Hidden in Plain Sight: Using Contact Information to Identify Sex Trafficking in Online Advertisements

Noelle Warkentin

School of Criminology
Simon Fraser University
Canada
noelle_warkentin@sfu.ca

Richard Frank

School of Criminology
Simon Fraser University
Canada
rfrank@sfu.ca

Abstract— Sex trafficking is an international crisis that has permeated the online realm, enabling traffickers to build a larger client base and evade law enforcement. The current study was conducted to identify sex trafficking by comparing advertisements (ads) with matching contact information. Ads with matching contact information and different descriptions of individuals may signify that one individual is controlling the sale of multiple people, potentially indicating trafficking. A customized web scraper was utilized to collect information from ads from the 'personals' category of Leolist.cc; those with social media handles, phone/WhatsApp numbers, email addresses, and 'click to view' numbers were extracted. Ads were grouped based on matching contact information and qualitatively analyzed; groups with different physical descriptions of girls were coded as potential trafficking. Ads with matching contacts and listings of multiple girls were searched with sex trafficking keywords and coded as potential trafficking or agency if at least half the ads contained over three keywords. From 7,328 unique contacts, 97 were coded as potential trafficking and 67 as potential trafficking or agency. The practical implication of this project is that an automated tool may be developed that uses these methods to identify trafficking advertisements to take them down, potentially aiding victims.

Keywords—Sex trafficking, classified ads, OSINT

I. INTRODUCTION

Sex trafficking within Canada and around the world is a major issue, with some estimating that globally, approximately 20 million adults and two million children are victims of this criminal enterprise [24]. However, these statistics are largely based on police reports, meaning that the true figure is probably much higher [22]. Further, it is estimated that between 95-98% of these sex trafficked victims are women and children [24, 22]. According to Public Safety Canada [22], "trafficking in persons, involves recruiting, transporting, transferring, receiving, holding, concealing, harbouring, or exercising control, direction or influence over that person, for the purpose of exploitation, generally for sexual exploitation or forced labour" (p.14). While this definition encompasses the different forms of human trafficking, for the purpose of this paper we are discussing trafficking for the intent of sexual exploitation. To elaborate then on this definition, one of the key elements of trafficking is coercion, which often involves the use or threat of violence, abuse of authority, deception, confinement, or preying on the victims' vulnerabilities [5, 22, 25].

While sex trafficking is an international issue, it is important to understand the scope of this activity within Canada. Millar & O'Doherty [19] identified that within Canada, 1,188 incidents of human trafficking resulted in 937 persons being charged between 2006 – 2017. Over three quarters of these persons were charged between 2014-2017, with the majority of cases coming from Ontario and Quebec [19]. Further, Indigenous women and girls are at an increased risk of being trafficked due to many circumstances surrounding the intergenerational trauma inflicted from Indian residential schools in Canada [23]. Due to the generational trauma this population has faced, there is a higher chance they have been involved in the child welfare system as well as the criminal justice system, both factors of which leads to a lack of self-esteem that makes them vulnerable to manipulation and recruitment from traffickers [23]. Traffickers, who take many different approaches to solicit their victims, are skilled at identifying the most vulnerable such as those who have experienced trauma or poverty [3, 16, 25].

Often, recruitment occurs through several avenues, such as online, through relationships, employment, nightlife, and from strangers [2]. Of these avenues, Baird et al. [2] identified that 36% of victims were recruited online through classifieds sites such as Kijiji and Craigslist, or through social media such as Instagram or Facebook. This finding highlights the use of technology in not only advertising the sale of victims for sexual exploitation, but in the recruitment of new victims. Further to this, technology enables traffickers to reach a larger audience of potential clients while also allowing them to evade law enforcement efforts [22, 13, 10]. For these reasons, new methods of identifying victims of sex trafficking online should be of vital importance. Therefore, the goal of the current study is to assess the usefulness of using different forms of contact information to identify potential instances of sex trafficking within online classifieds sites. This paper is structured as follows: firstly, a review of the literature on trafficking and technology, and methods used to identify victims will be presented, followed by the introduction of a novel methodology

of identifying traffickers used within the current study, and ending with results and discussion of the findings and their implications for law enforcement.

II. LITERATURE REVIEW

A. Sex Trafficking and Technology

Technology has led to many advancements in our society, both for the sharing of information as well as for the benefit of businesses around the world. While this shift towards a more connected society has affected positive change, it has also allowed nefarious actors to broaden their activities. Specifically, traffickers have benefitted in large part by using the internet to reach more clients [22, 13, 10]; likewise, with a perceived sense of anonymity, clients can now connect with like minded others, and have easier access to trafficked victims, meaning they are more likely to access these services without the fear of getting caught [13, 14]. In fact, the proof of traffickers using online classifieds sites as well as social media to advertise the sale of their victims is increasingly evident [17], and this is no small endeavor as trafficking is estimated to be a $150 billion industry [1].

