunto V é formado por diferentes tipos de vértices (e.g. pessoa e livros), e o conjunto E, por diferentes tipos de arestas (e.g. compra e leitura). Mais

[1]G é um multigrafo, caso tenha mais de uma aresta conectando o mesmo par de vértices, caso contrário, G é um grafo

[2]G é dirigido, se for possível, dado um par de vértices, identificar os vértices origem e destino. Caso contrário, G não é dirigido.

[3]Informações contextuais, topológicas e/ou temporais associados aos vértices ou arestas de G.

[4]G é homogêneo se possuir apenas um tipo de aresta e vértices.

[5]G é heterogêneo, caso possua mais de um tipo de vértice e/ou arestas.

formalmente, $V = V_1 \cup V_2 \cup ... \cup V_n$ e $E = E_1 \cup E_2 \cup ... \cup E_n$. Dessa maneira, V e E são formados pela união, respectivamente, de diferentes tipos de vértices e arestas dirigidas. Exemplificando, considere o multigrafo heterogêneo dirigido $G(V, E)$, onde $V = V_P \cup V_L$ e $E = E_{PL} \cup E_{LP}$. Em G, cada $v_P \in V_P$ e $v_L \in V_L$, são vértices do tipo Pessoa e Livro, respectivamente. Já cada $e_{PL} \in E_{PL}$ e $e_{LP} \in E_{LP}$, são conjuntos de arestas que representam os relacionamentos de compra e leitura, respectivamente. Assim, pode-se expressar que as pessoas v e u compraram o livro t e esse livro foi lido apenas por u, da seguinte forma: u e $v \in V_P$, $t \in V_L$, $\exists(v, t)$ e $\exists(u, t) \in E_{PL}$ e $\exists(t, u) \in E_{LP}$.

É comum, nos dois tipos de grafos apresentados, $v \in V$ e/ou $e \in E$ estarem associados a um ou mais atributos que podem representar diferentes tipos de informações acerca dos mesmos, por exemplo, informações temporais (informações cronológicas) topológica (se baseia na informação estrutural do grafo) e/ou contextual (informações de domínio). Esses atributos poderiam, por exemplo, associar a mensagem trocada entre um par de pessoas representadas no grafo.

2.1.2. Predição de Link

A Predição de Link (ou predição de ligações) é uma tarefa comumente utilizada em redes sociais, com o objetivo de identificar conexões na rede entre dois indivíduos não conectados [Liben-Nowell and Kleinberg 2007]. Essas podem ser conexões novas (conexões que poderão existir em um determinado momento) ou ausentes (deixaram de existir em algum momento na rede, porém deveriam existir) [Lü and Zhou 2011]. Essa tarefa pode ser utilizada de diversas maneiras, sendo elas: recomendação de compras virtuais [Santos 2015], de amizade em redes sociais [Aiello et al. 2012], ligações invisíveis de terroristas [Huang and Lin 2009], entre outras. Considerando a recomendação de amizades em redes sociais e as dificuldades em obter conjunto de dados com suspeitos nessas redes em português, na literatura a predição de link pode ser solução para esse problema.

Existem diversos métodos de predição de link, os quais são baseados em duas abordagens: não supervisionada e supervisionada [Wang et al. 2015]. Na abordagem não supervisionada são utilizadas métricas de similaridade (e.g. vizinhos comuns, coeficiente de jaccard, total de vizinhos, entre outras) com objetivo de gerar scores que demostrem, numericamente, o grau de similaridade entre dois indivíduos. Em seguida, os pares de vértices não conectados, são ranqueados de maneira decrescente, de acordo com seus scores, de tal maneira que os pares no topo desse rank são indicados com alta propensão se conectarem [Florentino et al. 2020a]. Diferentemente da abordagem não supervisionada, a abordagem supervisionada trabalha com classificação binária. Nessa abordagem, é necessário um conjunto de dados previamente rotulado informando quais são e não são os pares conectados. De posse dessa informação, são aplicadas métricas de similaridade, a fim de gerar scores, como citado na abordagem não supervisionada. Ao final, as informações referentes a classificação binária (conectados e não conectados), bem como os scores são disponibilizados a um algoritmo de aprendizado de máquina supervisionado (e.g. árvore de decisão, regressão logísticas, entre outros) para que ele possa aprender e, mais tarde, efetuar a predição [da Silva Soares and Prudêncio 2012].

2.2. Descoberta do Conhecimento: Mineração de Texto

Em redes sociais, informações contextuais, tais como mensagens, descrições, títulos, entre outros, são frequentemente analisadas. Devido a isso, a mineração de texto se torna uma abordagem interessante, uma vez que por meio dela é possível extrair informações uteis, utilizando técnicas computacionais [Berry Michael 2004].

A mineração de texto é um processo composto por diversas etapas, dentre as mais importantes está a etapa de indexação e normalização. Essa é composta por técnicas para reduzir e padronizar dados textuais. Dentre essas técnicas podem ser citadas, o $stemming$, que extrai o radical de um termo, por exemplo, termos "bonita" e "bonitinha" seriam reduzidos a "bonit". Uma outra técnica, pertencente a etapa de indexação e a normalização, é a remoção de $stop\ word$. Nessa técnica é feita a remoção de palavras com pouco ou nenhum significado, por exemplo, as palavras "uma", "que", entre outras.

Uma outra etapa, pertencente a técnica de mineração de texto, que possui grande importância, é o cálculo de relevância dos termos. Nessa etapa, são utilizadas medidas como a frequência do termo (da sigla em inglês, *TF*) e a frequência inversa do documento (da sigla em inglês, *IDF*). Essas medidas, respectivamente, verificam a frequência e a raridade de um termo em uma coleção de dados textuais, possibilitando representar numericamente a relevância do termo [Morais and Ambrósio 2007].

2.3. Representação do Conhecimento: Anotação Semântica

A anotação semântica, que utiliza vocabulários controlados (ou tesauros), ontologias entre outros, é uma abordagem muito usada para análise de textos [Corrêa et al. 2015] e que pode ser aplicada na análise de redes sociais [Moura 2009]. O vocabulário controlado é composto por um conjunto de termos descritores que estão semanticamente relacionados a um determinado domínio. Esses termos podem ser organizados de maneira hierárquica, ou seja, com relações hierárquicas (genérico/específico) entre si [Sales and Café 2009]. Por outro lado, uma ontologia é um recurso semântico mais sofisticado, que permite, além de representar hierarquicamente os termos de um determinado campo do conhecimento, representar vários tipos de relações entre eles, incluindo relações hierárquicas [Chandrasekaran et al. 1999].

3. Trabalhos Relacionados

Nesta seção, são apresentados trabalhos relacionados referentes a representação, considerando diferentes níveis de comunicações e informações do YouTube. Além disso, são apresentados diferentes trabalhos para identificação de suspeitos em redes sociais, bem como as limitações existentes. Tal situação justifica o uso do método INSPECTION [Florentino et al. 2021a] [Florentino et al. 2020b] para os experimentos.

3.1. O YouTube e as Representações por Níveis de Comunicação

No trabalho de [Benevenuto et al. 2008], define-se que: um vídeo é do tipo respondido se tiver pelo menos uma resposta; um usuário é dito respondido se pelo menos um dos seus vídeos é do tipo respondido; um usuário é responsivo se postou pelo menos uma resposta em um vídeo. Assim, levando em conta usuários respondidos e responsivos, essa

abordagem foca no nível (ii) de comunicação, evidenciando apenas as interações entre os usuários que postam o vídeo e as respostas diretas a ele. Por meio do algoritmo CRAWLER, é construído um multigrafo dirigido que representa as interações entre os usuários respondidos e responsivos. Em [Benevenuto et al. 2009], o algoritmo CRAWLER é utilizado para identificar usuários *spammers* e oportunistas (ou promotores de vídeo), que procuram ganhar visibilidade para um vídeo específico do YouTube.

Embora essa representação possa auxiliar na análise das interações entre usuários do Youtube, ela não abarca o nível (iii) de comunicação, onde ocorrem interações entre os usuários que respondem ao vídeo. Além disso, nesse trabalho leva-se em consideração várias informações sobre os vídeos (e.g. duração, visualização, etc.), usuários (e.g. qtd. de vídeos carregados, de amigos, etc.) e sobre as interações entre usuários e vídeos. A extração dessas informações possibilita um melhor conhecimento sobre os usuários, mas pode implicar em alto custo de processamento.

Considerando o crescimento do uso de redes sociais, com o objetivo de incitar o ódio e divulgação de conteúdo violento, em [Klausen et al. 2012] buscou-se identificar grupos jihadistas[6] que estavam usando canais do YouTube de Al-Muhajiroun com esse objetivo. Em seu trabalho, a representação da rede é feita com foco nas relações entre os canais e seus assinantes, e para isso, é utilizado um grafo heterogêneo. Desse modo, esse multigrafo é composto por dois tipos de vértices (canais e contas de usuários). Já as arestas representam a assinatura de uma conta a um canal. Com isso, é possível observar que apenas as interações entre usuários e canais são consideradas, e nesse caso também o nível (iii) de comunicação não é considerado.

Em [Kwon and Gruzd 2017], os autores buscaram analisar a agressividade em respostas a comentários e vídeos da campanha de Donald Trump à presidência dos EUA no YouTube. Para isso, além de contar com um conjunto de termos agressivos, foram definidos dois tipos de comentários: pai (comentários) e filho (respostas). Em cada comentário é verificado o uso de palavrões e a intensidades desses. Além disso, variáveis de controle dos comentários (e.g. gostei, informação temporal etc.) e vídeos (e.g. views, likes etc.) também são consideradas a fim de enriquecer a análise. Contudo, em seu trabalho, mesmo tendo identificado dois tipos de comentários, as interações entre pessoas por meio dos comentários (nível iii) não são exploradas.

Já em [Mariconti et al. 2019] foi proposto um método para identificação automatizada de vídeos no YouTube suscetíveis a ataques e discursos ódio de maneira coordenada. Nesse método, diversos aspectos dos vídeos são analisados (título, categoria, conteúdo, etc.). Esses aspectos permitem compreender os vídeos que são mais suscetíveis e são transformados em informações para o desenvolvimento de um modelo que possa ser proativo na identificação desses ataques e discursos. Diferentemente dos demais trabalhos já apresentados, [Mariconti et al. 2019] considera apenas o nível (i) de comunicação.

Em uma análise geral dos trabalhos relacionados foi observado que em nenhum deles foi considerada a interação entre diferentes pessoas, geradas a partir de comunicações feitas por meio de comentários e respostas (nível (iii) de comunicação).

[6]Grupos Jihadistas são grupos sunitas violentos [Klausen et al. 2012].

3.2. Métodos para Identificação de Suspeitos

Na literatura, é comum encontrar diversos trabalhos para identificação de suspeitos de crimes nas redes sociais. Tal situação se justifica pelo fato de cada vez mais pessoas estarem utilizando os recursos dessas redes para fins ilícitos. Muitos desses trabalhos se baseiam em uma análise contextual, ou seja, realizam uma análise do conteúdo das mensagens trocadas para identificar práticas criminais ou suspeitos.

No método, desenvolvido por [Pendar 2007] para identificação de predadores sexuais[7], se baseia em mensagens disponibilizadas por pessoas em uma rede social. Nesse trabalho, o autor conta um conjunto de dados previamente rotulado, informando predadores e não predadores. Tendo-se esse conjunto de dados rotulado, as mensagens, trocadas por essas pessoas na rede social, são tratadas com técnicas de mineração de texto (e.g. remoção de stop word, cálculo de relevância de termos, entre outros) e diferentes análises linguísticas são realizadas, por meio de Bag-of-words[8]. Em seguida, todas essas informações são fornecidas a um algoritmo de aprendizagem da máquina para gerar um modelo que caracterize as mensagens predadoras sexuais, o que permitirá a identificação de predadores sexuais em uma rede não rotulada.

Como em [Pendar 2007], os autores [Villatoro-Tello et al. 2012] e [Santos and Guedes 2019] seguiram a mesma linha em seus métodos. Em [Villatoro-Tello et al. 2012],buscou identificar tanto conversas suspeitas, quanto quais usuários são predadores sexuais. Difirindo de [Pendar 2007], o qual identifica apenas predadores sexuais. Por outro lado, de uma maneira geral, [Santos and Guedes 2019] é bem semelhante ao de [Pendar 2007]. Sua principal contribuição é que, tanto quanto sabemos, foi um dos primeiros métodos desenvolvidos para identificação de predadores sexuais em português.