Technology has proved extremely important for legitimate sex workers, who use online sites to advertise their services independent of third parties, while reducing any of the risks that are inherent with street work, including random violence from bystanders [6]. At the same time, sex workers advertising online can avoid arrest and they are less likely to come under aggression from perpetrators [6]. For these reasons, moving to the online space is an economical and safety benefit for those working in the industry, and the vast majority of the coordination of sex work takes place online [10]. The most common platforms sex workers use to advertise their services are classified sites with categories specifically for sex services, as well as social media sites such as Twitter and Instagram, and dating sites such as Plenty of Fish and Tinder [9, 6]. An unfortunate issue with this is that while sex workers are using the internet to their benefit, this offers traffickers another way to disguise their activities by creating advertisements with the pretence that the victim advertised is a non-trafficked sex worker [17, 18]. Traffickers use the same sites as sex workers to post their advertisements, and because these traffickers market their victims as sex workers, it is difficult to differentiate between advertisements that are legitimate and those that are criminal [14, 18, 1]. Often, traffickers will purport to be spa or massage parlors to further conceal their activities and post false ages or claims of working independently [1, 14]. Advertisements posted by traffickers will either be written in the first person or in third person, but in either case there is no mention of a pimp or trafficker [21], keeping the line blurry between who may be a victim and who may be a sex worker. Even though traffickers can hide in plain sight within these advertisements, they take further steps to disguise specific content such as contact information, so that it cannot be easily scraped by a web crawler [15, 16]. Further, traffickers benefit from using online advertising since they can easily facilitate the sale of very young victims, as the ages of these victims are not readily apparent to law enforcement [3].

Perhaps one of the more known examples of trafficking coming to light within these online spaces was when Craigslist came under fire for reports of sex trafficking within their adult services section [17]. Multiple news outlets had reported that the Craigslist adult section was being used by traffickers to advertise their victims,

and in the end, Craigslist shut down their adult services section [17]. However, while this shut down prevented traffickers from using this section on Craigslist, which was a major site for this activity, there are still many other websites which allow advertisements for sexual services, and traffickers made the move to these other spaces, such as Backpage and Myspace [17, 1]. Backpage specifically, which is a North American classifieds site, was catapulted into the spotlight for the sites extensive adult services section following the removal of the adult services section on Craigslist [12]. In fact, reports have indicated that Backpage is one of the largest online classifieds sites for adult services, with the site making approximately $27 million annually (Feyerick & Steffen, 2010 as cited in 12]. Therefore, it is clear that even though advertisements of these classifieds' sites may be removed, there will be others that can fill the gap.

Part of the reason law enforcement have such a difficult time identifying and prosecuting these cases is because these websites can be published anywhere, which means that jurisdictional boundaries are obscured, and because there is no accountability from internet service providers on what is published on their sites [10, 20]. Traffickers take advantage of this fact, especially knowing the benefits of using these services to reach a broader audience of clients [10, 20]. Thus, with the perceived sense of anonymity, and the knowledge that they can easily evade law enforcement within these online spaces, the supply and demand of this activity online will only continue to perpetuate [11, 14]. For these reasons, there has been a substantial shift towards trafficking within these online spaces [22, 11, 14], so methods to identify cases of trafficking should be considered of high importance.

B. Previous Methods to Identify Trafficking

As the identification of traffickers within these online spaces is a crucial endeavor, there have been many attempts by researchers to find ways to effectively detect those advertisements that may potentially be trafficking, and even so, there's never absolute certainty that the method is identifying real victims [9]. Given that distinguishing between advertisements from sex workers and trafficking victims is difficult [17, 18, 9], multiple methods have been employed in this venture. With these advertisements being so open within these online sites there is a plethora of data that can be used by anti-traffickers, researchers, and law enforcement to observe, monitor, track and research both the supply and the demand side of this activity [17, 9]. In fact, specific indicators have been established that can be used within these advertisements to pinpoint traffickers [4, 26].

1) Textual Analysis of Advertisements

Ibanez & Gazan [14] analyzed Backpage advertisements to establish physical environment indicators to classify a virtual sex trafficking indicator index. They did this by using a sample of 200 advertisements from three locations within the United States, measured the number of indicators present in each, along with the frequency of indicator types across the sample. Using the virtual indicators of sex trafficking they found that 75% of the advertisements in the sample were found to have these indicators [14]. Alternatively, Alvari et al. [1] first used an unsupervised approach to filter out data that is potentially trafficking. Data was then labeled by a law enforcement expert as well as a victim of sex trafficking to confidently state whether trafficking was occurring. Following this they trained a semi-supervised learner on

a portion of this data to identify labels for data that were not checked; the authors found that this approach in fact worked well at identifying trafficking within these advertisements [1].

2) Utilizing Machine Learning

Other researchers such as Dubrawski et al. [9] have used relevant text found within sex worker advertisements in combination with machine learning to identify which advertisements in their sample were trafficking. They were able to apply a classification algorithm to learn how to identify trafficking advertisements [9]. Using a different approach, L'Hoiry et al. [18] used a sex trafficking identification matrix (STIM) to flag advertisements with identified risk indicators specifically for police investigations. While the researchers noted that this method was useful, it is best when combined with the skills, knowledge, and experience of police officers. Lastly, using a dataset of 1,164,663 advertisements scraped from Backpage, Portnoff et al. [21] linked ads with co-occurring phone numbers and email addresses as ground truth to label whether the advertisements shared the same author. Following this, the authors developed a machine learning classifier to differentiate between advertisements that were posted by the same and different posters [21].