Diferente dos trabalhos anteriores, em [Fire et al. 2012] e [Wang 2010] foram desenvolvidos métodos para identificar spammers em redes sociais, considerando informações topológicas. O método proposto em [Fire et al. 2012] assume que uma pessoa altamente conectada a amigos, que pertencem a várias comunidades não conectadas, possui uma grande possibilidade de ser um spammer. Em [Wang 2010] , além das informações topológicas (por exemplo, a relevância de um usuário de acordo com o número de mensagens enviadas e recebidas), também leva em consideração o conteúdo das mensagens (por exemplo, a existência de links HTML, menções a outras pessoas, entre outros). Da mesma forma que em [Pendar 2007], ambos os métodos usam a abordagem supervisionada, ou seja, eles precisam usar conjuntos de dados previamente rotulados, informando spammers e não spammers.

Em [Elzinga et al. 2012] é apresentado um método não-automatizado, usando um sistema semântico relacional de temporal, com o objetivo de analisar mensagens com conteúdo pedófilo em salas de bate-papo por um certo tempo. Os autores identificaram

[7]Predador sexual é a pessoa que para satisfação sexual pratica pornografia infantil, a extorsão e o abuso, gerando consequências físicas e/ou mentais negativas a uma outra pessoa [Pendar 2007] [Santos and Guedes 2019]

[8]Bag-of-words ou Saco de Palavras é a representação do texto de diferentes formas, respeitando a ocorrência de palavras nele [Kejriwal et al. 2016].

sete categorias de termos usados por pedófilos: saudações doces, elogios, partes íntimas, manipulações sexual, cam e fotos, onde e quando. Essas categorias caracterizam como os pedófilos estabelecem uma conexão e escalam a conversa na rede, para alertar um encontro físico. Nessa análise, cada mensagem é enquadrada manualmente em uma dessas categorias, e então eles analisam a dinâmica da conversa. Entretanto, eles não relataram a implementação de seu método para automatizar a identificação do padrão de dinâmica da conversação de pedófilos em uma rede social. Desse modo, dependendo do tamanho dos conteúdos textuais e sala a serem analisadas, o enquadramento manual pode ser inviável.

Já em [Bretschneider et al. 2014], foi desenvolvido um método para identificar mensagens de assédio em redes sociais. Com base em um conjunto de palavras profanas, eles selecionam mensagens que as mencionam para verificação posterior. Se uma mensagem contém uma palavra profana que se dirige diretamente a algumas pessoas, então elas são rotuladas como mensagens de assédio. Em uma das versões do método, chamada Naive, os autores rotulam as mensagens que mencionam palavras profanas como mensagens de assédio. No entanto, em [Bretschneider and Peters 2016], uma nova versão do método foi desenvolvida para identificar a prática de Cyberbullying. Nesta versão, aqueles que enviaram pelo menos duas mensagens de assédio para a mesma pessoa, são rotulados como um infrator de Cyberbullying. Além disso, eles calculam o grau de ofensa com base em um multigrafo dirigido, onde nós e arestas, representam, respectivamente, as pessoas e as interações dessas pessoas. Neste multigrafo, as arestas e nós possuem atributos, os quais representam, respectivamente, o grau de assédio de interação (o número de mensagens de assédio recebidas e enviadas) e o grau ofensivo (o número de pessoas ofendidas e o número de mensagens de assédio enviadas). Em ambos os trabalhos, [Bretschneider et al. 2014] e [Bretschneider and Peters 2016] não diferem os termos presentes no conjunto de palavras profanas. Tal situação pode impactar negativamente o desempenho do método. Isso, se deve ao fato de existirem termos com diferentes graus periculosidade.

Em suma, considerando as limitações apontadas nos trabalhos para identificação de pessoas suspeitas em redes sociais, apresentados nesta seção, o uso do INSPECTION [Florentino et al. 2020b] [Florentino et al. 2021a] pode ser considerada uma boa solução. Tal situação se justifica pelo fato desse método não necessitar de um conjunto de dados previamente rotulado e focar no uso de termos suspeitos de acordo com o domínio da aplicação.

4. Identificando Suspeitos com o Método INSPECTION a Partir da Representação Dada Pelo Algoritmo TROY

Esta seção apresenta o algoritmo denominado TROY para extração de interações entre pessoas por meio de comentários e respostas (Nível III de comunicação) no YouTube. O algoritmo TROY se insere como etapa Recorte da Rede no método INSPECTION [Florentino et al. 2021a][Florentino et al. 2020b] apresentado na Figura 1, cujo objetivo é a identificação de suspeitos de crimes em redes sociais. A ideia é identificar e extrair essas interações para que, a partir de um vocabulário controlado e ponderado, seja possível analisar os conteúdos textuais. Depois, de acordo com os resultados da análise busca-se classificar as pessoas. Quanto maior for a posição na classificação, mais suspeito, o que

facilita a identificação de suspeitos de crimes.

A seguir, na Seção 4.1, é apresentada uma visão geral do método INSPECTION, com a nova etapa de recorte da rede, a Seção 4.2 ilustra a aplicação do método, considerando um exemplo de interações no Youtube. Vale destacar que a etapa recorte da rede é uma extensão do método INSPECTION.

4.1. INSPECTION: Um Método para Identificação de Suspeitos

Seja um vídeo v_n qualquer entre os mais diversos vídeos (V) existentes no YouTube, selecionado por um especialista. O método INSPECTION é apresentado na Figura 1.

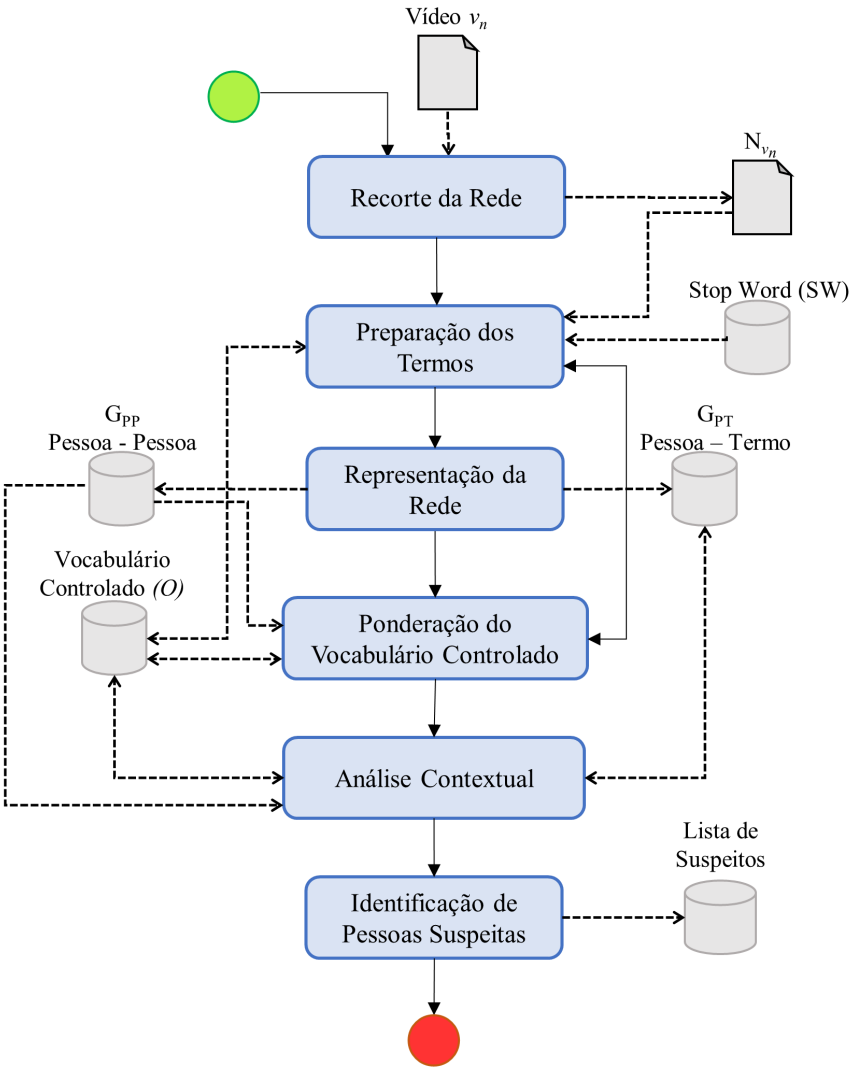

Figura 1. INSPECTION [Florentino et al. 2021a][Florentino et al. 2020b]: Método de Identificação de Suspeitos com a nova etapa de recorte da rede.

Na etapa de *Recorte da Rede*, busca-se identificar as interações sociais por meio de

dados textuais (ou mensagens), ou seja, pessoas que comentaram, que responderam e/ou foram respondidas sobre v_n (N_{v_n}). N_{v_n} é composto por um conjunto de mensagens M enviadas e/ou recebidas por pessoas de um conjunto U (N_{v_n} = [M, U]). Cada mensagem $m_x \in M$, é um conjunto ordenado de termos (m_x = $\{t_1, t_2, ..., t_n\}$)[9]. O presente método busca identificar em U, pessoas com comportamentos suspeitos, por meio dos termos utilizados nas mensagens de M ($t_j \in m_x$).

Na etapa de *Preparação dos Termos*, todos os termos t_j de cada $m_x \in M$ são tratados. Esse tratamento possibilita a padronização desses termos, criando-se um conjunto de mensagens com termos tratados, denominado M'. Com base nos conjuntos de mensagens com termos tratados M' e de pessoas U, na etapa de *Representação da Rede*, dois multigrafos são construídos a fim de representar a interação entre pessoas, e entre pessoas e termos.

Na etapa de *Ponderação do Vocabulário Controlado*, conta-se com um vocabulário controlado composto por termos suspeitos, previamente definido por um especialista, de acordo com um domínio da aplicação (e.g Pedofilia, Terrorismo, entre outros). Pressupõe-se que o vocabulário usado está dividido em grandes categorias, cobrindo os diferentes aspectos do domínio em questão. Antes de ponderar o vocabulário, inicialmente, conta-se com a visão do especialista para ponderar as categorias, que leva em conta a sua importância. Em seguida, pondera-se cada termo do vocabulário, de acordo com a sua ocorrência nas mensagens existentes no conjunto M'.

Posteriormente, na etapa de *Análise Contextual*, cada pessoa é analisada de acordo com os termos suspeitos, utilizados em mensagens enviadas, presentes no vocabulário. Ao final, a cada pessoa é atribuído um $score$, que representa numericamente o comportamento suspeito de uma pessoa. Uma vez calculados os $scores$ de todas as pessoas, na etapa de *Identificação de Pessoas Suspeitas*, essas pessoas são ordenadas de forma decrescente, onde as mais suspeitas de praticarem crimes virtuais, de acordo com o domínio da aplicação, estarão no topo da lista.

4.2. Representação do Nível de III de Comunicação por meio do Algoritmo TROY

O Algoritmo TROY (Alg. 1) implementa a etapa de *Recorte da Rede*, pois busca representar um vídeo v_n com o objetivo de gerar N_{v_n} a ser analisado. É válido ressaltar que outros algoritmos poderiam implementar essa etapa, por exemplo, o CRAWLER [Benevenuto et al. 2008].

Sendo V um conjunto responsável por conter os mais diversos vídeos postados no YouTube, cada $v_n \in V$ é representado por $v_n = [A, C, P, R]$, onde A é o autor do vídeo (ou responsável por publicá-lo), C é um conjunto de comentários, P é um conjunto de pessoas e R é um conjunto de respostas. Caso a opção de comentários de um vídeo esteja desabilitada C = ∅, P = ∅ e R = ∅. Caso contrário, cada $c \in C$ e $r \in R$ são, respectivamente, um comentário e uma resposta a um comentário feito em v_n. É válido ressaltar que uma resposta sempre estará atrelada a um comentário. Já $p \in P$ é uma

[9]Cada mensagem é um tupla $< o, d, m_x >$ contendo a pessoa origem (o), pessoa destino (d) e o conteúdo de uma mensagem (m_x). Simplificadamente, quando necessário, uma mensagem é representada por apenas m_x

pessoa envolvida no comentário ($c \in C$) e/ou resposta ($r \in R$). Cada comentário $c \in C$ é uma tupla $< id, text, temp, p >$, onde id é o identificador do comentário, $text$ é o conteúdo textual do comentário, $temp$ é a data e hora em que o comentário foi feito, e $p \in P$ é a pessoa responsável pelo comentário. Cada resposta $r \in R$ é uma tupla $< c_{id}, text, temp, p >$, onde c_{id} é o identificador de c ao qual r está conectado, $text$ é o conteúdo textual da resposta, $temp$ é a data e hora em que a resposta foi postada, e $p \in P$ é a pessoa responsável pela resposta.

Algoritmo .1: Algoritmo TROY

Entrada: Os conjuntos M e U vazios
Um vídeo $v_n \in V$, onde $v_n = [A, C, P, R]$
Saída: Conjunto U com todas as pessoas que tiveram as interações em comentários e respostas.
Conjunto M com todas as mensagens, bem como as pessoas envolvidas (emissor e receptor) pertencente a U.

```
1  início
2      #Identificando CI, RD e RDI
3      para cada c ∈ C faça
4          R_c = {r/r ∈ R e r.c_id = c.id}
5          para cada r_c ∈ R_c faça
6              M = M ∪ { < c.p, r_c.p, c.txt > }
7              M = M ∪ { < r_c.p, c.p, r_c.text > }
8          fim
9      fim
10     #Identificando RDI em respostas
11     para cada p ∈ P faça
12         Resg = {r/r ∈ R e contains(concat("@", p), r.text)}
13         para cada r_Resg ∈ Resg faça
14             R_resp = {r / r ∈ R e r.c_id = r_Resg.c_id e r.p = p }
15             para cada r_resp ∈ R_resp faça
16                 se (r_resp.temp ≤ r_Resg.temp) então
17                     M = M ∪ {< r_Resg.p, r_resp.p, r_Resg.text > }
18                     M = M ∪ {< r_resp.p, r_Resg.p, r_p.text > }
19                 fim
20             fim
21         fim
22     fim
23     U = {p/p ∈ P e ∃m ∈ M e p = m.o ou p = m.d}
24 fim
```

Um comentário $c \in C$ pode ser Impactante ou Não Impactante:

- Comentário Impactante (CI): Um comentário é dito impactante quando é respondido por uma ou mais pessoas. Formalmente, dado um $c \in C$, diz que c é impactante se $|\{r/r \in R \text{ e } c.id = r.c_{id}\}| > 0$;
- Comentário Não Impactante (CNI): Um comentário é dito não impactante quando

não tem nenhuma resposta. Formalmente, dado um $c \in C$, diz que c é não impactante se $|\{r/r \in R$ e $c.id = r.c_{id}\}| = 0$. Neste tipo de comentário, a pessoa $p \in P$ e o comentário $c \in C$ não são considerados na representação dos comentários e respostas.

Um comentário pode ser respondido por diversas pessoas. Com isso, considerando um comentário do tipo impactante, essas respostas podem ser do tipo:

- Direta (RD): É quando uma pessoa y responde o comentário feito por uma pessoa x, sem citar outras pessoas;
- Direta e Indireta (RDI): Neste tipo de resposta, uma pessoa z responde uma resposta feita por uma pessoa y.

De forma compacta, o Algoritmo TROY (Alg.1) é descrito em duas fases $F1$ e $F2$. Em $F1$ (linhas 3-9), são identificadas as trocas de mensagens do tipo: CI, RD e RDI. Para cada comentário CI, constrói-se o conjunto de respostas relacionadas àquele comentário (linhas 3 e 4). Em seguida, para cada resposta, são criadas as mensagens do conjunto M (linhas 5, 6 e 7). Uma delas parte da pessoa responsável pelo CI para a pessoa responsável pela resposta, com o conteúdo do comentário, e uma outra no sentido contrário com o conteúdo da resposta.

No conjunto de respostas R estão as respostas do tipo RD e RDI. As respostas RDI precisam ser tratadas pois embutem interações através de citações com @. Para identificar essas interações vamos para a fase $F2$ do algoritmo (linhas 11-24). Para cada pessoa $p \in P$ (linha 11), encontram-se as respostas do tipo RDI, em que a pessoa p foi citada (linha 12). Em seguida, para cada resposta com a citação $@p$ (linha 13), encontram-se as respostas cujo responsável é p e que referem-se ao mesmo comentário da resposta com a citação (linha 14). Entre essas respostas estão as que possivelmente motivaram a citação de p, e que caracterizariam a interação entre a pessoa p e quem a citou. Porém, somente as respostas que antecedem a resposta com a citação (linha 16), é que poderiam tê-la motivado. Assim, para cada resposta desse subconjunto, acrescenta-se ao conjunto M duas mensagens: uma originada da pessoa que foi citada com destino a quem a citou, e outra originada de quem citou com destino à pessoa citada, com os respectivos textos. Por último, U recebe o conjunto P, ou seja, todas as pessoas que realizaram comentários impactantes, respostas diretas e/ou indiretas (linha 18).

5. Exemplo de Aplicação

As subseções seguintes ilustram a aplicação do método INSPECTION e do algoritmo TROY, sobre uma breve conversação extraída do Youtube. O exemplo apresentado utiliza operações e fórmulas do método INSPECTION [Florentino et al. 2020b] [Florentino et al. 2021b], ilustrando cada uma de suas etapas. Cada operação e fórmula foi apresentada e brevemente explicada nas subseções seguintes

5.1. Recorte da Rede

Tendo-se como entrada um vídeo v_n, a etapa de recorte da rede busca explicitar as interações existentes, por meio de comentários e respostas em v_n, utilizando do Nível III de comunicação. Para isso, é utilizado o algoritmo TROY (Alg .1).

Exemplificando, considere como entrada um vídeo v_n, com os respectivos comentários e respostas, apresentado na Figura 2, mais as definições e fases apresentadas, é possível concluir que o usuário B fez um comentário impactante, X fez uma resposta direta, Y fez uma resposta direta e indireta e, por último, G fez um comentário não impactante.

Figura 2. Exemplificando os Tipos de Comentários e Respostas, de acordo com as definições apresentadas na Seção 4.2 .

Utilizando o Algoritmo 1 e o vídeo v_n, exemplificado acima, é possível construir N_{v_n} = [M, U], onde U=B,X,Y e M= <B, X, Novinha na praia que gataaa!>, <, B, Ridiculooooooooooooo!!!>, <B, Y, Novinha na praia que gataaa!>, <Y, B, @X Mas é uma gata mesmo.>, <X, Y, Ridiculooooooooooooo!!!>, <Y, X, @X Mas é uma gata mesmo.> . Em M, os elementos 1, 2, 3, 4 foram identificados na fase $F1$ do Algoritmo TROY (Alg. 1), já os elementos 4 e 5 na fase $F2$. Abaixo é apresentada N_{v_n} (Figura 3).

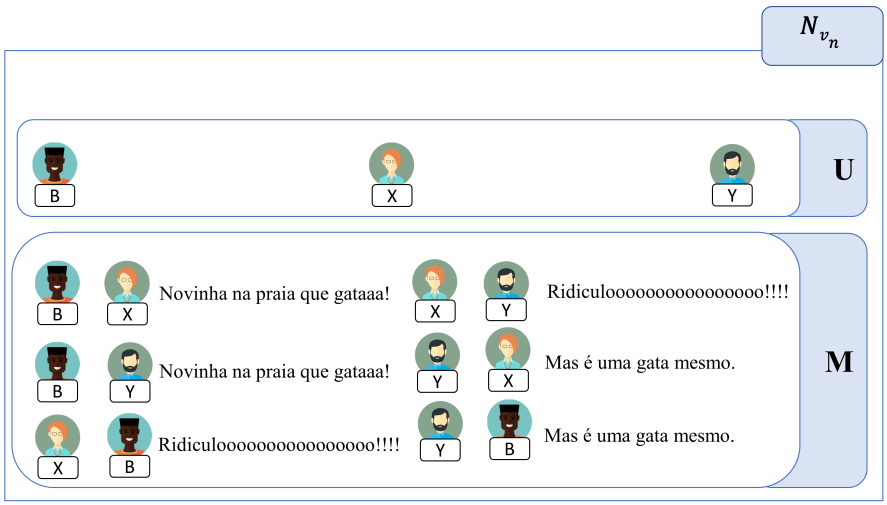

Figura 3. Representando N_{v_n} gerado por meio do TROY.

5.2. Preparação dos Termos

Nesta etapa, as subetapas de Normalização e Extração de Conteúdo Textual, Remoção de *Stop Words* e *Stemming* são executadas. Na subetapa de Normalização e Extração de Conteúdo Textual, devido à informalidade existente ao se trabalhar com dados textuais em redes sociais, para cada $t_j \in m_x$ são removidas letras repetidas, pontuações entre outros. Já na remoção de *Stop Words* (SW), busca-se remover $t_j \in m_x$ com pouco ou nenhum significado ($m_x = m_x - SW$), consequentemente, isso remove o número de termos a serem analisados. Por último, na subetapa e *Stemming*, considerando as variações de um termo, são extraídos os radicais de cada $t_j \in m_x$ ($t'_j = stem(t_j)$). Com isso M é convertido em M' = {<B, X, novinh, prai, gat>, <X, B, ridicul>, <B, Y, novinh, prai, gat>, <Y, B, gat>, <X, Y, ridicul>, <Y, X, gat>}.

5.3. Representação da Rede

A partir de M' e U, são construídos dois multigrafos dirigidos: G_{PP}, representando somente as pessoas que trocam mensagens, e G_{PT}, representando as pessoas e relacionando-as aos termos usados nas mensagens trocadas entre elas. Esses multigrafos, além de possibilitarem a identificação das pessoas emissoras e receptoras de uma ou mais mensagens ou termos, possibilitam também, realizar as análises que serão feitas nas próximas etapas.

Em $G_{PP}(V_{PP}, E_{PP})$, V_{PP} = {B, X, Y} e E_{PP} = {(B, X), (X, B), (B, Y), (Y, B), (X, Y), (Y, X)}. Em cada aresta, representamos o atributo T, que contém o texto da mensagem. Por exemplo, e_{PP_1} = (B, X), $e_{PP_1}.T$ = {novinh, prai, gat}. A Figura 4 mostra graficamente a representação de G_{PP} para o exemplo.

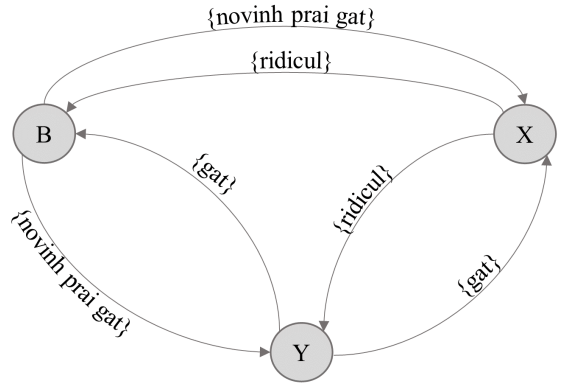

Figura 4. Representação gráfica do Multigrafo Pessoa-Pessoa (G_{PP}) construído a partir do exemplo.

Já em $G_{PT}(V_P \cup V_T, E_{PT} \cup E_{TP})$, apresentado na Figura 5, V_T é o conjunto de vértices brancos, enquanto V_P é o conjunto de vértices cinzas. As arestas dos multigrafos correspondem às ligações entre os vértices de ambos os conjuntos, V_p e V_T. Por exemplo, a aresta (Y, gat) $\in E_{PT}$, enquanto (gat, X) $\in E_{TP}$.