3) Harnessing Phone Numbers

Phone numbers have been found to be important not only for reviewers of escort services to access profiles [13], but also for the investigation of researchers and law enforcement. For instance, Giommoni and Ikwu [11] searched a sample of UK based sex work advertisements with ten indicators of human trafficking to identify whether trafficking was occurring. One of these indicators was phone numbers, which they used to correlate between the different advertisements to detect whether multiple advertisements were linked to the same number. While the authors did find 30 phone numbers associated with multiple girls, overall, their analysis did not find many advertisements within the website searched as being indicative of trafficking, in that the majority did not have three or more keyword indicators found [11]. Using a different method, Ibanez and Gazan [13] used phone numbers to track the movement of traffickers online by using area codes to gather when they moved to different locations; trafficking that moves to different locations is a sign of more sophisticated organized crime. By using advertisements from Backpage, these researchers were able to track patterns in movement throughout the United States, and by including a social network analysis they were able to identify the most prominent locations [13].

III. THE CURRENT STUDY

The current study was conducted to identify sex trafficking advertisements within a Canadian classifieds site. To the best of our knowledge, Canadian specific sites have not been investigated as thoroughly as websites specific to other countries. In addition, while phone numbers and email addresses have previously been investigated as being useful in detecting suspected trafficking [13, 11], other types of contact information have yet to be utilized in research of this type. Since social media is known to be used by traffickers [2, 1], there's the possibility that social media handles may be connected to traffickers within these advertisements. In addition, following a pre-analysis overview of the collected data, WhatsApp numbers were discovered within some of the advertisements; since WhatsApp is an encrypted

communication platform, it's reasonable to assume that traffickers will take advantage of the use of this platform, and so these numbers were deemed appropriate for inclusion.

Advertisements that share the same contact information, yet have different descriptions of individuals, may signify that there is one individual controlling the sale of multiple people. This scenario could be evidence that someone other than the advertised individual is in control of the sale, potentially indicating trafficking. The aim of the current study was to identify suspected sex trafficking within online Canadian classifieds sites by connecting different advertisements with matching contact information: Email addresses, phone numbers, WhatsApp numbers, and social media handles. To accomplish this, each type of contact information was extracted from a dataset of advertisements scraped from the Canadian classifieds site Leolist.cc.

IV. METHODS

The purpose of this study was to determine whether sex trafficking could be identified in a sample of online classifieds ads by comparing contact information and the physical descriptions posted within advertisements. Since we were specifically interested in trafficking within Canada, the online Canadian classifieds site Leolist.cc was chosen for analysis. Leolist.cc was determined to be appropriate for analysis as it boasts a large number of postings [4], specifically those within the 'personals' category which contains subsections related to sex work, such as: Gigs & Jobs, Dom & Fetish, Female Escorts, Female Massage, Male Escorts, Transsexual Escorts, Sugar Babies, & Hook Up. Further, Leolist.cc was flagged as a site that potentially hosts sex trafficking within its advertisements [4], therefore, we determined that this would be an appropriate site for analysis.

A. Data Capture

The 'personals' category of Leolist.cc was scraped of all content using a customized web crawler (The Dark Crawler). The crawler, which used python coding to scrape the site, was designed to be used specifically on Leolist.cc. Information from every advertisement posted under the 'personals' category was captured up until September 10, 2022, at which point data was saved into a database and downloaded into a .txt file for analysis. Data capture included the following information: Advertisement ID, Subject ID, Availability, Tagline, Description, Subject ID, Link, Post date, Age, Gender, Hair colour, Eye colour, Height, Weight, and Body Stats.

B. Data Analysis

The scraped dataset was parsed using python coding to extract details from all ads with instances of the different types of contact information: Email addresses, phone numbers, WhatsApp numbers, social media handles, and 'click to view' numbers. Since the Regex code for phone numbers and WhatsApp numbers is very similar, these two types of contact information were included together. In addition, many of the advertisements in our sample had contact information that was hidden behind 'click to view' CAPTCHAs (which stands for "completely automated public Turing test to tell computers and humans apart"). These 'click to view' lines were often followed by four digits, or an email domain name. Since it could be possible to match the last four digits in these 'click to view' lines, we decided it could be beneficial to include them in the analysis. However, 'click to view' email domain names

were not included as it would not be feasible to verify each email address with no other uniquely identifying information.

Python coding was used to search the dataset, with specific regular expressions (Regex) used to extract each type of contact information separately (see Table 1 for Regex used). This was done so that each distinct contact type could be analyzed independently. The four new files, separated by type of contact information extracted, were then exported into Excel for analysis. Each file contained all the previously mentioned information originally collected with the advertisements (see Data Capture), along with a new column that contained the extracted contact information. Criterion sampling was then employed on the dataset; we were only interested in analyzing groups of matching contact information, and so data were first sorted in ascending order, which made it easier to identify multiple advertisements with the same contact information. We then copied any advertisements that did have matching contact information and pasted them into a new tab (this was done separately for each unique contact).

Table 1. REGULAR EXPRESSIONS USED TO SEARCH FOR CONTACT INFORMATION

Contact Information	**Regex**
Emails	r'([a-zA-Z0-9_.+-]+@[a-zA-Z0-9-]+\.[a-zA-Z0-9-.]+)'
Phone/WhatsApp numbers	r'(?:\+\d{2})?\d{3,4}\D?\d{3}\D?\d{3}'
Social media handles	r'@([A-Za-z0-9_]+)'
Click to view numbers	r'click to view \-[\d]+[\d]+[\d]+[\d]'

1) Coding Round One

Following the criterion sampling of advertisements with multiple contact information into new tabs, qualitative analysis was conducted to determine whether the advertisements with the associated contact information were potentially trafficking. Advertisements were read through and compared to identify whether the physical descriptions matched across advertisements or were different. To determine a difference between descriptions within the advertisements, we focused mainly on body stats, height, weight, and ethnicity. These descriptors would show clear differences in the individuals advertised yet would be more difficult to lie about compared to name or age. Based on this information, groups with matching contact information were assigned to one of two deductively chosen codes: Potentially trafficking or agency. Deductive coding was chosen as the best option for analysis as the main factor we were looking to identify was whether trafficking could be found within the advertisements. Since we were basing this on the requirements of matching contact information of advertisements with different descriptions of individuals, we only needed codes that could label these specific instances.

If advertisements with the same contact information were found to differ in the descriptions of individuals, they were coded as being potential trafficking; if advertisements contained similar descriptions they were not included in the coding as they could potentially be a sex worker advertisement. However, if the advertisement made mention of multiple girls in any way, they were initially deductively coded as agency (in reference to any escort agency, massage parlor, or spa).

a) Verifying ‘Click to View’ Numbers

Since the captured ‘click to view’ numbers only provided the last four digits, we needed to manually confirm the entire number to ensure that groups of advertisements indeed shared the same number. To verify these numbers, an extra step was included specifically for the ‘click to view’ numbers, where the URL link that accompanied the ad was followed, and the CAPTCHA was completed to view the entire number. This was completed on a random number of advertisements with matching last four digits to confirm the number was the same. Groups of numbers that were found to not match were then excluded from analysis, as this did not meet our criteria.

2) Coding Round Two

Once the initial round of coding was complete, a second round of coding, which involved the use of keywords was employed on those groups coded as agency. Since previous research has identified that traffickers may obfuscate their activities by claiming to be massage parlors or spas [1], determining which of those coded under agency may potentially be trafficking was important. While emojis have been found in the past to be useful indicators of trafficking [8], they were not captured in the current study as the focus was on contact information. Since textual information from the ads were captured, keywords were the chosen method to differentiate potential trafficking from agency ads. To complete this, a deductive keyword search was conducted on the groups of advertisements coded as agency. Specifically, 50 keywords that have been found to be indicators of sex trafficking were used for this search; these keywords were borrowed from Cartwright et al. [4] and UNODC [26], who had both identified them as being general and reliable indicators for the purpose of identifying sex trafficking, and had been used to identify trafficking previously.

Based on the findings from their report, a minimum of three keywords within an advertisement was required for it to be considered as potential sex trafficking [4]. For this reason, all 50 keywords were searched in all the advertisements within the groups coded as agency, and a count of how many keywords were found per advertisement was conducted. Since advertisements may differ, there would not be consistency with the number of keywords found per advertisement. Therefore, if a group of advertisements that shared contact information had approximately half of the advertisements with three or more keywords found, we determined this was enough to indicate potential trafficking. Since the requirement of at least three keywords found in one advertisement could imply trafficking, the same contact information tied to at least some advertisements that met this criterion could signify that trafficking is occurring. If far less than half of advertisements met the minimum requirement for keywords, the group would continue to be coded as agency. Thus, criterion sampling was employed and only those agency coded groups that had approximately half of advertisements with three or more keywords were re-coded as potential trafficking or agency.

Following this keyword search, those coded as agency were no longer used in the remainder of the analysis. Since the purpose of this study was to ascertain whether contact information could be used to identify trafficking, any advertisements that were thought to be a sex worker or an agency were no longer needed. After distinguishing between agency and potential trafficking, advertisements within the codes

of potential trafficking and potential trafficking or agency were checked once more to ensure that coding was completed accurately.

V. RESULTS

A total of 198,679 advertisements were extracted from the 'personals' category in Leolist.cc. From this sample, we used python coding to extract email addresses, phone/WhatsApp numbers, social media handles, and 'click to view' numbers, and discovered a total of 126,250 advertisements with at least one. Unfortunately, we were unable to gather a large enough sample of email addresses, as our extraction only identified eight advertisements with this information. It's possible that the code we used was unable to capture any instances where the poster may have inserted spaces into the address or other ways to disguise the address, thus we did not include these advertisements in our analysis.

Of the contact information types that were kept for analysis, 97 were coded as potential trafficking, and 67 were coded as potential trafficking or agency (see Table 2 for a breakdown). From the advertisements that were coded as potential trafficking and potential trafficking or agency, the number of advertisements in each group ranged from two to 1,176, with one group of matching phone numbers having 4,968 advertisements. The 4,968 advertisements associated with one phone number were found to be extremely similar, if only for minor differences in spacing or a change of character or capitalization. It's important to note however that even though some advertisements differed in their description of individuals, some were repeat posts with the same descriptions. This indicates that although contact information may be useful in identifying trafficking within these spaces, the number of advertisements connected may not be an indicator of how many victims are being trafficked in connection to that contact.