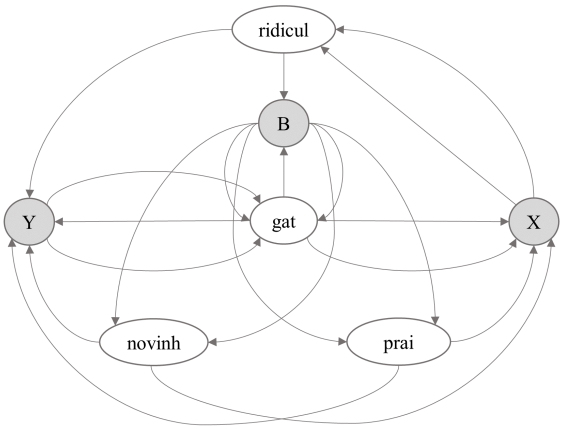

Figura 5. Representação gráfica do Multigrafo Pessoa-Termo (G_{PT}) construído a partir do exemplo.

5.4. Ponderação do Vocabulário Controlado

No método INSPECTION assume-se a existência de um vocabulário controlado formado por um conjunto de termos C_s, normalmente utilizados em mensagens de pedófilos. Este vocabulário está organizado de acordo com algumas categorias genéricas (classes C_r). Para este exemplo C_r = {Onde, Elogios, Roupas}. A Figura 6 mostra os termos utilizados neste exemplo, e a sua organização como subclasses das categorias em C_r.

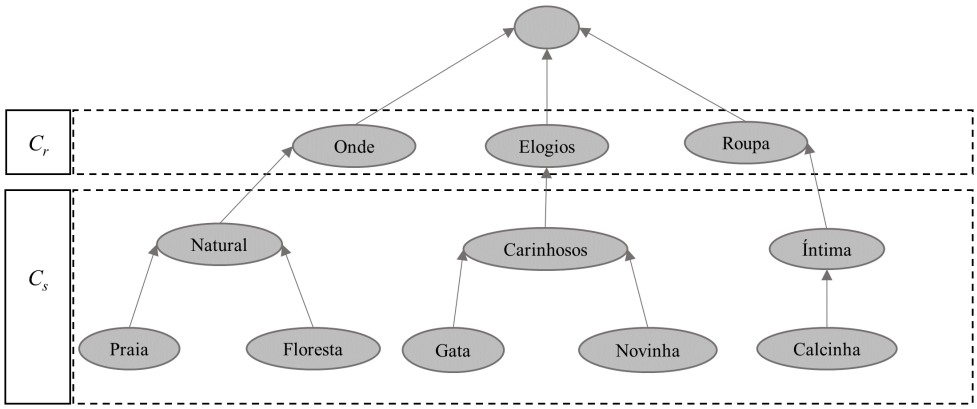

Figura 6. Vocabulário Controlado O.

Nesta etapa, cada uma destas classes raízes é ponderada por um especialista, com valores 1, 2 e 3, respectivamente. Já os termos de vocabulário (C_s) são ponderados de acordo com a sua utilização nas mensagens em análise (V_T). Portanto, para evitar o custo da ponderação de todos os termos, primeiro é necessário identificar quais deles precisam ser ponderados. Depois, a fim de os normalizar da mesma forma dos termos da mensagem, cada termo no conjunto C_s é submetido a as subetapas normalização e extração de conteúdo textual e *stemming* da etapa de Preparação dos Termos, gerando ao final o conjunto C'_s.

No exemplo dado, para realizar a ponderação das C_s, são extraídos todos os termos existentes em V_T pertencentes ao vocabulário. Para isso, é utilizada a Operação A, descrita abaixo:

$$A = C'_s \cap V_T \tag{1}$$

Exemplificando, sendo C'_s = {natur, prai, florest, carinho, gat, novinh, intim, calc} e V_T = {ridicul, gat, novinh, prai}, assim A = {gat, novinh, prai}.

Depois, para levar em conta o conteúdo das mensagens trocadas na ponderação dos termos do vocabulário, inspirada na ponderação de termos IDF (inverse document frequency), calcula-se então o Peso Global, dado por GW_{t_j} (Equação 2).

$$GW_{t'_j} = log_2 \left(\frac{|E_{PP}|}{|n_{t'_j}|} \right) \tag{2}$$

$GW_{t'_j}$ representa a frequência inversa do termo t'_j, ou a raridade do termo, no conjunto de mensagens do recorte da rede. Portanto, quanto menos o termo aparecer no conjunto de mensagens, maior é a sua importância. Desse modo, $|E_{PP}|$ é o número total de arestas em G_{PP}, e $|n_{t'_j}|$ é o número de arestas em E_{PP} que tenham o termo t'_j em mensagens. Com isso, formalmente, $n_{t'_j}$ é apresentada abaixo:

$$n_{t'_j} = \{\forall e_{pp_i} / t'_j \in e_{pp_i}.T\} \tag{3}$$

Dessa maneira, para cada $t'_j \in A$ é aplicado GW_{t_j}. Seguindo com o exemplo, tem-se:

- $GW_{novinh} = log_2 \left(\frac{6}{2}\right) = 1,58$
- $GW_{gat} = log_2 \left(\frac{6}{4}\right) = 0,58.$
- $GW_{prai} = log_2 \left(\frac{6}{2}\right) = 1,58.$

Posteriormente, para normalizar esses pesos, é feito um ajuste em relação ao peso global máximo atribuído, e em relação aos pesos de cada categoria. Inicialmente, para normalizar de acordo com o peso global máximo (HGW), este é representado pela Equação 4:

$$HGW = log_2 \left(\frac{|E_{PP}|}{1} \right) \tag{4}$$

Seguindo com o exemplo, utilizando a Equação 4, conclui-se que HGW = 2,58 ($HGW = log_2 \left(\frac{6}{1}\right) = 2,58$). Tendo-se HGW e a fim de obter as taxas de peso, para cada termo é calculado $GW^{\%}_{t'_j}$, por meio da Equação 5:

$$GW^{\%}_{t'_j} = \frac{GW_{t'_j} * 1}{HGW} \tag{5}$$

Considerando o exemplo até momento, têm-se:

- $GW^{\%}_{novinh} = \frac{1,58}{2,58} = 0,61$
- $GW^{\%}_{gat} = \frac{0,58}{2,58} = 0,23.$
- $GW^{\%}_{prai} = \frac{1,58}{2,58} = 0,61.$

Em seguida, esses valores são utilizados para recalcular o peso de cada termo e normalizá-los de acordo com os pesos das suas categorias correspondentes. Sabendo-se que *novinh* e o *gat* estão ligadas a c_r *Elogios* ($w = 2$) e *prai* ligada a c_r *Onde* ($w = 1$). Para cada uma dessas c_r é verificado o intervalo ***Max(***c_{r_j}***)*** e ***Min(***c_{r_j}***)***, respectivamente, apresentados pelas Operações 6 e 7:

$$\textit{Max(}c_{r_j}\textit{)} = c_{r_j}.w \tag{6}$$

$$Min(c_{r_j}) = Max(\{c_{r_i}.w | c_{r_i} \in C_r - \{c_{r_j}\} \wedge c_{r_i}.w < c_{r_j}.w\} \cup \{0\})). \tag{7}$$

Considerando o exemplo, $Max(Elogios) = 2$ e $Max(Onde) = 1$, já o $Min()$ é calculado abaixo:

- ***Min(Onde)*** = $Max(\{\} \cup \{0\})$ = 0
- ***Min(Elogios)*** = $Max(\{1,0\} \cup \{0\})$ = 1,0.

Finalmente, utiliza-se $GW^N_{t'_j}$ (Equação 8) para obter o Peso Global final para cada termo, dentro do intervalo da sua categoria correspondente.

$$GW^N_{t'_j}(GW^N_{t'_j} = ((Max(c_{r_k}) - Min(c_{r_k})) \times GW^{\%}_{t'_j}) + Min(c_{r_k})) \tag{8}$$

Ao considerar o exemplo, têm-se:

- $GW^N_{novinh} = ((2,0 - 1,0) \times 0,61) + 1,0$ = 1,61
- $GW^N_{gat} = ((2,0 - 1,0) \times 0,23) + 1,0$ = 1,23.
- $GW^N_{prai} = ((1,0 - 0) \times 0,61) + 0$ = 0,61.

Por último, atribui-se:

- $novinh.w$ = GW^N_{novinh} = 1,61
- $gat.w$ = GW^N_{gat} = 1,23
- $prai.w$ = GW^N_{prai} = 0,61

A Figura 7 mostra o vocabulário controlado do exemplo ponderado e normalizado:

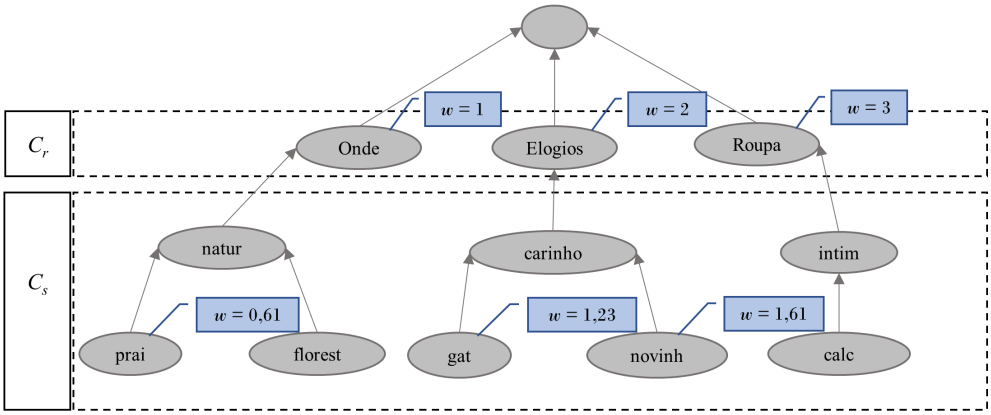

Figura 7. Vocabulário Controlado O'.

5.5. Análise Contextual

Com o vocabulário controlado ponderado e normalizado, é possível calcular os $scores$ de cada pessoa em V_P. Para isso, inicialmente, são resgatados todos os termos v_T usados por uma pessoa $v_P \in V_P$, conforme apresentado pela Operação 9 ($C_{v_T}(v_P)$). Em seguida, é verificado quais termos v_T são subclasses (C'_s) no vocabulário controlado, ou seja, são termos considerados de risco, conforme a Operação 10 ($C^{\cap}_{v_T}(v_P)$).

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

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

A Tabela 1 resume os resultados da aplicação das operações 9 e 10, considerando o exemplo:

Tabela 1. Exemplo da Análise Contextual.

v_p	$C_{v_T}(v_P)$	$C's$	$C^{\cap}_{v_T}(v_P)$
B	{novinh, prai gat}	{natur, prai, florest, carinho, gat, novinh, intim, calc}	{novinh, prai, gat}
X	{ridicul}		
Y	{gat}		{gat}

Tendo-se $C^{\cap}_{v_T}(v_P)$ de cada $v_p \in V_p$, neste momento, utilizam-se as métricas $\mathcal{M}_{GW}(v_P)$ e $\mathcal{M}_{FGW}(v_P)$. A primeira realiza a soma da raridade de cada termo suspeito usado por uma pessoa nas mensagens, normalizado pelo peso de uma classe raiz, à qual esse termo está conectado. Já a segunda, é realizada a soma da multiplicação entre raridade, informada anteriormente, e a frequência do termo, indicando a importância desse termo para uma pessoa em relação a todas as mensagens analisadas. A Tabela 2 apresenta as etapas para o cálculo dos $score$, por meio das métricas $\mathcal{M}_{GW}$ e $\mathcal{M}_{FGW}$.

Tabela 2. Cálculo dos *Scores*, por meio das métricas $\mathcal{M}_{GW}$ e $\mathcal{M}_{FGW}$

v_p	$C^{\cap}_{v_T}(v_P)$	$c_{s_i}.w$	$\mathcal{M}_{GW}(v_P)$	$W(v_P, c_{s_i})$	$\mathcal{M}_{FGW}(v_P)$
B	gat	1,23	3,45	1	3,45
	novinh	1,61		1	
	prai	0,61		1	
X	-	0	0	0	0
Y	gat	1,23	1,23	1	1,23

Com isso, $v_P.st = \mathcal{M}_{GW}(v_P)$ ou $v_P.st = \mathcal{M}_{FGW}(v_P))$

5.6. Identificação de Suspeitos

Em ambas as métricas, verificou-se que B obteve os maiores $scores$ ($\mathcal{M}_{GW}(B)$ = 3,45 e $\mathcal{M}_{FGW}(B)$ = 3,45). Isso indica que ele é o mais suspeito entre as pessoas do conjunto U. Por outro lado, X não utilizou nenhum termo considerado suspeito em seu vocabulário. Portanto, em ambas as métricas, seu $score$ foi 0.