Table 2. BREAKDOWN OF CONTACT INFORMATION FOUND, ALONG WITH TOTAL CODED

Contact Information	Total Unique	Potential Trafficking	Potential Trafficking or Agency
Phone/WhatsApp number	2,672	44	59
Social media handles	868	5	6
Click to view numbers	3,788	48	2
Total	7,328	97	67

A. Phone/WhatsApp Numbers

In total, 19,557 advertisements were extracted that had either phone numbers, WhatsApp numbers, or both. So long as the advertisements had the same number for either of these they were grouped together as one. From the total number of advertisements, 2,672 of the numbers were unique. Of these unique numbers, in the first round of coding 44 were coded as potentially trafficking and 79 as agency. Following this initial round of coding a second round of coding was applied which distinguished between agency and potential trafficking or agency. Through this second round we identified 59 as potentially trafficking or agency. Upon examining the numbers provided in the final coded groups, it was identified that these were all phone numbers, meaning none of the matching sets of WhatsApp numbers were coded as potential trafficking or potential trafficking or agency. Table 3 depicts

samples of advertisements with matching phone numbers coded as potential trafficking.

Table 3. EXAMPLES OF DIFFERENT ADVERTISEMENTS FOUND WITH SAME PHONE NUMBER[1]

Code	EX1	EX2
Potential Trafficking	I am [redacted]. 5f3,36d natural 116 lbs .slim and Pretty Face with long hair, soft silky skin. Independent beautiful & slim with very thin waist. I provide a very sensual massage with lots of teasing and build up. I like to make you relax , happy and we can enjoy good time together. You will sure be satisfied I do very good massage first Open-Minded Menu and Shower Together (No Rush) 4days in surrey txt only Ivy 581~309~[redacted]	I am [redacted] 5f5 ,36d natural 135 lbs .slim and Pretty Face with long hair, soft silky skin. Independent beautiful & slim with very thin waist. I provide a very sensual massage with lots of teasing and build up. I like to make you relax , happy and we can enjoy good time together. You will sure be satisfied I do very good massage first Open-Minded Menu and Shower Together (No Rush) 4days in hotelÂ southgateÂ txt only Nina 581~309~[redacted]

As can be seen in Table 3, advertisements coded as potential trafficking contain information about the individuals such as height, weight, and name that differ. The fact that the ads contain the same phone/WhatsApp numbers for contact, and yet descriptions of the individuals differ, could indicate trafficking. For those coded as potential trafficking or agency, while the ads were relatively similar and suggestive of being an escort agency or spa, at least three keywords associated with sex trafficking were identified in the ads. Because of the shared phone/WhatsApp number as contact information and the identified keywords there is a possibility that these are in fact ads of sex trafficked individuals.

B. Social Media Handles

A total of 3,582 advertisements extracted had what appeared to be social media handles, meaning they had an asperand followed by a string of characters that is indicative of a social media handle. So long as the advertisements had the same handle they were grouped together as one. From the total number of advertisements, 868 of the numbers were unique, and following the first round of coding, five groups of matching social media handles were coded as potentially trafficking and 14 as agency. After this initial round, the second round of coding established that six of the agency coded groups could be re-coded as potentially trafficking or agency based on keywords found. Table 4 depicts samples of advertisements with matching social media handles coded potential trafficking.

Table 4. EXAMPLES OF DIFFERENT ADVERTISEMENTS FOUND WITH THE SAME SOCIAL MEDIA HANDLES

Code	EX1	EX2

[1] For anonymity purposes names, numbers and social media handles have been hidden.

Potential Trafficking	click to view -2274 Â Â @tempg[redacted] Hi gents I'm a beautiful curvaceous blonde with fun bubbly personality (and booty)that will put you at ease. I love to laugh and have an amazing time in the presence of my companion, let's meet xo 5f7 36DD all naturalÂ Dirty Blonde Hair & Green Eyes 180hh 215.45m 260hrÂ APP'S / voip/ private numbers will be ignored, Screening is required for safety, respectful gents only, shower is mandatory upon arrival xoxo please do not book if you are feeling unhealthy in anyway	To inquire about my availability & Prebooking please text click to view -2274 See you soon xo @tempg[redacted] @date[redacted] Onlyfans: @date[redacted] Introducing [redacted] Love A petite lovers dream Sensual, warm, playful companion who loves to make lasting connections with people from all walks of life Until we meet [redacted] Love xo Stats: 5' 95lbs 32C-24-32 Well proportioned petite figure all natural, no enhancements and no tattoos Rates: 220hh 300hr OUTCALLS to select hotels Downtown + travel fee.Â NICE & respectful gents only, Screening is required for safety APPS / VOIP/ PRIVATE #'S will be ignoredÂ please do not book if you are feeling unhealthy in anyway

As can be seen in Table 4, advertisements coded as potential trafficking contain information about the individuals such as height and body measurements that differ. The fact that the ads contain the same social media handles for contact, and yet descriptions of the individuals differ, could indicate trafficking. While those coded as potential trafficking or agency had relatively similar ads that were suggestive of being an escort agency or spa, at least three keywords associated with sex trafficking were identified in the ads. Because of the shared social media handles across ads and the identified keywords there is a possibility that these are in fact ads of sex trafficked individuals.