6. Experimentos e Resultados

6.1. Conjunto de Dados

Para os experimentos foram extraídos comentários e respostas de 4 (quatro) vídeos v_n de um canal pertencente a uma cantora menor de idade. O método INSPECTION não necessita de um conjunto de dados previamente rotulado (Seção 4.1). Todavia, para que seja possível a validação desse método com o algoritmo Troy, ou outro algoritmo, essa rotulação se faz necessária. Dessa maneira, dada a dificuldade em se obter um conjunto de dados previamente rotulado, principalmente em português, foi escolhido um canal de vídeos que pudesse ser rotulado. Além disso, a escolha do canal mencionado justifica-se pois os vídeos ali publicados tratam de temas com maior probabilidade de terem conteúdos textuais suspeitos.

Na Tabela 3 são apresentados alguns dados estatísticos de cada um dos vídeos v_n:

Tabela 3. Dados Estatísticos de Cada v_n utilizado nos experimentos.

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

Cada vídeo v_n é submetido à etapa de Recorte da Rede, onde utilizando o Algoritmo TROY (Alg 1.), o qual, considerando o nível III de comunicação, explicita as interações entre as pessoas. Ao final N_{v_n} = (U, M) é gerado, onde M é um conjunto de mensagens e U um conjunto de pessoas responsáveis por essas mensagens.

O mesmo conjunto de vídeos foi submetido ao algoritmo CRAWLER ao invés do TROY, a fim de comparar o desempenho do INSPECTION com ambos os algoritmos. É

válido ressaltar que o algoritmo CRAWLER considera apenas o Nível II de comunicação. Desse modo, todas as respostas e comentários (impactantes) foram direcionados ao autor do vídeo.

Para avaliar o desempenho do método é necessário que cada pessoa $u \in U$ em N_{v_n}, esteja marcada como suspeito ou não suspeito, conforme o caso. Foram preparados dois cenários para marcação de pessoas suspeitas em N_{v_n}, gerados a partir de cada vídeo v_n. No primeiro cenário, denominado CD01, realiza-se a marcação de suspeitos por meio de uma análise minuciosa a partir de cada mensagem disponibilizada por uma pessoa em qualquer um dos vídeos v_n selecionados. Já no segundo cenário, denominado *CD02*, considerando a grande dificuldade em se obter um conjunto de dados com suspeitos de pedofilia rotulados, em português, e a fim de ter uma melhor precisão nessa rotulação, são utilizados como suspeitos os pedófilos "reais" presentes no conjunto de dados PAN-2012-BR [Andrijauskas et al. 2017][dos Santos and Guedes 2020]. A integração de N_{v_n} e os pedófilos existentes no PAN-2012-BR é feita por meio da tarefa de predição de link, utilizando a similaridade contextual entre os pedófilos e demais pessoas em cada N_{v_n} gerado a partir do v_n. No cenário *CD02*, mais uma vez, a escolha do canal e dos vídeos utilizados nos experimentos, para extração de comentários e respostas, são justificadas. Isso se deve ao fato da dificuldade em identificar reais pedófilos em um conjunto de dados com pessoas que são mais propensas a terem conteúdos textuais suspeitos, uma vez que podem compartilhar da mesma similaridade contextual. Com isso, é possível avaliar o desempenho do método e algoritmo diante dessa dificuldade.

6.1.1. Cenário *CD01*

No cenário *CD01*, para marcação de suspeitos e não suspeitos, todos os conteúdos textuais dos comentários e respostas postados nos 4 (quatro) vídeos v_n, por cada pessoa, foram analisados. Com isso, as pessoas que, em um contexto geral, tiveram mais comentários e respostas com conteúdos impróprios e/ou inadequados são marcadas como suspeitas, bem como as mensagens desses conteúdos.

A Figura 8, apresenta a etapa de recorte da rede considerando o cenário *CD01* para marcação de suspeitos.

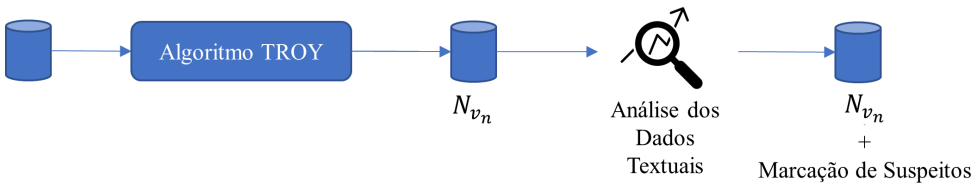

Figura 8. Marcação de suspeitos no Cenário *CD01*.

As Tabelas 4 e 5 apresentam dados estatísticos de cada N_{v_n} gerado a partir de cada v_n utilizando o cenário *CD01* e, respectivamente, os algoritmos TROY e CRAWLER. É

válido relembrar que o TROY utiliza o Nível III de comunicação. Já o algoritmo CRAWLER, utiliza o Nível II de comunicação, ou seja, todos os conteúdos são direcionados ao autor do vídeo.

Tabela 4. Dados Estatísticos de Cada Vídeo N_{v_n} utilizando o Algoritmo TROY e o Cenário *CD01*

Vídeo	|U|	|M|	|U| Suspeitas	|M| Suspeitas
v_1	1.767	7.639	33	469
v_2	67	145	1	3
v_3	3.649	14.084	92	1.567
v_4	5.216	15.760	101	1.096

Tabela 5. Dados Estatísticos de Cada Vídeo N_{v_n} utilizando o Algoritmo CRAWLER e o Cenário *CD01*

Vídeo	|U|	|M|	|U| Suspeitas	|M| Suspeitas
v_1	1.767	3.647	33	240
v_2	67	106	1	1
v_3	3.649	6.259	92	489
v_4	5.216	7.486	101	343

6.1.2. Cenário *CD02*

Considerando o fato de os vídeos selecionados serem mais suscetíveis ao interesse de praticantes de pedofilia, conforme já dito, e a fim de ter uma melhor precisão na marcação de suspeitos, foram utilizados os suspeitos de pedofilia presentes do conjunto de dados PAN-2012-BR [Andrijauskas et al. 2017] [dos Santos and Guedes 2020]. Esse conjunto de dados embasado em conversas disponibilizadas fornecido pelo Ministério Público Federal de São Paulo (MPF-SP) em parceria com o Centro Universitário da Fundação Educacional Inaciana (FEI). Desse modo, o cenário *CD02* é apresentado na Figura 9.

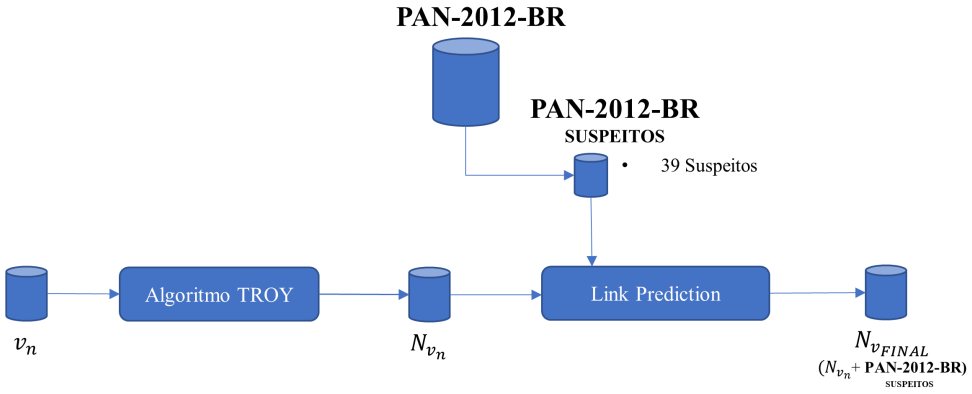

Figura 9. Inclusão de suspeitos no Cenário *CD02*.

Inicialmente, todos os suspeitos são extraídos do conjunto de dados PAN-2012-BR, gerando o PAN-2012-BR SUSPEITOS. Tendo-se os suspeitos, a etapa de Link Prediction é responsável por realizar as ligações entre os suspeitos PAN-2012-BR SUSPEITOS e as pessoas em N_{v_n}, bem como o direcionamento das mensagens entre essas pessoas. Para efetuar tais conexões e direcionamentos, tendo-se N_{SUSP} gerado do PAN-2012-BR SUSPEITOS, é apresentado o Algoritmo 2 composto pelas seguintes etapas:

- *Preparação dos Termos* (linha 3 - 8 do Algoritmo 2) . Sabendo-se que $N_{SUSP} = \{M_{SUSP}, U_{SUSP}\}$, onde U_{SUSP} é um conjunto de pessoas envolvidas em mensagens $m \in M_{SUSP}$ e M_{SUSP} é um conjunto de mensagens disponibilizadas por cada pessoa $u \in U_{SUSP}$, em PAN-2012-BR SUSPEITOS. Para cada mensagem $m \in M_{SUSP}$ de N_{SUSP} (linha 3 do Algoritmo 2) essa é preparada por um conjunto de técnicas de mineração de texto (função *prepara_termo()* na linha 4 do Algoritmo 2), tais como: remoção de *stop word*, *stemming*, entre outros. Por outro lado, o mesmo acontece com cada $m \in M$ de N_{v_n} (linha 6 do Algoritmo 2), sendo assim, cada mensagem m também é preparada pelo mesmo conjunto de técnicas de mineração de texto ($m \in M$) (linha 7 do algoritmo 2).
- *Predição:* A partir dos conteúdos textuais tratados, busca-se verificar a similaridade textual de cada pessoa $u \in U_{SUSP}$ de N_{SUSP} com cada $v \in U$ de N_{v_n} (U_{SUSP} x U, linha 10-11 do Algoritmo 2). Essa similaridade é representada por um $score$, obtido por meio de uma métrica (e.g. coeficiente de Jaccard, Vizinhos Comuns, entre outras) (linha 12 do Algoritmo 2). No presente trabalho, foi escolhida a métrica Coeficiente de Jaccard (*CJ*, Eq. 11). Essa métrica diz que quanto mais termos comuns existirem entre u e v, maior será a similaridade contextual entre eles, consequentemente, maior a possibilidade de conexão. Desse modo, na equação 11, reg é responsável por resgatar todos os termos de mensagens ($m \in M$) utilizadas por uma pessoa [Muniz et al. 2018]

$$metrica(u, v) = CJ(u, v) = \frac{|reg(u) \cap reg(v)|}{|reg(u) \cup reg(v)|} \tag{11}$$

  O $score$ de cada pessoa u e v, respectivamente pertencentes a N_{SUSP} e N_{v_n} , obtido por meio de uma métrica é armazenado em uma $lista$, bem como as pessoas envolvidas (linha 13 do Algoritmo 2).
- *Ranqueamento:* Tendo-se a $lista$ com a similaridade contextual, por meio de $scores$, de cada pessoa v e u, respectivamente pertencentes a N_{v_n} e N_{SUSP}. Neste momento é efetuado ranqueamento dessas pessoas, de acordo com $score$ obtido. Para isso, são eliminados da $lista$ pessoas que tiveram $score$ menor ou igual a 0, ou seja, não tiveram nenhuma similiaridade textual (linha 16 do Algoritmo 2). Em seguida, utilizando os $scores$, realiza-se o ranqueamento de maneira decrescente, de tal maneira que quem está no topo possui uma maior propensão a conexões (linha 18 do Algoritmo 2). Feito isso, a $lista$ é dividida em decis e, ainda na lista, é criado um novo campo, denominado como "decile", responsável por armazenar em que decil, cada u e v, com seu respectivo $score$, pertence (linha 19 do Algoritmo 2). Por último, apenas as pessoas que se encontram no primeiro decil são selecionadas, pois possuem a maior similaridade textual (linha

20 do Algoritmo 2). Além disso, em experimentos preliminares no PAN-2012-BR [Andrijauskas et al. 2017] [dos Santos and Guedes 2020], verificou-se que o primeiro decil tinham mais reais conexões entre pessoas.