C. 'Click to View' Numbers

Lastly, 103,110 advertisements were extracted that contained the ‘click to view’ numbers. As this would be an extraordinary number of advertisements to parse through, it was decided that only half of this data would be analyzed. Once ‘click to view’ numbers were sorted in ascending order to find matching sets, the first 50,409 of these advertisements were copied into a new sheet for analysis. From this new sample of advertisements, 3,788 were unique, and from the first round of coding 48 were coded as potentially trafficking and 26 as agency. In the second round of coding, two of the groups coded as agency were then re-coded as potentially trafficking or agency. Therefore, the majority of those coded as agency were not found to be potentially trafficking, which may indicate that those numbers hidden behind ‘click to view’ CAPTCHAs are not marketing themselves as a type of agency. Table 5 depicts samples of advertisements with matching ‘click to view’ numbers coded as potential trafficking.

Table 5. EXAMPLES OF DIFFERENT ADVERTISEMENTS FOUND WITH SAME 'CLICK TO VIEW' NUMBER

Code	EX1	EX2
Potential Trafficking	Hi guys! My name is [redacted] , very sweet and pretty girl 20y ears old. 5'3, 105lbs, 34D-24-34. All Nature Beautiful package with firm fit sexy body Exotic Asian with smooth milk skin and delicious curves, smoky hot body. You'll love my service and sweet personality The place is clean and comfortable. Easy parking Rate 180/hh 200/45mins 240/hr dongguan style or Nuru special service 340/hr never in a hurry and always ready to cater to your coming I do best massage,bbj ,fs and c1m,never rush! Why wait ? Please call or text : Â click to view -0810 Location:Â Richmond	Hi, gentsÂ I am [redacted] I'm in gp at the momentÂ Long waisted black hair Petite 5'1hour glass figure Beautiful eyes with natural big lips .No block calls pleaseÂ My pics are 100% Real and verified. Don't miss out on this once in a life time experience. .Â (click to view -0810 new number

As can be seen in Table 5, advertisements coded as potential trafficking contain information about the individuals such as height and body description that differ. The fact that the ads contain the same 'click to view' phone numbers for contact, and yet descriptions of the individuals differ, could indicate trafficking. Ads coded as potential trafficking or agency were relatively similar and suggestive of being an escort agency or spa. However, because of the shared 'click to view' phone numbers and the identification of at least three keywords associated with sex trafficking there is a possibility that these are in fact ads of sex trafficked individuals.

VI. DISCUSSION

The aim of the current study was to identify suspected sex trafficking within online Canadian classifieds sites by connecting different advertisements with matching contact information: Email addresses, phone numbers, WhatsApp numbers, and social media handles. Previous researchers have used phone numbers to identify potential trafficking within online advertisements, or to track the movement of suspected traffickers [11; 13], but to date, there has been a lack of research looking into different types of contact information. Phone numbers may not be as stable as other contact methods, since many of these may be connected to burner phones, therefore, detecting multiple types of contact information that could be connected to traffickers may be a valuable approach. In the current study, we used social media handles, phone numbers, and 'click to view' numbers (phone numbers hidden behind CAPTCHAs) to identify 164 contacts that may be trafficking, from a total of 7,328 unique contacts. Specifically, 97 unique contacts were identified as potential trafficking due to having differing descriptions of physical characteristics of the individuals listed. Another 67 were identified as potential trafficking or agency due to having multiple girls listed in each advertisement, while also containing keyword indicators of trafficking [4, 26] in approximately half of the advertisements. While

the number of advertisements identified as potential trafficking in this study are undoubtedly a significant underestimation of the true number of trafficking cases, the methods used in the current study may still be a valuable addition to assisting with detection of suspected traffickers within online sites.

Unfortunately, we were only able to extract eight email addresses, and so were not able to use this contact type for analysis. Due to either not using email addresses in the advertisement, or obfuscation techniques such as adding extra spacing between characters, our coding was not able to extract email addresses from the collected data. Obfuscating contact information is not new, as previous researchers have discovered various techniques posters use on phone numbers [15, 16]. Finding different ways of capturing this information is important, as this could aid in identifying advertisements that may not have other types of contact information. Of the unique contact information types that were extracted, the majority were phone/WhatsApp numbers (2,672) and 'click to view' numbers (3,788), with only 868 unique social media handles identified. Upon completion of coding, it was discovered that none of the coded phone/WhatsApp numbers contained any WhatsApp numbers; while there were WhatsApp numbers in the 2,672 unique numbers captured, none of these were connected to advertisements in either of the final codes. This could indicate that traffickers may not use WhatsApp often, or that their advertisements weren't detected with the methods employed in this study.

Another interesting finding was that there were slightly more contacts that were found to be potential trafficking or agency compared to those coded as potential trafficking of the social media handles and phone numbers. The largest difference was in the phone numbers in which the majority (59) of the 79 coded as agency were re-coded as potential trafficking or agency; the split was even in the social media handles, with six being re-coded of the 14 agency codes. In comparison, there were far more coded as potential trafficking within the 'click to view' numbers. Of the 26 originally coded as agency, only two of these were re-coded as potential trafficking or agency. This could potentially signify that traffickers marketing themselves as a type of agency are less likely to hide their contact information behind CAPTCHAs; conceivably, traffickers marketing themselves as an agency of some type, whether that be an escort service, spa, or massage parlor, might feel as though they have enough cover and would rather not take the time to conceal their contact information through CAPTCHAs.

A. Limitations

One thing to consider is that while these methods were efficient at capturing the contacts of those advertisements with different physical descriptions, keywords were still needed to differentiate between those that may be legitimate agencies, and those that may be trafficking. Therefore, the methods we employed may be useful to law enforcement at identifying certain advertisements, however, a keyword search would still need to be applied to those being marketed as agencies. Further to this, it must be noted that even with keywords it is difficult to ascertain with more certainty that an advertisement is posted by a trafficker rather than an agency. As some of the keywords identified as indicators of trafficking could also easily be explained as a business. For example, an agency might use terms that we identified as indicators such as "girls", "7 days", and "male operated" as legitimate descriptors

of their business; these agencies will have multiple girls, they may be open 7 days a week, and they may be male operated. So, while these terms have been found to indicate trafficking, it is also important to note how complicated it can be trying to distinguish agencies that employ non-trafficked sex workers between those that are trafficked. Since traffickers have been known to imitate spa or massage parlors to conceal their activities and post false ages or claims of working independently (1, 14], the distinction between agencies and traffickers is extremely blurry.

Another limitation in the study was that we were only able to collect and analyze data from one Canadian classifieds site; by analyzing all classifieds sites we may be able to determine which sites have the greatest number of suspected trafficking cases, and this could narrow down the scope of investigations for law enforcement. In addition, we were not able to extract enough email addresses for analysis, and so this is something that should be improved on in future studies, as using email addresses may aid in identifying more cases of potential trafficking within these sites. Finally, it is not possible to know for certain whether an advertisement is trafficking or not [9], therefore, we cannot state that these contacts are trafficking, we can only surmise based on our findings that they are potentially trafficking, so this is where law enforcement may be able to take over.

B. Recommendations for Law Enforcement

A logical recommendation would be to hold internet service platforms accountable for the advertisements posted on their sites; indeed Delateur [7] proposes holding these websites responsible and prosecuting those at the top. However, there are jurisdictional issues which may not be easily mitigated. If a site is hosted in a foreign country. For example, Leolist.cc is hosted in Budapest, Hungary, so local law enforcement in that country must be willing to cooperate to shut down the site. Alas, this might be something that takes years to accomplish, and may not be feasible as an immediate solution. Perhaps one recommendation that may have slightly more of an affect is by contacting the known legitimate businesses who pay for advertisements within these sites. Feasibly, the knowledge of a company being associated with a site that aids in the facilitation of trafficking will deter them from continuing their relationship for fear of public and financial repercussions (similar to what happened with Craigslist [17]). Nevertheless, it is recommended that the methods identified in the present study be utilized by law enforcement to aid in identifying active cases of potential trafficking. By proactively monitoring these sites, law enforcement may be able to keep track of, and possibly intervene on, specific cases if they are able to simultaneously gather additional information.

VII. CONCLUSION

With the rise of technology, there has been a significant shift in traffickers utilizing online spaces to broaden their client base while also remaining hidden from law enforcement [22, 11; 14]. The issue with identifying these traffickers is that they will hide under the guise of sex worker advertisements, often falsely claiming to be independent workers, a spa, or massage parlor [14, 18, 17]. With these concealing techniques it is difficult to distinguish between sex worker advertisements and those with trafficked victims [17, 18, 9]. For this reason, efficacious methods for identifying potential trafficking advertisements are extremely important. Using a customized web scraping tool, we captured all ads from the ‘personals’ category of the

Canadian classifieds site Leolist.cc and extracted all advertisements that contained phone/WhatsApp numbers, social media handles, and 'click to view' numbers. Using this information, we were able to identify multiple advertisements connected to matching phone numbers and social media handles. This method may be used to build an automated tool that can be used by law enforcement to extract advertisements with matching contact information in real time, thus aiding in investigations at a more efficient rate. To conclude, this study has contributed a novel approach to identifying suspected sex traffickers within online classifieds sites and may be useful for future researchers to expand on, or to assist law enforcement in their efforts.

REFERENCES

[1] Alvari, H., Shakarian, P., & Snyder, J. E. K. (2016). A Non-Parametric Learning Approach to Identify Online Human Trafficking (arXiv:1607.08691). arXiv. http://arxiv.org/abs/1607.08691

[2] Baird, K., McDonald, K. P., & Connolly, J. (2020). Sex trafficking of women and girls in a southern Ontario region: Police file review exploring victim characteristics, trafficking experiences, and the intersection with child welfare. Canadian Journal of Behavioural Science / Revue Canadienne Des Sciences Du Comportement, 52(1), 8–17. https://doi.org/10.1037/cbs0000151

[3] Barrett, N. A. (n.d.). AN ASSESSMENT OF SEX TRAFFICKING. 58.