- *Seleção:* Tendo-se a $lista$, com as pessoas que se encontram no primeiro decil, verificou-se que um suspeito $u \in U_{SUSP}$ de N_{SUSP}, pode ter mais interações com outras pessoas do que cada pessoa $v \in U$ de N_{v_n}. Essa situação pode trazer ruídos nas análises e fazer com que $u \in U_{SUSP}$ de N_{SUSP} tornem-se *outliers* na criação dessas conexões. Desse modo, é necessário normalizar as interações de $u \in U_{SUSP}$ de N_{SUSP} predito no primeiro decil. Para isso, em N_{v_n}, a média de conexões entre todas as pessoas $v \in U$, é verificada. Essa média serve como um limite de conexões em que cada suspeito $u \in U_{SUSP}$ de N_{SUSP} poderá ter. Para o cálculo dessa média, inicialmente, cria-se max_c, esse é responsável por ter todas as conexões existentes em M de N_{v_n} (linha 22 do Algoritmo 2). Feito isso, são eliminadas as possíveis conexões duplicadas (linha 23 do Algoritmo 2). Em seguida, é realizado o agrupamento de cada v contando o número de conexões que ela possui (linha 24 do Algoritmo 2), esse agrupamento é representado por "*total_connection*" no algoritmo. Tendo-se o total de conexões de cada pessoa v é identificada a média arredondada de interações em N_{v_n}, por meio da linha 25 do Algoritmo 2, essa média é dada por max_c nesse Algoritmo.

  Tendo-se a $lista$, considerando o 1º decil, e a média de interações, representada por max_c, em N_{v_n}, neste momento, se inicia a normalização de $u \in U_{SUSP}$ de N_{SUSP} pertencentes a $lista$. Para isso, é feito um novo ranqueamento considerando cada u e seus respectivo score, bem como a ordem que ele aparece na $lista$ (linha 26 do Algoritmo 2), uma vez que já foi ordenado na linha 18 do Algoritmo 2. Por último, com o *rank*, na $lista$, são considerados apenas as posições que sejam inferior ou igual a max_c (linha 27 do Algoritmo 2), ou seja, a média de interações.

- *Conexões:* Neste momento, tendo-se $lista$ com as conexões preditas, considerando a média de interações em N_{v_n}, são feitas as conexões entre as pessoas u e v, respectivamente, pertecentes a N_{SUSP} e N_{v_n}. Dessa maneira, para cada par de conexão predita na $lista$ (linha 29 do Algoritmo 2) são resgatadas todas as mensagens disponibilizadas por u em M_{SUSP} de N_{SUSP} (linha 30 do Algoritmo 2) e v em M de N_{v_n} (linha 31 do Algoritmo 2). Esses resgates, são respectivamente, representados por X e Y. Feito isso, o objetivo é efetuar a ligação entre as pessoas preditas, assim como o conteúdo textual das mensagens (linha 33 e 37 do Algoritmo 2), gerando ao final M_{FINAL}. Contudo, no que tange às mensagens das pessoas $v \in U$ de N_{v_n}, apenas os comentários impactantes e respostas diretas são direcionadas aos suspeitos. Tal situação se deve ao fato de as respostas diretas e indiretas serem mais direcionadas as pessoas específicas. Devido a isso, mensagens que se iniciam com "@" não são consideradas, pois é a forma em que, normalmente, pessoas são citadas. Além disso, o conjunto de U_{FINAL} é criado a partir de pessoas presentes na $lista$ (linha 40 -41 do Algoritmo 2). Ao final, é gerado $N_{v_{FINAL}}$.

Algoritmo .2: Construção do CD02

Entrada: N_{SUSP} e N_{v_n}
Saída: $N_{v_{FINAL}}$

início
  #————————-Preparação dos Termos
  para cada $m \in M_{SUSP}$ *de* N_{SUSP} **faça**
    m = $prepara\_termo(m)$
  fim
  para cada $m \in M$ *de* N_{v_n} **faça**
    m = $prepara\_termo(m)$
  fim
  #————————-Predição
  para cada $u \in U_{SUSP}$ *de* N_{SUSP} **faça**
    para cada $v \in U$ *de* N_{v_n} **faça**
      $score$ = $metrica(u, v)$
      $lista$ = $lista \cup < u, v, score >$
    fim
  fim
  $lista$ = $lista[lista.score > 0]$
  #————————-Ranqueamento
  $lista$ = $sorted(lista, by =$"$score$"$)$
  $lista.decile$ = $decile(lista, by =$"$score$"$)$
  $lista$ = $lista[lista.decile = 1]$
  #————————-Seleção
  $max\_c =< v, a >\in M$ *de* N_{v_n}
  $max\_c = max\_c.drop\_duplicates()$
  $max\_c = max\_c.groupby($"v"$).count(as =$"$total\_connection$"$)$
  $max\_c = round(max\_c.total\_connection.mean())$
  $lista.rank$ =
  $lista.groupby(lista.v)[lista.score].rank($"$first$"$, ascending = False)$
  $lista = lista[lista.rank \leq max\_c]$
  #————————-Conexões
  para cada $< u, v >\in lista$ **faça**
    X = $\{< m_x > /o = u$ e $m_x \in M_{SUSP}$ de $N_{SUSP}\}$
    Y = $\{< m_y > /o = v$ e $m_y \in M$ de $N_{v_n}\}$
    para cada $m_x \in X$ **faça**
      M_{FINAL} = $M_{FINAL} \cup < u, v, m_x >$
    fim
    para cada $m_v \in Y$ **faça**
      se *notstartwith("@", m_y)* $e < v, u, m_y > \notin M_{FINAL}$ **então**
        M_{FINAL} = $M_{FINAL} \cup < v, u, m_y >$
      fim
    fim
    U_{FINAL} = $U_{FINAL} \cup \{< u > /u \notin U_{FINAL}\}$
    U_{FINAL} = $U_{FINAL} \cup \{< v > /v \notin U_{FINAL}\}$
  fim
  $N_{v_{FINAL}}$ =$\{U_{FINAL}, M_{FINAL}\}$
fim

Considerando cada um dos vídeos v_n utilizado nos experimentos, o algoritmo TROY e o cenário *CD02*, foi possível chegar nos dados estatísticos apresentados na Tabela 6. Por outro lado, com o algoritmo CRAWLER e o *CD02* , os dados estatísticos extraídos são apresentados na Tabela 7.

Tabela 6. Dados Estatístico de Cada $N_{v_{FINAL}}$ utilizando o Algoritmo TROY e o Cenário *CD02*

	$\mathbf{N}_{v_{FINAL}}$				
Vídeo	|U|	|M|	Média de Interações	|U| Suspeitas	|M| Suspeitas
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

Tabela 7. Dados Estatístico de Cada $N_{v_{FINAL}}$ utilizando o Algoritmo CRAWLER e o Cenário *CD02*

	$\mathbf{N}_{v_{FINAL}}$				
Vídeo	|U|	|M|	Média de Interações	|U| Suspeitas	|M| Suspeitas
v_1	1.806	3.911	3	39	264
v_2	87	209	2	20	103
v_3	3.688	6.523	3	39	264
v_4	5.255	7.751	3	39	264

6.2. Vocabulário Controlado

Para construção do vocabulário controlado (O_1 e O_2) foram utilizadas, como classes raízes, seis categorias citadas em [Elzinga et al. 2012]: "onde", "quando", "partes íntimas", "manipulações sexuais", "fotos e câmera" e "elogios". No vocabulário controlado O_2, é inserida a classe raiz "roupa", com o objetivo de verificar se com enriquecimento do vocabulário controlado é possível obter melhores resultados. Além disso, a inserção dessa classe raiz se deve ao fato de haver vários termos relacionados, comumente utilizados por pedófilos, como exemplo, calcinha, sutiã, saia, cueca, entre outros.

É válido ressaltar que o vocabulário é composto por diversos termos suspeitos que são comuns a ambos os sexos. Mas, na classe raiz partes íntimas, por exemplo, existem subclasses dividas em períneo masculino e feminino. Além disso, derivações informais dos termos também são consideradas, podendo citar como exemplo, o termo pênis (termos formal) e pinto (termo informal/alternativo), ambos pertencentes ao períneo masculino. Além disso, todas as classes raízes possuem termos comumente utilizados no domínio em questão, a pedofilia. Sendo, dessa maneira, essencial para identificação de pessoas suspeitas de pedofilia.

Em relação à etapa de ponderação das classes raízes do vocabulário controlado, isso é feito com o apoio de um especialista no domínio em questão. Dessa maneira, neste experimento, houve apoio de um Policial Federal de Aracaju, que é especialista e atua na investigação de casos de pedofilia há 11 anos. Após a apresentação do método a ele, foi

solicitado que as ponderações das classes fossem realizadas no intervalo de 1 até o número máximo de classes raízes. Sendo assim, a ponderação das classes raízes dos vocabulários O_1 e O_2 variaram, respectivamente, de 1 a 6 e de 1 a 7.

Segundo o especialista, a interação de um sujeito que pratica pedofilia, inicialmente, tem o objetivo de ganhar a sua confiança. Em seguida, o sujeito começa a demandar fotos, filmes, etc. Assim sendo, normalmente utilizam-se muitas vezes termos elogiosos e carinhosos, como gatinha, fofinha, etc. Por esse motivo, atribui-se bastante importância a termos desta classe raiz. Com relação aos termos das classes quando e onde, estes não são tão importantes, não só por serem muito comumente usados em qualquer conversa, mas especialmente porque o objetivo principal de um pedófilo não é marcar encontros, mas obter imagens, fotos e filmes. De acordo com o Estatuto da Criança e do Adolescente (Art. 241. Redação dada pela Lei no 11.829, de 2008), a transmissão ou armazenamento de fotos pornográficas que envolvam crianças já caracteriza crime de pedofilia. Assim, a essa classe raiz (Fotos e Câmera) o especialista atribuiu o maior valor. Termos relacionados a Partes Íntimas também são importantes, e por isso também ganharam um valor significativo. Do mesmo modo, os termos de manipulação sexual são relevantes, mais importantes que os elogios, porém ainda abaixo das classes raízes de fotos e partes íntimas que são mais largamente usados por esses sujeitos.

Foi solicitado ao especialista que as ponderações das classes raízes fossem feitas de duas maneiras, utilizando números decimais e/ou inteiros e, outra, utilizando apenas números inteiros. O objetivo é verificar o desempenho do método tendo-se um vocabulário ponderado de maneira enrijecida (por meio de números inteiros) e mais flexível (utilizando números decimais e/ou inteiros). A Tabela 8 apresenta as ponderações desses vocabulários, segundo o especialista.

Tabela 8. Ponderações dos Vocabulários Controlados e origens das subclasses

	O_1^{INT}	O_2^{INT}	O_1^{REAL}	O_2^{REAL}	
c_r	Peso (w)	Peso (w)	Peso (w)	Peso (w)	C_s
Quando	1	1	1.5	1.5	[Scheider and Kiefer 2018]
Onde	2	2	2.3	2.3	[Hobbs and Pan 2006]
Elogios	3	3	4.0	4.0	[Neves nd]
Fotos e Câmera	4	5	6.0	6.0	[Mukherjee and Joshi 2013]
Partes Íntimas	5	6	5.7	5.7	[Rosse and Mejino 2008]
Manipulações Sexuais	6	7	5.5	5.5	[Kronk et al. 2019]
Roupa	-	4	-	5.0	[Kuang et al. 2018]

Desse modo, 4 (quatro) vocabulários controlados foram desenvolvidos e ponderados: O_1^{INT} (não considera a classe raiz roupa e a ponderação foi feita utilizando números inteiros), O_2^{INT} (considera a classe raiz roupa e a ponderação foi feita utilizando números

inteiros), O_1^{REAL}(não considera a classe raiz roupa e na ponderação podem ser utilizados números decimais) e O_2^{REAL}(considera a classe raiz roupa e na ponderação podem ser utilizados números decimais).

6.3. Avaliação do Desempenho

Com o objetivo de avaliar o desempenho do método INSPECTION [Florentino et al. 2020b][Florentino et al. 2021a] utilizando os algoritmos TROY [Florentino et al. 2021b], e o CRAWLER [Benevenuto et al. 2008], com os diferentes conjuntos de dados e ponderações dos vocabulários controlados, foi utilizada a Área sobre a Curva (AUC) [Li et al. 2018]. Nos experimentos, essa medida calcula a probabilidade de um suspeito de pedofilia sempre ter um score superior a um não suspeito, ambos escolhidos aleatoriamente n vezes. Nos experimentos, n foi fixado em 100 (cem).

6.4. Resultados

6.4.1. Cenário *CD01*

A Tabela 9 apresenta os resultados obtidos considerando o cenário *CD01*. É válido relembrar que nesse cenário é feita uma análise minuciosa de todo o conteúdo textual disponibilizado por uma pessoa em todos os vídeos utilizados. Com isso aquelas pessoas que disponibilizam mais conteúdos textuais impróprios e/ou inadequados, por meio de comentário e respostas, são marcadas como suspeitas.

Em relação aos resultados, inicialmente foi observado que tanto TROY (que utiliza o Nível III de comunicação), quanto CRAWLER (que utiliza Nível II de comunicação), em todos os cenários e com ambas as métricas, obtiveram uma AUC superior a 0.5. Desse modo, proporcionaram resultados superiores ao preditor aleatório ($AUC > 0,5$), mostrando a capacidade do método INSPECTION em identificar suspeitos, com as representações dadas por ambos os algoritmos. A menor AUC foi obtida com a representação dada pelo CRAWLER, com a métrica $\mathcal{M}_{FGW}$ e o vocabulário O_2^{INT}.

No comparativo TROY e CRAWLER, considerando todas as configurações (vocabulário, ponderação do vocabulário e métricas), o TROY obteve melhores resultados em todos os vídeos, uma melhora geral, considerando esses resultados, de 3,2%. O melhor resultado, obtido com o TROY (AUC = 0,995) e CRAWLER (AUC = 0,985), foi no vídeo v_2 com a métrica $\mathcal{M}_{GW}$ utilizando a ponderação e métrica O_2^{INT} (destacado em negrito e itálico, na Tabela 9).

Em uma análise por vídeo, constatou-se que no vídeo v_1, com o TROY, a melhor AUC foi obtida com o vocabulário e ponderação O_1^{REAL} com a métrica $\mathcal{M}_{GW}$ (AUC = 0,810, destacado em negrito). Com o CRAWLER, no mesmo vídeo, a melhor AUC foi obtida com a mesma métrica, vocabulário e ponderação (AUC = 0,805, destacado em negrito). Em v_2, como já informado anteriormente, os melhores resultados foram obtidos com a métrica $\mathcal{M}_{GW}$, vocabulário e ponderação O_1^{INT}. Em v_3, com o TROY e CRAWLER, a melhor AUC foi obtida com a métrica $\mathcal{M}_{GW}$. Mas a melhor ponderação e vocabulário para o TROY foi O_2^{REAL} (AUC = 0,830, destacado em negrito), já para o CRAWLER foi o O_1^{INT} (AUC = 0,780, destacado em negrito). Por último, em v_4, no

TROY, foram as ponderações e vocabulários O_2^{REAL} e O_1^{REAL}, respectivamente, obtidos com as métricas $\mathcal{M}_{GW}$ e $\mathcal{M}_{FGW}$ que levaram a melhor AUC (AUC = 0,870, destacado em negrito na Tabela 9). Com o CRAWLER a melhor AUC (AUC = 0,845, destacado em negrito e itálico) foi obtida também usando vocabulário e ponderação O_1^{REAL}, mas com a métrica $\mathcal{M}_{GW}$.

Resumidamente, considerando os melhores resultados por vídeo, descrito acima, foi possível observar que a ponderação com números inteiros e/ou decimais, no TROY, levaram a melhores resultados em três dos quatro vídeos. Já com o CRAWLER, os resultados ficaram similares com ambos os tipos de ponderação. Em relação ao vocabulário, no TROY os melhores resultados foram obtidos tanto com O_1 (sem a classe raiz "roupa") e O_2 (com a classe raiz "roupa"), porém com uma maior predominância para o vocabulário O_1 . Com o CRAWLER, foi o vocabulário O_1 que levou os melhores resultados por vídeo. Em ambos os algoritmos a melhor métrica foi a $\mathcal{M}_{GW}$.

De uma maneira geral, foram 32 (trinta e dois) resultados para cada representação, utilizando os algoritmos TROY ou CRAWLER. Em 28 (vinte e oito) desses resultados, o TROY obteve uma AUC superior ao CRAWLER (resultados em que o CRAWLER obteve resultados superiores estão destacados em vermelho abaixo). Apena no vídeo v_1 com a métrica $\mathcal{M}_{FGW}$, vocabulários e ponderações O_2^{REAL} e O_1^{REAL}, ambos os algoritmos empataram (célula destacada em verde, na Tabela 9). Em relação as configurações em que o CRAWLER obteve melhores resultados (dois casos, destacados em vermelho na Tabela 9). Esses foram obtidos com as métricas $\mathcal{M}_{FGW}$ com vocabulários bem distintos (O_1^{REAL} e O_2^{INT})

No cenário *CD01*, conclui-se que o TROY, em um comparativo com o CRAWLER, se saiu melhor em 87,5% dos resultados.

Tabela 9. Resultados dos Experimentos Considerando o Cenário *CD01*

		TROY				CRAWLER			
	Métrica	O_1^{INT}	O_2^{INT}	O_1^{REAL}	O_2^{REAL}	O_1^{INT}	O_2^{INT}	O_1^{REAL}	O_2^{REAL}
v_1	$\mathcal{M}_{GW}$	0,730	0,765	**0,810**	0,785	0,715	0,745	**0,805**	0,755
	$\mathcal{M}_{FGW}$	0,750	0,665	0,755	0,760	0,720	0,660	0,755	0,760
v_2	$\mathcal{M}_{GW}$	***0,995***	0,965	0,925	0,910	***0,985***	0,950	0,875	0,850
	$\mathcal{M}_{FGW}$	0,990	0,990	0,950	0,980	0,960	0,965	0,925	0,920
v_3	$\mathcal{M}_{GW}$	0,815	0,765	0,780	**0,830**	**0,780**	0,755	0,760	0,775
	$\mathcal{M}_{FGW}$	0,785	0,820	0,710	0,785	0,715	0,745	0,750	0,730
v_4	$\mathcal{M}_{GW}$	0,835	0,820	0,865	**0,870**	0,830	0,800	**0,845**	0,820
	$\mathcal{M}_{FGW}$	0,835	0,805	**0,870**	0,855	0,790	0,815	0,840	0,840

6.4.2. Cenário *CD02*

Nesta seção, são apresentados os resultados obtidos considerando o cenário *CD02*. É válido relembrar que nesse cenário são utilizados, como suspeitos de pedofilia, os 39

pedófilos existentes no conjunto de dados PAN-2012-BR. Esses suspeitos são integrados a cada um dos vídeos v_n utilizados, por meio da tarefa de predição de link. Nessa tarefa, é verificada a similaridade contextual entre as pessoas e as que tiverem maiores similaridades, e para esses pares de pessoas similares, as conexões são criadas.

Neste cenário, como no cenário *CD01*, todos os algoritmos proporcionaram uma AUC superior a 0.5. Dessa forma, mostrando a capacidade do INSPECTION em identificar suspeitos com ambas as representações mais uma vez. Em um comparativo entre o TROY e CRAWLER, considerando todas as configurações (vocabulário, ponderação do vocabulário, e métricas), o TROY obteve melhores resultados em todos os vídeos, uma melhora geral, considerando esses resultados, de 2,3%. O melhor resultado, obtido com o TROY, foi no vídeo v_1 com uma AUC de 0,905 (destacado em negrito e itálico na Tabela 10). Já com o CRAWLER, a melhor AUC foi de 0,892 (destacado em negrito e itálico na Tabela 10) no vídeo v_4.

Realizando uma análise por vídeo, no vídeo v_1 a melhor AUC foi obtida com o vocabulário e ponderação O_2^{REAL} com a métrica $\mathcal{M}_{FGW}$ (AUC = 0,905, destacado em negrito e itálico). Com o CRAWLER, a melhor AUC foi obtida com a mesma métrica, mas desta vez, com o vocabulário e ponderação O_2^{INT} (AUC = 0,875, destacado em negrito). Em v_2, com o TROY e CRAWLER, os melhores resultados foram obtidos com a métrica $\mathcal{M}_{GW}$. Porém, enquanto com o TROY o melhor resultado foi com o vocabulário e ponderação O_1^{REAL} (AUC = 0,885, destacado em negrito), com o CRAWLER, foi com o O_2^{REAL} (AUC = 0,842, destacado em negrito na Tabela 10). Em v_3, com o TROY a melhor AUC foi obtida com a métrica $\mathcal{M}_{FGW}$ (AUC = 0,897, destacado em negrito) com a ponderação e vocabulário O_1^{REAL}. No CRAWLER, foi a métrica $\mathcal{M}_{GW}$ que levou a melhores resultados (AUC = 0,850, destacado em negrito na Tabela 10). No último vídeo, v_4, foi a ponderação e vocabulário O_2^{INT} que levou a melhores resultados (AUC = 0,900, destacado em negrito). Sendo obtidos com a métrica $\mathcal{M}_{GW}$ no TROY e $\mathcal{M}_{FGW}$ com o CRAWLER (AUC = 0,892, destacado em negrito e itálico na Tabela 10).

Resumidamente, com o TROY as métricas $\mathcal{M}_{FGW}$ e $\mathcal{M}_{GW}$ tiveram resultados similares, mas em relação ao vocabulário e ponderação, a maior relevância foi para o O_1^{REAL}. Já com o CRAWLER, as métricas tiveram o mesmo desempenho do que no TROY, mas dessa vez com o vocabulário e ponderação O_2^{INT}, podendo ser identificados em 3 (três) dos 4 (quatro) vídeos. Dessa maneira, é possível concluir que no CRAWLER a classe raiz "roupa" foi responsável por levar a melhores resultados. Já com o TROY, o comportamento foi bem parecido para ambos os vocabulários, sem e com a classe raiz "roupa". Em relação a ponderação do vocabulário, no TROY os melhores AUC foram obtidas utilizando números inteiros e/ou decimais em três vídeos. Como no TROY, o CRAWLER também teve melhores resultados em três vídeos, mas, nesse momento, com a ponderação feita por números inteiros.

Ainda, fazendo uma análise por métrica, vocabulário e ponderação, foram 32 (trinta e dois) resultados. Em 22 (vinte e dois) desses resultados, o TROY obteve uma AUC superior ao CRAWLER (resultados em que o CRAWLER obteve resultados superiores estão destacados em vermelho na Tabela 10) e em apenas no vídeo v_3 com a métrica $\mathcal{M}_{GW}$ e vocabulário O_2^{REAL} ambos os algoritmos empataram (célula destacada em verde

na Tabela 10). Em relação as configurações em que o CRAWLER obteve melhores resultados (destacados em vermelho na Tabela 10), houve uma maior predominância (4 de 7 resultados) no vocabulário ponderado com números inteiros, métrica $\mathcal{M}_{FGW}$ e no vocabulário sem a classe raiz roupa O_1.

Tabela 10. Resultados dos Experimentos Considerando o Cenário *CD02*

		TROY				CRAWLER			
	Métrica	O_1^{INT}	O_2^{INT}	O_1^{REAL}	O_2^{REAL}	O_1^{INT}	O_2^{INT}	O_1^{REAL}	O_2^{REAL}
v_1	$\mathcal{M}_{GW}$	0,845	0,865	0,823	0,865	0,833	0,810	0,841	0,860
	$\mathcal{M}_{FGW}$	0,836	0,849	0,880	***0,905***	0,854	**0,875**	0,825	0,860
v_2	$\mathcal{M}_{GW}$	0,824	0,860	**0,885**	0,845	0,795	0,765	0,804	**0,842**
	$\mathcal{M}_{FGW}$	0,707	0,835	0,755	0,810	0,755	0,810	0,805	0,750
v_3	$\mathcal{M}_{GW}$	0,854	0,819	0,842	0,840	0,800	**0,850**	0,833	0,840
	$\mathcal{M}_{FGW}$	0,868	0,870	**0,897**	0,861	0,785	0,830	0,825	0,845
v_4	$\mathcal{M}_{GW}$	0,865	**0,900**	0,895	0,856	0,862	0,886	0,872	0,875
	$\mathcal{M}_{FGW}$	0,883	0,898	0,888	0,888	0,856	**0,892**	0,855	0,884

De modo geral, no cenário *CD02*, é possível verificar que o TROY se saiu melhor em 75% dos resultados obtidos com os experimentos. Isso quando comparado ao CRAWLER.