[4] Cartwright, B., Frank, R., Sharifi, T., & Pannu, M. (2022). Deploying artificial intelligence to detect and respond to the use of digital technology by perpetrators of human trafficking [Unpublished report].

[5] Chew, L. (2022). Global Trafficking in Women: Some Issues and Strategies. 9.

[6] Cunningham, S., Sanders, T., Scoular, J., Campbell, R., Pitcher, J., Hill, K., Valentine-Chase, M., Melissa, C., Aydin, Y., & Hamer, R. (2018). Behind the screen: Commercial sex, digital spaces and working online. Technology in Society, 53, 47–54. https://doi.org/10.1016/j.techsoc.2017.11.004

[7] Delateur, M. J. (n.d.). From Craigslist to Backpage.com: Conspiracy as a Strategy to Prosecute Third-Party Websites for Sex Trafficking. SANTA CLARA LAW REVIEW, 56, 63.

[8] Diba, P., Antonopoulos, G., & Papanicolaou, G. (2017). Improving and sharing knowledge on the Internet's role in the human smuggling and trafficking process.

[9] Dubrawski, A., Miller, K., Barnes, M., Boecking, B., & Kennedy, E. (2015). Leveraging Publicly Available Data to Discern Patterns of Human-Trafficking Activity. Journal of Human Trafficking, 1(1), 65–85. https://doi.org/10.1080/23322705.2015.1015342

[10] Farley, M., Franzblau, K., & Kennedy, M. A. (n.d.). ONLINE PROSTITUTION AND TRAFFICKING. Albany Law Review, 77, 57.

[11] Giommoni, L., & Ikwu, R. (2021). Identifying human trafficking indicators in the UK online sex market. Trends in Organized Crime. https://doi.org/10.1007/s12117-021-09431-0

[12] Gow, G. A., Barlott, T., Quinn, K., Linder, J., Soler, A., Edwards, G., & Hossain, S. (2015). Project Backpage: Using Text Messaging to Initiate Outreach

Support for Victims of Human Trafficking and Sexual Exploitation. Journal of Human Trafficking, 1(4), 259–274. https://doi.org/10.1080/23322705.2015.1016769

[13] Ibanez, M., & Gazan, R. (2016a). Detecting sex trafficking circuits in the U.S. through analysis of online escort advertisements. 2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM), 892–895. https://doi.org/10.1109/ASONAM.2016.7752344

[14] Ibanez, M., & Gazan, R. (2016b). Virtual indicators of sex trafficking to identify potential victims in online advertisements. 2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM), 818–824. https://doi.org/10.1109/ASONAM.2016.7752332

[15] Kejriwal, M., & Szekely, P. (2017). An Investigative Search Engine for the Human Trafficking Domain. Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 10588 LNCS, 247–262. https://doi.org/10.1007/978-3-319-68204-4_25

[16] Keskin, B. B., Bott, G. J., & Freeman, N. K. (2021). Cracking Sex Trafficking: Data Analysis, Pattern Recognition, and Path Prediction. Production and Operations Management, 30(4), 1110–1135. https://doi.org/10.1111/poms.13294

[17] Latonero, Mark, Human Trafficking Online: The Role of Social Networking Sites and Online Classifieds (September 1, 2011). http://dx.doi.org/10.2139/ssrn.2045851

[18] L'Hoiry, X., Moretti, A., & Antonopoulos, G. A. (2021). Identifying sex trafficking in Adult Services Websites: An exploratory study with a British police force. Trends in Organized Crime. https://doi.org/10.1007/s12117-021-09414-1

[19] Millar, H., & O'Doherty, T. (2020). Racialized, Gendered, and Sensationalized: An examination of Canadian anti-trafficking laws, their enforcement, and their (re)presentation. Canadian Journal of Law and Society / Revue Canadienne Droit et Société, 35(1), 23–44. https://doi.org/10.1017/cls.2020.2

[20] Moser, M. (2021). Human Trafficking and Online Platform Liability. University of Toledo Law Review, 53, 21.

[21] Portnoff, R. S., Huang, D. Y., Doerfler, P., Afroz, S., & McCoy, D. (2017). Backpage and Bitcoin: Uncovering Human Traffickers. Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 1595–1604. https://doi.org/10.1145/3097983.3098082

[22] Public Safety Canada (2019). National Strategy to Combat Human Trafficking, 2019-2024. Government of Canada.

[23] Sexual Exploitation and Trafficking of Aboriginal Women and Girls: Literature Review and Key Informant Interviews: Final Report. (2014). Native Women's Association of Canada.

[24] Shelley, F. M., & Metz, R. (2017). Prostitution, the Sex Trade, and Trafficking. In Geography of trafficking: From drug smuggling to modern-day slavery. essay, ABC-CLIO.

[25] United Nations Office on Drugs and Crime. (2020a). Global Report on Trafficking in Persons 2020

[26] United Nations Office on Drugs and Crime. (2020b). Human Trafficking Indicators.