7. Considerações Finais

A Análise de Redes Sociais tem despertado grande interesse sócio econômico de instituições públicas e privadas, pois por meio dela é possível extrair padrões comportamentais e características dessas redes [Dorogovtsev and Mendes 2002]. Dentre os mais diferentes temas de estudos existentes na análise de redes sociais, a identificação de pessoas suspeitas de crimes vem se tornando um dos temas mais importantes [Pendar 2007]. Tal situação se justifica pelo fato de um crescente número de indivíduos utilizarem essas redes a fim de propagarem e/ou praticarem atos ilícitos [dos Santos and Guedes 2020] [Costa 2019] [Bretschneider et al. 2014].

Existem diversos métodos na literatura para identificação de suspeitos, mas muitos desses se baseiam em comunicações feitas em redes sociais, e acabam não considerando as interações existentes dentro de comentários e respostas, que tipicamente ocorrem no YouTube [Benevenuto et al. 2008] [Benevenuto et al. 2009] [Klausen et al. 2012]. Diante deste cenário, foi levantada a ideia de que ao analisar essas interações, seria possível melhorar a identificação de suspeitos, uma vez que se teria um maior conhecimento comportamental de um individuo e do impacto de um comentário e/ou resposta dentro da rede analisada.

A fim de investigar a ideia levantada, foi proposta a inclusão da etapa de *Recorte da Rede* [Florentino et al. 2021b] ao método INSPECTION [Florentino et al. 2021a] [Florentino et al. 2020b]. A funcionalidade dessa etapa é implementada pelo algoritmo TROY que define e representa as interações entre pessoas, por meio de comentários e

respostas, considerando o Nível de III de comunicação, no YouTube. Posteriormente, as demais etapas do INSPECTION são utilizadas a fim identificar as pessoas suspeitas.

Em experimentos preliminares, no domínio da pedofilia, foi observado que a identificação de suspeitos por meio de comentários e respostas, utilizando o Nível III de comunicação, através do uso do Algoritmo TROY, possibilitou melhores resultados [Florentino et al. 2021b]. Desse modo, a fim de obter uma maior confiança nos resultados alcançados em [Florentino et al. 2021b], no presente artigo, foram feitos novos experimentos considerando 4 (quatro) novos vídeos com alta propensão a comentários e respostas feitos por suspeitos de pedofilia. Devido à dificuldade em se obter um conjunto de dados previamente rotulado, em português, com suspeitos de pedofilia, primeiramente preparou-se o cenário *CD01*, no qual a marcação de suspeitos é feita de maneira minuciosa, de acordo com cada conteúdo textual disponibilizado por meio de comentários e respostas. Neste artigo, apresentamos também o cenário *CD02*, que integra pedófilos, presente no conjunto de dados PAN-2012-BR [Andrijauskas et al. 2017] [dos Santos and Guedes 2020], a pessoas presentes em cada um dos vídeos extraídos, utilizando a tarefa de predição de link.

As escolhas dos vídeos, em ambos os cenários, se justificam pelo fato de tratarem de temas com maior probabilidade de ocorrência de comentários e/ou respostas suspeitas. Consequentemente, essa característica facilitou a marcação de pessoas suspeitas. Além disso, é válido ressaltar que a marcação de suspeitos é utilizada apenas para avaliar o desempenho do método INSPECTION, a partir das representações dadas pelo TROY e CRAWLER, os quais, respectivamente, utilizam o nível de comunicação III e II.

Os resultados obtidos com os experimentos no domínio da pedofilia apontaram para a adequação do método proposto e mais uma vez confirmam a hipótese levantada, i.e., de que o algoritmo TROY, levando em conta o nível III de comunicação, apresenta melhores resultados que o algoritmo CRAWLER [Benevenuto et al. 2008], o qual leva em conta somente o nível II de comunicação. Utilizando o TROY, no cenário *CD01*, foram obtidos melhores resultados em 87,5% dos experimentos. Já no cenário *CD02*, o qual utiliza pedófilos reais, os melhores resultados foram em 75% dos experimentos.

O Algoritmo TROY pode ser utilizado em outras redes sociais, como por exemplo o Instagram e o Facebook, que usam o mecanismo de comentários e respostas em fotos e vídeos. Dessa maneira, pretende-se fazer experimentos futuros considerando essas redes. É válido ressaltar que para representação de vídeos e/ou fotos, considerando o nível III de comunicação por meio do Algoritmo TROY, em outras redes sociais, é necessário representá-lo como v_n, conforme descrito na seção 4.2. Ainda, em trabalhos futuros, incluem-se a definição de uma linha de corte adequada para o ranking de suspeitos, a exploração do aspecto topológico, realização de experimentos em diferentes domínios, teste de hipóteses e ponderação das classes raízes sem a necessidade de um especialista no domínio da aplicação.

8. Agradecimentos

Em primeiro lugar, os autores agradecem ao Paulo Renato da Costa Pereira, D.Sc., especialista em crimes de pedofilia da Policia Federal Brasileira, pelo seu grande apoio durante

os experimentos. Os autores agradecem também ao Projeto ARCo (Análise de Redes Complexas) do Instituto Militar de Engenharia, pelo apoio dado.

Referências

Aiello, L. M., Barrat, A., Schifanella, R., Cattuto, C., Markines, B., and Menczer, F. (2012). Friendship prediction and homophily in social media. *ACM Transactions on the Web (TWEB)*, 6(2):1–33.

Andrijauskas, A., Shimabukuro, A., and Maia, R. F. (2017). Desenvolvimento de base de dados em língua portuguesa sobre crimes sexuais. *VII Simpósio de Iniciação Científica, Didatica e de Ações Sociais da FEI*.

Benevenuto, F., Duarte, F., Rodrigues, T., Almeida, V. A., Almeida, J. M., and Ross, K. W. (2008). Understanding video interactions in youtube. In *Proceedings of the 16th ACM international conference on Multimedia*, pages 761–764.

Benevenuto, F., Rodrigues, T., Almeida, V., Almeida, J., and Gonçalves, M. (2009). Detecting spammers and content promoters in online video social networks. In *Proceedings of the 32nd international ACM SIGIR conference on Research and development in information retrieval*, pages 620–627.

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.

Corrêa, P., Gomes, C., de Carvalho Moura, A. M., and Cavalcanti, M. C. (2015). A multi-ontology approach to annotate scientific documents based on a modularization technique. *J. Biomed. Informatics*, 58:208–219.

Costa, A. O. (2019). Ciberterrorismo. *Intertem@ s ISSN 1677-1281*, 38(38).

da Silva Soares, P. R. and Prudêncio, R. B. C. (2012). Time series based link prediction. In *The 2012 international joint conference on neural networks (IJCNN)*, pages 1–7. IEEE.

Dorogovtsev, S. N. and Mendes, J. F. (2002). Evolution of networks. *Advances in physics*, 51(4):1079–1187.

dos Santos, L. F. and Guedes, G. (2020). Identificação de predadores sexuais brasileiros em conversas textuais na internet por meio de aprendizagem de máquina. *iSys-Brazilian Journal of Information Systems*, 13(4):22–47.

Dynel, M. (2014). Participation framework underlying youtube interaction. *Journal of Pragmatics*, 73:37–52.

Elzinga, P., Wolff, K. E., and Poelmans, J. (2012). Analyzing chat conversations of pedophiles with temporal relational semantic systems. In *2012 European Intelligence and Security Informatics Conference*, pages 242–249. IEEE.

Fernández, A. (2011). Clinical report: The impact of social media on children, adolescents and families. *Archivos de Pediatría del Uruguay*, 82(1):31–32.

Figueiredo, D. R. (2011). Introdução a redes complexas. *Atualizações em Informática*, pages 303–358.

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.

Florentino, É., Goldschmidt, R., and Cavalcanti, M. (2021a). Identifying suspects on social networks: An approach based on non-structured and non-labeled data. In *Proceedings of the 23rd International Conference on Enterprise Information Systems - Volume 1: ICEIS,*, pages 51–62. INSTICC, SciTePress.

Florentino, É. d. S., Goldschmidt, R. R., and Cavalcanti, M. C. R. (2021b). Exploring interactions in youtube to support the identification of crime suspects. In *XVII Brazilian Symposium on Information Systems*, pages 1–8.

Florentino, É. S., Cavalcante, A. A., and Goldschmidt, R. R. (2020a). An edge creation history retrieval based method to predict links in social networks. *Knowledge-Based Systems*, 205:106268.

Florentino, É. S., Goldschmidt, R. R., and Cavalcanti, M. C. (2020b). Identifying criminal suspects on social networks: A vocabulary-based method. In *Proceedings of the Brazilian Symposium on Multimedia and the Web*, pages 273–276.

Hobbs, J. R. and Pan, F. (2006). Time ontology in owl. *W3C working draft*, 27:133.

Huang, Z. and Lin, D. K. (2009). The time-series link prediction problem with applications in communication surveillance. *INFORMS Journal on Computing*, 21(2):286–303.

Kejriwal, N., Kumar, S., and Shibata, T. (2016). High performance loop closure detection using bag of word pairs. *Robotics and Autonomous Systems*, 77:55–65.

Klausen, J., Barbieri, E. T., Reichlin-Melnick, A., and Zelin, A. Y. (2012). The youtube jihadists: A social network analysis of al-muhajiroun's propaganda campaign. *Perspectives on Terrorism*, 6(1):36–53.

Kronk, C., Tran, G. Q., and Wu, D. T. (2019). Creating a queer ontology: The gender, sex, and sexual orientation (gsso) 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.

Kwon, K. H. and Gruzd, A. (2017). Is aggression contagious online? a case of swearing on donald trump's campaign videos on youtube. In *Proceedings of the 50th Hawaii International Conference on System Sciences*.

Lévy, P. and Feroldi, D. (1999). *Cybercultura: gli usi sociali delle nuove tecnologie*. Feltrinelli.

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.

Liben-Nowell, D. and Kleinberg, J. (2007). The link-prediction problem for social networks. *Journal of the American society for information science and technology*, 58(7):1019–1031.

Lü, L. and Zhou, T. (2011). Link prediction in complex networks: A survey. *Physica A: statistical mechanics and its applications*, 390(6):1150–1170.

Mariconti, E., Suarez-Tangil, G., Blackburn, J., De Cristofaro, E., Kourtellis, N., Leontiadis, I., Serrano, J. L., and Stringhini, G. (2019). "you know what to do"proactive detection of youtube videos targeted by coordinated hate attacks. *Proceedings of the ACM on Human-Computer Interaction*, 3(CSCW):1–21.

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 detesauros 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*, 156:129–137.

Neves, F. (n.d.). Elogios de a a z (*in Portuguese*).

Pendar, N. (2007). Toward spotting the pedophile telling victim from predator in text chats. In *International Conference on Semantic Computing (ICSC 2007)*, pages 235–241. IEEE.

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 tesauros e ontologias (*in Portuguese*). *Perspectivas em Ciência da Informação*, 14(1):99–116.

Santos, D. (2015). Predição de links em redes de coautoria: um estudo utilizando a teoria da evolução espectral em redes complexas. *Projetos e Dissertações em Sistemas de Informação e Gestão do Conhecimento*, 4(1).

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. In *CLEF (Online Working Notes/Labs/Workshop)*, volume 1178.

Wang, A. H. (2010). Don't follow me: Spam detection in twitter. In *2010 international conference on security and cryptography (SECRYPT)*, pages 1–10. IEEE.

Wang, P., Xu, B., Wu, Y., and Zhou, X. (2015). Link prediction in social networks: the state-of-the-art. *Science China Information Sciences*, 58(1):1–38.